

**OVERBLOG**

<http://seo-optimisation.over-blog.net/>

## **L'optimisation des pages Web pour le référencement**

Aujourd'hui, chaque page publiée sur Internet a besoin d'être adaptée aux besoins des moteurs de recherche. Pour optimiser ses pages, un Webmaster peut s'adresser à un professionnel, participer aux conférences ou s'inscrire dans une formation online.

Je ne parle donc pas de "comment" optimiser ses pages mais j'explique les mots utilisés en optimisation. Vous pouvez donc, peut-être pas chaque jour, mais quand même régulièrement apprendre le vocabulaire des référenceurs. Je propose donc un approche un peu différent à ce métier.

Les textes publiés dans le blog ainsi que celles téléchargés en format PDF sont protégés par le droit d'auteurs européens ainsi que par les copyrights internationaux et servent exclusivement à l'utilisation privée. La reproduction digitale ou imprimé est interdite, mais nous acceptons et permettons chaque lien qui cible n'importe quelle page du blog sans autorisation supplémentaire. La seule condition et à nommer dans le lien le titre de l'article, l'adresse du blog ou l'adresse de la page ciblée.

### [abréviations](#) (09/05/2009 )

Presque chaque page Web utilise des abréviations sans que l'auteur du texte réfléchisse sur les problèmes qu'une abréviation cause pour le référencement.

Il n'y a pas seulement Firefox et Internet Explorer mais aussi de navigateurs à synthèse vocale qui ont besoin des explications dès que le Webmaster de la page utilise une abréviation, un acronyme ou un mot dans une langue étrangère. Ces explications se transmettent à l'aide des balises, des attributs et des valeurs.

C'est aussi la raison pourquoi les règles de l'accessibilité obligent l'auteur d'une page Web à expliquer les abréviations par de valeurs et de faire, en plus, une différence entre une abréviations dans la lange par défaut de la page et une abréviation dans une autre langue.

Des informations plus détaillées sur les abréviations se trouvent par le lien suivant : [L'empli des abréviations et l'optimisation](#)

### [accessibilité](#) (29/04/2009 )

Même si les algorithmes des moteurs de recherche prennent les paramètres de l'accessibilité des pages Web en considération, la majorité des Webmasters ne sait pas encore ce que cela implique lors de l'optimisation pour le référencement.

Accessibilité ne signifie rien d'autre qu'éviter à discriminer au moins 10 pour cent des internautes. Pour être plus claire, respecter l'accessibilité signifie à créer et programmer les pages Web de la manière que chaque internautes, même s'il a de problèmes à voir les couleurs "correctement", s'il a de problèmes d'ouïe ou s'il a un autre handicap courant, puisse profiter entièrement de sa visite.

Respecter l'accessibilité signifie donc, par exemple, maîtriser les couleurs, adapter ses arrière-plans, utiliser les balises qui marquent des abréviations et acronymes, ne pas oublier les textes alternatifs des images etc.

Plus d'informations sur l'accessibilité se trouvent dans la page suivante : [L'accessibilité des pages Web](#)

### [adjacence](#) (24/03/2008 )

Si un internaute cherche une expression qui se compose de deux ou plusieurs mots dans un moteur de recherche, le moteur affiche en premier lieu les pages où ces mots sont assez proches dans le texte. On parle ici de l'adjacence.

Si l'internaute cherche donc les mots "optimisation" et "lien" la page qui utilise l'expressions "... optimisation des liens ..." occupera une place plus haute que la page avec l'expression "... optimisation des pages web les liens ...".

Il est donc important dans l'optimisation des pages Web de prévoir plusieurs possibilités de recherche et de penser à l'adjacence. C'est en règle générale le rôle du *web copywriting*, donc d'un rédacteur du contenu Web.

### [adresse IP](#) (25/03/2008 )

Chaque nom de domaine est représenté par une adresse IP numérique. Des nombreux sites n'ont toutefois pas de IP unique mais une IP partagé vu qu'ils ne sont pas hébergés dans un propre serveur mais se situent dans des serveurs partagés.

Lors de l'optimisation plusieurs problèmes devront donc être pris en considération. Si la même adresse IP est ou été utilisé par un spammeur, il est quasiment impossible d'entrer dans une base de données des moteurs de recherche.

L'optimisation idéale se base donc sur le fait qu'un site dispose d'une adresse IP propre, sans faille et qui ne se faisait jamais remarqué par une utilisation frauduleuse.

### [adword site](#) (19/01/2008 )

*On parle d'un adword site (= MFA site = **Made For Adsense**)* si le site entier est exclusivement construit pour favoriser le système *Google Adsense*, donc pour gagner de l'argent dès qu'un internautes cliquent sur les annonces affichés.

De tels sites ne rapportent en règle générale pas le résultat espéré et risquent d'être mal référencé vu que les méthodes de l'optimisation souvent ne peuvent pas être appliqués à de tels sites. Il s'y ajoute le problème que la majorité des internautes n'aiment pas les sites qui débordent de publicité.

Il y a toutefois des exceptions. Une page qui présente seulement de programmes TV, cinéma ou autres se prête par excellence comme *adword site*. Les annonces publiées sont toujours en relation avec le programme et donnent donc en règle générale du résultat. La majorité des *adword sites* ne sont pas conforme au contracte avec Google et font donc partie du *Black HAT SEO*.

### [algorithme](#) (14/01/2008 )

L'algorithme est la base de la structure d'un programme qui cherche à évaluer la valeur neutre d'une page Web avant de l'indexer. Il y donc de centaines de valeurs qui font partie d'un algorithme.

Comme les référenceurs cherchent a profiter des algorithmes pour placer leurs pages dans des endroits privilégiés, aucun moteur de recherche publie les valeurs quelles prennent en compte. Pour connaître les éléments décisifs les Webmasters tracent donc les placements et échangent leurs observations.

Les moteurs de recherche, de l'autre côte, observent les agissements des référenceurs pour modifier leurs algorithmes sans cesse et pour limiter de cette manière la triche. Les algorithmes utilisés sont exclusivement connus par les ingénieurs des moteurs de recherche.

Plus d'informations sur les algorithmes se trouvent dans la page [Les algorithmes des moteurs de recherche](#)

### [allintitle](#) (15/03/2008 )

Les titres sont depuis toujours des éléments majeurs dans l'optimisation des pages Web pour le référencement. Il est donc nécessaire de choisir de titres qui définissent le plus exactement le contenu de la page et qui contiennent le ou les mots clés avec lesquels les internautes cherchent la page en question.

Pour connaître les titres que d'autres sites et surtout les sites concurrentiels utilisent, Google permet la recherche des titres. Nous apprenons par ce moyen en même temps les mots clés qui se trouvent dans ces titres. Il suffit d'employer la commande *allintitle:mot-clé* dans la fenêtre de recherche.

La recherche *allintitle:mot-clé* permet de savoir combien des sites utilisent déjà les mêmes titres et les mêmes mots-clés dans leurs titres. Sans cette connaissance personne n'est capable d'optimiser ses pages, et surtout ses titres, pour les moteurs de recherche.

### [allinurl](#) (12/03/2008 )

Le nom du domaine est un élément majeur dans l'optimisation des pages Web pour le référencement. Il est donc nécessaire de choisir un nom de domaine qui contient le ou les mots clés qui définissent le contenu du site.

Pour connaître les variantes des noms des domaines encore libres et en même temps les noms de domaines existants qui utilisent les mêmes mots clés que nous ciblons, Google permet la recherche des noms de domaines qui utilisent certains mots clés. Il suffit d'employer la commande *allinurl:mot-clé* dans la fenêtre de recherche.

La recherche *allinurl:mot-clé* permet de savoir combien des sites utilisent déjà les mêmes mots-clés et de connaître les entreprises qui sont les concurrents directs. Sans cette connaissance personne n'est capable d'optimiser ses pages pour les moteurs de recherche.

### [alt \(attribut\)](#) (06/12/2007 )

L'utilisation de l'attribut *alt* est aujourd'hui obligatoire pour des raisons multiples. La valeur de l'attribut s'affiche au cas où l'image ne s'affiche pas, ce même texte est utilisé par les *voice browsers* et les validateurs exigent également l'emploi de l'attribut *alt* (texte alternatif).

Il s'y ajoute encore le fait que les règles de l'accessibilité demande l'utilisation de l'attribut *alt* et que les robots des moteurs de recherche utilisent sa valeur pour le référencement. Il ne se pose donc pas la question d'omettre cet élément lors de l'optimisation de ses pages.

Il se pose plutôt la question quel texte s'y trouve. Il est hors question de répéter seulement les mots clés ou d'ajouter des phrases à chaque image qui ne représente rien d'autre qu'un point minuscule. Les robots se rendraient compte et l'effort de l'optimisation sera anéanti. Il est donc nécessaire d'utiliser les valeurs de l'attribut *alt* intelligemment.

### [alt texte](#) (20/02/2008 )

Chaque balise <img> dispose de l'attribut alt. La valeur de cet attribut s'appelle alt texte. Il s'agit donc de l'information qui décrit le contenu de l'image en question.

Le alt texte sert principalement à deux buts parallèles. Il s'affiche chez les internautes qui ont bloqué l'affichage des images pour les inciter à ouvrir, alors à regarder l'image que le Webmaster a placée dans la page Web. De l'autre côté le alt texte donne aux robots les informations nécessaires pour pouvoir mieux référencer la page. Vu qu'un robot ne peut pas voir les images, il doit se contenter du alt texte livré par le Webmaster.

Pour optimiser une page Web il est donc nécessaire d'utiliser le alt texte correctement et de ne jamais oublier de l'utiliser. Il est toutefois interdit d'utiliser les mêmes alt textes plusieurs fois à l'intérieur d'un site Web.

### [annuaire](#) (04/12/2007 )

Contrairement aux moteurs de recherche qui affichent un résultat sur la base des mots clé, les annuaires proposent aux Internaute une recherche selon catégorie et sous-catégorie. La recherche est donc arborescente.

Les annuaires utilisent en règle générale pas de robots pour fouiller l'Internet mais exigent la soumission par les Webmasters dans la catégorie qui correspond le mieux au contenu des pages.

Il y a deux groupes d'annuaires. Le premier groupe extrait le contenu des balises meta et utilise les valeurs pour leur base de données. Ces annuaires ont donc automatisés l'insertion des sites. Le deuxième groupe utilise de rédacteurs qui visitent le site et rédigent ensuite un propre texte, parfois en se basant sur les metas. Ce deuxième groupe d'annuaires a souvent une très bonne réputation et refuse donc chaque site qui n'est pas conforme aux exigences du webdesign, du contenu ou de l'accessibilité.

Il y a de nombreux annuaires qui demandent un lien de retour avant d'accepter un site ce qui, logiquement, diminue leur crédibilité. Il y en a même quelques annuaires qui exigent un lien de retour sans toutefois offrir un lien en dur aux inscrits.

### [arborescence](#) (04/02/2008 )

L'arborescence est l'organisation hiérarchique des pages Web et d'autres éléments inclus dans ces pages dont la base est la page d'accueil, donc la page avec l'information principale. Le développement du site est donc construit comme un arbre.

La page d'accueil pointe vers les pages secondaires ou de sections d'un site des lesquelles, par exemple, les informations pointues seront accessibles. L'arborescence est donc la base de la structuration de chaque site Web et doit exister avant que la première page soit programmée. Même les pages qui seront ajoutés ultérieurement doivent déjà faire partie du système de l'arborescence.

Il ne faut toutefois pas oublier que les robots ne peuvent pas suivre plus que trois

liens consécutives. L'arborescence, la structuration et le système de navigation d'un site rencontrent donc assez vite ses limites sans préparation détaillée du site.

[ASCII](#) (25/04/2009 )

Comme un serveur et un navigateur ne parlent, en principe, que l'américain, une page Web qui utilise les caractères français pose de problèmes d'affichage dans de nombreux pays.

Cela concerne aussi les mailing-lists, listes de distributions en français, programmés dans un langage dynamique vu qu'aussi les serveurs mail n'utilisent que l'américain comma langue standard.

Pour remédier à ces problèmes d'affichage chaque page Web a besoin d'une information quel set de caractères la page utilise. Un Webmaster utilise dans ces cas soit la norme ISO 8859-1 ou, alternativement la norme UTF-16.

### [bac à sable](#) (19/12/2007 )

Dans plusieurs moteurs de recherche, dont Google, un nouveau site n'est en règle générale jamais affiché dans les premières pages de recherche pendant les premiers mois de son existence. Cet effet s'appelle bac à sable (*sandbox* en Anglais).

L'algorithme de Google met donc les nouveaux sites en épreuve avant de leur attribuer une place dans les premières pages. Il s'agit, selon contenu et technique employé jusqu'aux quelques mois. Le temps s'augmente si un nouveau site possède dès le début un nombre "exagéré" de liens de retour (*backlinks*) ou si les robots pensent que le site soit suroptimisé.

Les autres critères qui décident sur l'effet de bac à sable peuvent être de liens de nature commerciale, une limite de sites à référencer dans un espace défini ou un classement qui favorise de annonces payés.

### [banner blindness](#) (08/02/2008 )

Pour mettre en valeur certaines publicités/annonces ou pour inciter les internautes de visiter certaines pages du propre site les Webmasters ont généralement utilisé de *banners*. Le résultat était ce qu'on appelle *banner blindness*, l'effet que la majorité des Internautes ne "voit" plus les *banners* et les considère donc comme "non existants".

Le deuxième pas est maintenant que les Webmaster utilisent du texte, mis en valeur par les méthodes du webdesign et de l'optimisation pour insérer de la publicité dans leurs pages. Le risque est toutefois que les internautes développent de cette manière également une *text blindness* ce qui aurait comme suite que même de textes importantes seront moins bien lu qu'aujourd'hui et le contenu d'une page perdra du valeur. On peut donc parler d'un conflit entre commerce électronique et site d'information, donc de contenu.

L'utilisation de la publicité en forme de texte, surtout si ces textes utilisent l'emphase ou d'autres techniques de l'optimisation augmentera aussi l'effet que les moteurs de recherche indexent les pages avec de la publicité plutôt selon les mots clé de la publicité et falsifient donc souvent le résultat que les internautes espèrent. L'utilisation de la publicité en forme de texte doit donc être complètement intégré dans l'optimisation des pages Web.

### [bells-and-whistles](#) (06/03/2008 )

On dit qu'un site web a trop de *bells-and-whistles* si le Webmaster utilise de l'animation (flash, gif animé etc.) qui n'est pas nécessaire pour la compréhension de la page et ralentit donc seulement l'affichage.

*Bells-and-whistles* attire les yeux et empêche donc la lecture correcte du texte. Elles essaient souvent à cacher un contenu pauvre et sont donc, si l'animation ne fournit pas une information supplémentaire, absolument à éviter.

Pour l'optimisation des pages Web les *bells-and-whistles* signifie plutôt de perdre du *PageRank* et, en cas d'exagération de ne pas être accepté par les moteurs de

recherche, même si cette méthode en soit ne soit pas considérée comme *Black Hat SEO*.

[blog](#) (02/04/2008 )

Le mot *blog* provient de l'expression web log et était à l'origine un journal personnel ou un carnet de voyage qui était mis à jour quotidiennement pour faire partager ses expériences et pensées avec ses amis.

Comme les *bloggers*, donc les internautes qui entretenaient un *blog* ne connaissent rien de la programmation, les divers plateformes proposaient de logiciels qui permettaient de créer et d'entretenir un *blog* sans aucune connaissance de programmation.

Même qu'il y encore de nombreux *bloggers* qui poursuivent cette idée, les référenceurs ont découvert leur utilité pour l'optimisation des pages Web et les emploient donc couramment pour promouvoir "leurs" sites. L'apparition des moteurs de recherche comme *technorati* (international) ou blog populaire (français) favorisent cette méthode de promotion de sites par les professionnelles.

[bloptimization](#) (22/03/2008 )

Les *blogs* jouent chaque jour un rôle plus important dans l'optimisation des pages Web pour le référencement. Les *blogs* fournissent un nombre important de liens de qualité pour acquérir un *PageRank* et de nombreux *blogs* occupent déjà les premières pages des résultats.

Mais vu que les *blogs* ne suivent pas les mêmes règles qu'un site standard et ne se basent pas sur une structuration logique, il est nécessaire de les adapter aux besoins du référencement. Cette optimisation des *blogs* s'appelle *bloptimization*.

Un *blog* qui se compose d'une seule page qui change quotidiennement a besoin d'un autre traitement qu'un *blog* qui se compose des nombreuses pages. Et un *blog* avec ne nombreux commentaires n'a pas les mêmes exigences qu'un *blog* qui fournit exclusivement des informations.

[bombardement Google](#) (06/01/2008 )

Le bombardement Google (*Google bombing* en Anglais) utilise la faille de l'algorithme qui détermine le *PageRank* d'une page. Vu que Google utilise surtout le texte des liens de retour pour fixer le *PageRank*, il est assez facile de l'influencer et de le falsifier.

Si donc, par exemple, plusieurs sites ou seulement certains blogs utilisent la même expression pour les liens pointant vers une certaine page, le robot de Google "pense" que cette page ciblée aurait une importance supérieur aux autres pages de, peut-être meilleur qualité, et chaque personne qui cherche cette expression en question aura comme premier résultat la page ciblée par ces liens.

Google essaie toutefois de pister cette tricherie et de modifier l'algorithme de la manière que de tels pages ciblées seront écarté des résultats de recherche.

L'emploi exagéré du bombardement Google peut donc aussi avoir le résultat que le site ciblé atterrit sur la liste noir des moteurs de recherche.

[buzz](#) (02/05/2009 )

Le *buzz*, aussi appelé bourdonnement en français, est une stratégie du web marketing qui se base sur la technique à faire du bruit autour d'un produit, d'un service ou d'un site encore avant sa sortie ou publication officielle.

Le *buzz* joue donc avec la curiosité des internautes et crée un besoin déjà avant sa parution. Pour créer du *buzz* on emploie surtout les listes de distribution, les forums et les blogs.

Créer du *buzz* doit être bien réfléchi vu qu'un produit ou un service connu par un rumeur doit, ensuite, remplir les promesses et tenir la route. Un visiteur non content peut facilement créer du *buzz* inverse par les mêmes réseaux et créer une mauvaise image, même sans fondement.

Plus d'informations sur le buzz se trouvent par le lien suivant :

[Le buzz et le référencement](#)

### [cadre](#) (29/12/2007 )

Les cadres, plus connu sous le terme anglais *frame*, permettent de diviser une page Web en plusieurs parties et d'actualiser, selon la technique employé, une seule fenêtre ou plusieurs fenêtres à la fois sans que le reste de la page soit modifié.

Même si cette technique a certains avantages, l'emploi de cadres est aujourd'hui, surtout à cause de problèmes du référencement, déconseillé. En règle générale les cadres bloquent la navigation des robots et empêchent donc de cette manière l'indexation du site. Mais un site qui n'existe pas pour les moteurs de recherche n'existe pas pour les internautes non plus.

Le système de *frames* a été remplacé par les tableaux et les pages dynamiques qui, aujourd'hui ont été remplacés par les feuilles de style et les pages en *xhtml* ou dynamiques ce qui permet un référencement optimisé.

### [click tracking \(tracage de cliques\)](#) (16/12/2007 )

Le *click tracking* est assez discuté dans le domaine de l'optimisation des pages Web. D'un côté, même de nombreux moteurs de recherche l'utilisent, d'un autre côté, les résultats peuvent interférer avec la privacité des internautes.

Le *click tracking* permet d'analyser le comportement des visiteurs et fournir des informations assez détaillées : Combien de temps un internaute reste sur une page, le nombre de cliques effectués, quels liens ont été utilisés etc. Ces informations sont intégrées dans les algorithmes des moteurs de recherche pour, par exemple, définir le PageRank de Google ou pour détecter un mauvais voisinage.

Mais le *click tracking* est également utilisé pour calculer les cliques effectuées sur certains banner et d'analyser le comportement des internautes, par exemple, pour calculer les cliques vers les partenaires ou d'afficher de messages intempestifs.

### [closed loop](#) (28/03/2008 )

On parle d'un *closed loop*, d'un circuit fermé, si un group de sites utilise les liens exclusivement pour lier ses propres pages à l'intérieur de la communauté. Le but est d'obtenir un meilleur emplacement dans les moteurs de recherche et d'obtenir un meilleur *PageRank*.

Vu qu'un *closed loop* évite d'ajouter de liens qui ciblent de sites hors de la communauté, cette méthode fait partie de la *Black Hat SEO* et est puni par les moteurs de recherche.

Actuellement personne ne sait combien des liens extérieurs un *closed loop* nécessite pour que les robots les acceptent dans leurs bases de données. Il est donc conseillé d'éviter de se joindre à des *closed loops*.

### [clustering](#) (04/01/2008 )

Pour éviter qu'un seul site occupe la ou les premières pages des résultats, les moteurs de recherche utilisent la technique du *clustering*.

Si un site possède donc de nombreuses pages optimisées pour les mêmes mots clés, les moteurs affichent seulement une ou deux pages du même site. L'effet d'éliminer les autres pages des résultats s'appelle *clustering*. L'internaute a toutefois souvent la possibilité d'utiliser un lien qui lui permet d'afficher aussi les autres pages de ce site.

Il n'est donc pas trop astucieux d'optimiser plusieurs pages en ciblant les mêmes mots clés mais de choisir d'autres mots clés dominants pour chaque page Web du même site.

#### [commentaires](#) (24/01/2008 )

Pendant quelques années les référenceurs ont utilisé les balises de commentaires, par exemple `<!-- le commentaire -->` pour ajouter des mots clés ou phrases clés dans leur pages et ainsi d'améliorer leur positionnement.

Les algorithmes d'aujourd'hui reconnaissent les balises de commentaires et considèrent cette méthode en tant que *keyword stuffing* ce qui augmente le risque de se retrouver en bas de l'échelle ou même d'être bannie d'un moteur de recherche au lieu d'obtenir le résultat espéré. Il est donc exclu d'utiliser les commentaires pour l'optimisation des pages Web.

Vu qu'un commentaire mal formaté détruit la mise en page et risque toujours d'attirer l'attention des robots, il vaut mieux de ne pas les utiliser s'il n'y a pas une nécessité absolue de le faire.

#### [contenu](#) (02/12/2007 )

Le contenu de la page est le texte que les internautes voient sur l'écran sans que la mise en page ou le design soit pris en compte. Le contenu est donc la partie qui, la plupart de temps, est rédigé à l'aide d'un traitement de texte.

Les moteurs de recherche ainsi que la majorité des annuaires demandent aujourd'hui un contenu minimum par pages. Il ne sert donc plus à rien d'écrire quelques lignes de texte et d'espérer que cette page soit mise en avant dans les moteurs de recherche. Pour qu'une page soit acceptée par les moteurs il faut compter avec un strict minimum de 300 mots par page. Une page optimisée comporte donc entre 600 et 1000 mots.

Vu qu'un site a aussi besoin d'être indexé par les annuaires il est logique que ces textes ne doivent pas comporter trop de fautes d'orthographe ou de grammaire et que leur style doit rester compréhensible pour la majorité des internautes.

Une page trop longue, qui dépasse donc 5000 mots, n'est pas acceptée par les robots des moteurs de recherche non plus, vu que de textes longues ne sont plus lus par la majorité des internautes, même si le contenu soit d'un haut niveau.

#### [contenu dynamique](#) (01/04/2008 )

Contrairement à un site statique où chaque page est créée en avance par le Webmaster et comporte donc toujours le même contenu et une navigation identique,

Le contenu dynamique utilise des pages qui se créent au fur et à mesure selon la navigation et la demande de l'internaute.

Le contenu dynamique n'utilise pas de pages avec l'extension .html mais se reconnaît par les extensions .php, .asp, cgi, .cfm etc. Les pages elles-mêmes se composent d'un code qui "cherche" le contenu selon besoin de la base de donnée auquel il est lié.

Comme les pages dynamiques ne suivent pas le même schème comme les pages statiques, l'optimisation de ses pages doit être également adapté à cette technique et pose donc des problèmes très spécifiques. Seulement une structuration de base adaptée au contenu dynamique permet l'optimisation de ses pages.

### [crawling ciblé](#) (31/01/2008 )

Comme il est impossible que les robots des moteurs de recherche visitent chaque jour chaque site, ils utilisent un certain rythme de base. Google, par exemple, a besoin d'un mois pour visiter les sites indexés dans sa base de données et utilise donc ce rythme pour un site récent.

Mais il y a des sites qui sont quotidiennement mis à jour (les sites de la presse) et d'autres qui sont quasiment statique ou ne sont actualisés qu'une fois par année. C'est pourquoi la majorité des moteurs de recherche utilise le *crawling* ciblé (*focused crawling* en Anglais). Les moteurs visitent donc les sites actives plus souvent que les sites inactives et utilisent pour chaque site indexé une propre formule.

L'algorithme du *crawling* ciblé ne sert toutefois pas seulement à définir le cycle de visite des moteurs mais influence aussi le positionnement des sites. Seulement un site qui est régulièrement mis à jour et qui fournit chaque semaine des informations supplémentaires a donc la chance d'atterrir ou de rester sur les premières pages des résultats.

### [deep crawl](#) (16/02/2008 )

Selon l'actualisation du site, l'utilisation du Google sitemap, le nombre des pages ajoutées par jour et quelques autres facteurs, Google visite un site parfois chaque jour pour ajouter les nouvelles pages, d'autres fois le robot passe seulement une fois par an.

Mais ces visites à part, GoogleBot, le robot du *deep crawl* de Google fait chaque mois une fois un *deep crawl*, c'est-à-dire que ce robot ne se limite pas à regarder les pages ajoutées mais qu'il suit tous les liens jusqu'au troisième niveau à l'intérieur du site pour, de cette manière actualiser sa base de données, donc d'ajouter des pages mais aussi pour effacer d'autres pages ou sites.

Mais le *deep crawl* n'a rien à voir avec l'attribution du *PageRank* de Google. Le *PageRank* est actualisé dans un autre rythme vu qu'il ne dépend pas de la qualité d'un site et même pas des nombres de visites uniques ou de la taille d'un site.

### [deep submitting](#) (01/02/2008 )

On parle d'un *deep submitting* si un Webmaster ne soumet pas seulement sa page d'accueil mais chaque page individuellement dans un moteur de recherche en espérant d'augmenter ainsi sa chance d'être indexé avec toute page du site.

Les moteurs de recherche n'aiment pas cette agissement vu qu'un site doit, selon leur algorithme, être construit de la manière que chaque page soit à trouver à partir de la page d'accueil. Ils n'aiment pas non plus de devoir fouiller un site à partir, par exemple 200 pages, si une seule suffit pour trouver les autres. Chaque essai de faire travailler les robots plus que nécessaire est donc puni d'une manière ou d'autre.

Le *deep submitting* a donc souvent le résultat que les moteurs de recherche déclassent un site soumis par cette manière. La méthode est donc déconseillée en optimisation et ne rapporte certainement pas le résultat espéré.

### [détecteur d'intrusion](#) (27/04/2009 )

Un détecteur d'intrusion désigne l'ensemble d'un système logiciel et matériel qui permet à détecter une intrusion le même moment que l'essai d'intrusion a lieu.

Le but du détecteur d'intrusion est à neutraliser l'attaque sur l'ordinateur, le réseau ou le serveur immédiatement. Actuellement on emploie deux méthodes de détection en parallèle.

Les SDI disposent d'une base de données de signatures connues et surveillent donc en permanence l'attaque par une de ces signatures. La détection d'intrusion surveille en même temps chaque anomalie dans le système d'exploitation, donc le changement de mémoire, l'utilisation excessive du CPU etc. pour pouvoir bloquer chaque essai d'intrusion en temps réel.

### [document expert](#) (23/03/2008 )

Les moteurs de recherche considèrent un document qui parle d'un certain sujet et qui propose exclusivement de liens vers d'autres documents qui offrent des informations

supplémentaires sur le sujet en question comme document expert (*expert document* en Anglais).

La base est donc que le document expert fournit des informations profondes sur un certain sujet et ne pas seulement quelques phrases et qu'un rédacteur choisisse les documents à lier qui également fournissent des informations approfondies sur le même sujet.

Il y a des moteurs de recherche qui prennent exclusivement les liens des documents experts en considération pour déterminer la popularité des liens d'un site. L'optimisation des pages Web prend cet effet en considération lors de la rédaction du contenu.

#### [domaine satellite](#) (23/12/2007 )

Vu que les moteurs de recherche favorisent les noms de domaines qui ont une relation avec le contenu, de nombreux anciens sites, même déjà bien établis, sont donc d'une certaine manière pénalisés.

Il était donc logique qu'un nombre important d'entreprises s'achetaient de noms de domaines supplémentaires qui correspondaient mieux à leurs activités et aux contenus des pages. Ces domaines n'étaient jamais destinés à contenir du contenu mais étaient seulement destinés à renvoyer les visiteurs vers le site d'origine. Ces domaines sont appelés domaines satellites.

Les robots découvrent toutefois les sites et l'emploi excessif de domaines satellites est donc sanctionné par la majorité des moteurs de recherche.

#### [domaine virtuel](#) (10/01/2008 )

Généralement un serveur dispose d'une à dix adresses IP qui sont partagées parmi tous les noms de domaines installés dans le même serveur. Les noms de domaines, donc de sites, ne disposent donc généralement pas d'un propre IP. On dit alors que ces sites disposent d'un domaine virtuel.

Le nombre de domaines virtuels installés dans un serveur est décisif pour la vitesse de transfert des pages entre serveur et navigateur et la qualité des sites qui partagent le même IP joue parfois un rôle pour le référencement des pages, surtout, si on partage l'IP avec un spammeur ou si on est installé dans un serveur qui dispose de failles de sécurité.

Il est donc important pour chaque Webmaster de bien choisir son hébergeur et de se renseigner sur la quantité de domaines virtuels installés sur les serveurs pour ne pas avoir des surprises négatives lors du référencement.

#### [duplication de contenu](#) (18/12/2007 )

Les moteurs de recherche ont de plus en plus la tendance à demander du contenu et placent donc un site de 300 pages, dont chaque page contient au moins 600 mots beaucoup plus haut qu'un site de 30 pages avec 300 mots.

Le résultat était évidemment que des Webmasters peu scrupuleux ont commencé à

copier du contenu des autres sites au lieu d'écrire leurs propres textes. Il y en a qui utilisent de textes protégés par la loi, d'autres copient les informations libres de droits en "oubliant" d'indiquer la source. Copier d'autres pages est considéré comme malhonnête par les moteurs de recherche et provoque des conséquences.

Dès qu'un robot détecte une page de contenu double, ils effacent la page copiée simplement de leur base de données et affichent les autres pages du site plus bas dans le résultat de recherche.

### [emphase](#) (07/02/2008 )

La balise `<em>` est utilisée pour mettre en valeur certains mots ou certaines parties d'un texte. L'emphase (*emphasis* en Anglais) joue donc aussi un rôle dans l'optimisation des pages Web.

Il est donc important de bien et de correctement utiliser les emphases vu que les moteurs de recherche n'aiment pas les exagérations et, il ne faut pas oublier qu'il y a des annuaires n'acceptent pas de pages dont la mise en page souffre, ce qui est évidemment le cas si 30% d'un texte soit entouré des balises `<em>` et `</em>`.

Optimiser une page signifie donc de trouver un équilibre entre le texte visible par les internautes et annuaires et le code visible par les robots. Il faut évidemment informer les robots des mots ou phrases importants mais sans toutefois détruire l'aspect visible.

### [ergonomie du site](#) (11/01/2008 )

L'ergonomie d'un site désigne l'adaptation d'un site aux besoins des internautes. Il s'agit donc de lier l'efficacité de la navigation et de l'utilisation d'un site et le confort des utilisateurs.

L'ergonomie ne s'arrête toutefois pas à bien situer, par exemple, la navigation mais inclut aussi de permettre aux Internaute d'utiliser le clavier au lieu de la souris, d'utiliser la tabulation pour se situer dans un formulaire et de fournir des informations supplémentaires en utilisant les attributs *alt* et *title*.

L'ergonomie joue aujourd'hui également un rôle de plus en plus important lors du référencement et les algorithmes utilisés par les moteurs de recherche sont capable d'évaluer plus que 100 éléments qui traitent seulement l'ergonomie d'un site.

### [erreur 404](#) (11/12/2007 )

Dès qu'un internaute cherche une page qui n'existe pas sous l'adresse indiquée, le serveur répond avec un message "erreur 404". L'internaute ne voit pas forcément ce message d'erreur mais, par exemple, un texte qui l'informe seulement du fait que la page n'existe pas.

Pour les robots il s'agit, à cause, déjà à cause de l'emplacement de ce fichier, toujours de l'erreur 404 ce qui signifie pour les moteurs de recherche que le Webmaster en question n'a pas travaillé correctement. Si alors un seul lien d'un site ne fonctionne pas, tout le site est considéré comme "en travail" en n'atterrit donc pas sur les premières pages de recherche.

La même erreur est produite par un serveur si le Webmaster n'utilise pas les fichiers robots.txt et sitemap.xml. Il arrive donc souvent qu'un site démarre déjà avec deux erreurs avant que les robots trouvent la page d'accueil.

### [fautes d'orthographe](#) (07/05/2009 )

Les pages qui contiennent un nombre important de fautes d'orthographe pose généralement un problème de référencement vu que les algorithmes des moteurs de recherche passent les pages par un vérificateur d'orthographe.

Mais les fautes d'orthographe empêchent aussi les moteurs de recherche à trouver une page dont la requête se base sur l'orthographe correcte. Il est donc important à publier exclusivement de pages Web dont l'orthographe soit conforme avec les dictionnaires.

Un référenceur sait toutefois aussi quelles fautes d'orthographe comptent parmi les plus courantes et utilise donc le mot clé en question une seule fois avec une faute d'orthographe dans la page Web et augmente ainsi la chance que les internautes la trouvent.

### [ferme de liens](#) (10/12/2007 )

Pour atterrir sur une des premières pages d'un moteur de recherche, un site a besoin de liens de retour (*backlinks*) ce qui peut être assez difficile pour un nouveau site qui ne dispose pas encore de *PageRank*.

C'est pourquoi les fermes de liens (*link farms* en Anglais) ont fait leur apparition. Il s'agit en fait de sites qui insèrent automatiquement les liens qu'un Webmaster leur soumet. Ces pages se composent en règle générale exclusivement de liens sans aucune système et sans que les pages aillent le moindre rapport avec les sites ajoutés.

Les robots reconnaissent entre temps ces sites et ont adaptés leur algorithme à ces pseudo-liens de retour. Un tel lien ne rapporte donc plus rien aux Webmasters qui risquent toutefois que ses pages atterrissent sur la liste noire des moteurs de recherche.

### [fenêtre de politesse](#) (27/02/2008 )

La majorité des robots qui visitent un site ne scrutent pas le site entier lors de cette visite mais seulement quelques pages pour revenir un ou deux jours plus tard pour scruter d'autres pages. L'espace entre ces deux visites des robots s'appelle fenêtre de politesse (*politeness window* en Anglais).

Les moteurs de recherche utilisent cette technique pour que les serveurs ne soient pas surchargés par chaque visite. Il est évident que cette méthode ne s'applique pas pour de sites d'une dizaine des pages mais seulement pour des "vrai" sites selon les critères des moteurs de recherche qui dépassent alors une centaine des pages.

L'emploi des fenêtres de politesse est aussi la raison pourquoi les moteurs n'affichent pas chaque résultat après avoir visité un site et avoir chargé le fichier *sitemat.xml*.

### [fenêtre pop-under](#) (28/02/2008 )

Fenêtres *pop-under* sont la méthode mieux cachée des *pop-up*. La fenêtre qui contient souvent de la publicité n'est pas immédiatement visible par l'internaute comme c'est le cas avec les *pop-up*, mais elle se cache en dessous des fenêtres déjà ouvertes et apparaît donc dès que l'internaute ferme les autres fenêtres.

Même si les fenêtres *pop-under* ne dérangent pas lors de la navigation, leur emploi n'est pas apprécié par les internautes qui se sentent encore plus ennuyé en découvrant des fenêtres cachées.

Vu que l'utilisation des *pop-under* fait également partie du *Black Hat SEO*, les sites qui les emploient risquent de se voir bannis par la majorité des moteurs de recherche.

### [FFA](#) (21/02/2008 )

Les trois caractères signifient *Free For All*, gratuit pour tout le monde, et se réfère aux pages similaires aux annuaires qui proposent de listings de sites, soit catégorisé soit en vrac.

La différence entre un annuaire et un listing *FFA* est que les annuaires sélectionnent les pages qu'ils insèrent, qu'il y a donc un traitement rédactionnel des données et les *FFA* acceptent les liens de tout le monde qui veut s'inscrire. Les *FFA* ressemble donc plutôt aux fermes des liens et comportent les mêmes dangers. La majorité des *FFA* publient exclusivement de liens sans aucun autre contenu.

Il faut donc vérifier la valeur du site *FFA* en question avant d'y ajouter son *URL* vu qu'il y a très peu des *FFA* qui peuvent servir lors de l'optimisation des pages Web pour les moteurs de recherche.

### [fréquence de visite](#) (12/02/2008 )

La fréquence de visite (*Crawl Frequency* en Anglais) indique dans quel rythme un site soit visité par le robot du même moteur de recherche mais ne dit rien sur l'indexation des pages Web.

La fréquence de visites dépend de nombreux facteurs. Un site dont s'ajoutent chaque jour des pages supplémentaires, sous condition qu'il s'agit d'un site de confiance, pourrait être visité chaque jour ou, du moins chaque semaine. Un site qui comporte surtout des liens, peut-être même de moindre qualité, ou qui fait partie des programmes d'affiliation ne sera pas visité plus qu'une fois par année, même si le contenu change.

On peut donc partir du principe qu'un site de haute qualité (sans liens qui mènent vers un mauvais voisinage), qui est régulièrement actualisé, où le contenu augmente souvent et où chaque page propose environ 500 mots de contenu unique sera visité assez souvent, ce qui peut être un facteur important de l'optimisation des pages Web pour le référencement.

[FreshBot](#) (03/03/2008 )

*FreshBot* et le nom du robot que *Google* utilise pour les sites dont le contenu change souvent, c'est-à-dire plusieurs fois par semaine. Il est sans importance si les pages soient actualisées (ajouté de contenu) ou qu'il y a des pages entières qui s'ajoutent.

Un référenceur ou un Webmaster ne peut pas demander *Google* à utiliser *FreshBot*, c'est seulement l'algorithme que *Google* emploi qui définit le rythme des visites d'un site. La meilleur méthode reste donc de travailler activement sur le contenu. Il ne suffit toutefois pas de modifier chaque jour que trois mots par page !

La visite du *FreshBot* ne signifie toutefois pas que les pages soient toujours indexées le lendemain de la visite. L'algorithme utilise aussi ici plusieurs paramètres qui décident quelles pages *Google* ajoute à sa base de données et à quel moment. Le rythme standard des visites de *GoogleBot* est environ un mois.

### [géolocalisation](#) (24/12/2007 )

La géolocalisation désigne l'effet qu'un robot utilise pour situer un serveur d'après l'IP qui est envoyé ensemble avec la page Web pour l'indexer dans le pays correspondant.

Comme la géolocalisation est encore à son stade expérimentale, l'algorithme des moteurs de recherche ne leur donnent pas encore trop d'importance et se fient plus à l'extension du domaine, le nom du domaine et le nom des dossiers utilisées. Mais la géolocalisation prend de plus en plus de place dans les algorithmes et les Webmasters ne doivent donc pas négliger son importance.

Un internaute en France qui cherche donc un site d'après certains mots clés trouve donc, selon ce principe, en premier lieu les sites hébergés en France et en haut de la liste les sites qui, en plus, disposent de l'extension .fr.

### [Google bowling](#) (18/01/2008 )

On parle de *Google bowling* si un Webmaster utilise de moyens déloyaux pour faire baisser le *PageRank* d'un site concurrent et essaie de le faire descendre dans les résultats de recherche.

Le *Google bowling* montre parfois assez vite ses résultats, si, par exemple, quelqu'un place de liens de retour (vers le site "à descendre") dans un mauvais voisinage. Google ne peut pas contrôler qui a placé ces liens et réagit donc de la même manière si quelqu'un aurait placé lui-même ses liens de retour dans des sites bannis. Dans le pire des cas, le site honnête pourrait donc également atterrir sur la liste noire des moteurs de recherche.

Même s'il s'agit d'un acte malhonnête, déloyale et criminel, l'optimisation des pages Web montre parfois sa face cachée pour faire descendre un concurrent.

### [googlebot](#) (22/01/2008 )

*Googlebot* est le nom du robot que Google utilise pour retrouver les pages Web et pour construire ensuite un index que le moteur de recherche de Google puisse utiliser pour afficher les résultats de recherche dans un ordre logique et selon l'importance du contenu.

Comme *googlebot* se base exclusivement sur un algorithme il n'est pas capable de "comprendre" les idées d'un Webmaster. Pour que les pages soient correctement indexées le site doit donc se baser sur une structuration logique qui n'utilise jamais plus que trois niveaux. Il est également impératif que les bons mots clés soient utilisés, qu'il y a du contenu, que le code soit valide et que chaque page soit optimisée pour nommer que quelques éléments de l'optimisation.

Sans que les règles de l'optimisation soient appliquées il y a risque que les pages ne soient pas indexées ou qu'elles soient introuvables par les internautes.

### [Gray Hat SEO](#) (10/05/2009 )

La Gray Hat SEO désigne chaque méthode de l'optimisation qu'un référenceur utilise en espérant que l'algorithme des moteurs ne découvre pas la Black Hat SEO qu'il, en réalité, utilise.

Les limites de la Gray Hat SEO sont assez floues. Elle se base toujours sur de méthodes que les moteurs de recherche, en principe, pénalisent même si son emploi ne mène pas forcément au bannissement.

La Gray Hat SEO commence déjà avec l'emploi de quelques mots clés de trop dans un texte et continue avec la création de quelques forums et blogs en relation avec le contenu du site qui ajoutent de liens de retour.

Le problème de la Gray Hat SEO est justement que les moteurs de recherche ne bannissent pas le site mais, dès qu'ils ont de soupçons, ne l'indexent pas dans les premières pages des résultats.

### [guillemets](#) (29/03/2008 )

Il y a une différence entre le fait de chercher simplement des mots ou de chercher des mots entourés de guillemets. La recherche "optimisation des sites" fournit un autre résultat que la recherche des mots 'optimisation' et 'sites'.

La majorité des moteurs de recherche prennent les guillemets comme délimiteurs et fournissent donc seulement les résultats où l'expression correcte se trouve dans les pages Web cherchées et ne pas seulement les mots en question.

Cet effet joue aussi un rôle dans l'optimisation des pages Web pour le référencement vu que les internautes "professionnels" se servent en règle générale des guillemets, contrairement aux internautes "standards". Dans cette partie de l'optimisation l'effet de l'adjacence doit également être pris en considération.

### [GYM](#) (30/03/2008 )

Ces trois lettres représentent les moteurs de recherche Google, Yahoo et MSN, donc ceux que les internautes dans le monde entier consultent le plus fréquemment.

Chaque page Web doit donc être optimisé au mois pour ces trois moteurs. Selon le sujet et le public ciblé ils s'y ajoutent d'autres, par exemple technorati pour les *blogs* ou DMOZ si on parle des annuaires dont un site devrait être référencé.

La présence dans un des moteurs de recherche joue également un rôle lors de l'indexation, le placement et même le *PageRank* dans un autre moteur. Négliger d'optimiser ses pages pour GYM signifie donc quasiment de ne pas optimiser ses pages du tout.

### [header](#) (05/04/2008 )

Le *header* (l'en-tête en Français) est la première partie du code d'une page Web et contient surtout les balises `<TITLE>`, `<META>` et `<LINK>` ainsi que l'information sur la version du HTML/xhtml employé et la langue de base du document.

Les moteurs de recherche considèrent la partie du header comme plus important que la partie du body et surtout la balise `<TITLE>` joue un rôle dominant lors de l'optimisation des pages Web pour le référencement.

Nombreux moteurs de recherche fouillent également les balises `<META>` pour extraire des informations sur le contenu de la page ou même pour comparer les mots utilisés dans la description et les mots clé suggéré par le Webmaster avec le contenu effectif de la page. Le *header* doit donc être rédigé avec un soin extrême.

### [hit](#) (20/01/2008 )

Chaque Webmaster doit surveiller quelles pages du site attirent le plus de visiteurs, combien d'Internauts visitent le site etc. aussi pour, éventuellement, optimiser certaines pages pour le référencement ou de les adapter aux besoins des utilisateurs.

Les analyses standard sur les serveurs livrent diverses statistiques, parmi d'autres donnent-ils le nombre de *hits*. Cette indication ne sert en règle générale à rien de tout. Un *hit* est chaque élément qui est chargé du serveur, donc la page, mais également chaque image, même celles qui servent à la mise en page et qui sont placées dans les feuilles de style. Une seule page peut donc produire 20 à 30 *hits*, même s'il ne s'agit ni d'un visiteur unique qui passe ni de quelqu'un qui resté plus que 3 secondes sur la page.

Pour évaluer le succès d'un site, le nombre de *hits* ne dit donc rien. Un seul visiteur qui se balade un peu sur le site peut déjà engendré 200 ou 300 *hits* sans apporter quoi que ce soit au site lui-même.

### [hot linking](#) (04/03/2008 )

On parle du *hot linking* si un Webmaster emploi de liens qui affichent des fichiers qui se trouvent sur d'autres sites (vidéo, son, animations flash etc.) ou pour charger des fichiers ou programmes qui se trouvent sur un autre serveur.

Le *hot linking* utilise donc les capacités et la bande passante d'autres serveurs, ce qui constitue, même si le contenu affiché ou charge soit légal, un abus et fait donc aussi partie du *Black Hat SEO*.

L'emploi du hot linking peut donc être nuisible pour l'optimisation des pages Web pour le référencement et peut mener au bannissement du site. Il y a toutefois quelques rares exceptions ou le *hot linking* puisse être envisageable. Seulement un professionnel du référencement est capable de conseiller un tel pas.

### [HTTP\\_REFERER](#) (09/01/2008 )

Si on parle du *HTTP\_REFERER* on parle souvent seulement du *referrer*, se qui définit la page Web précédente de la page actuellement ouverte.

Le transfert des pages entre serveur *HTTP* et navigateur permet, grâce aux *headers* de chaque page, d'enregistrer l'*URL* de laquelle l'Internaute vient. L'analyse de la variable enregistrée permet, entre autre, de savoir de quel moteur de recherche l'internaute se servait et les mots clés qu'il a utilisé lors de sa recherche.

Le *HTTP\_REFERER* ne sert malheureusement pas seulement aux moteurs de recherche mais est souvent utilisé pour tracer les habitudes des Internautes et d'utiliser les résultats pour des buts divers et parfois malhonnêtes.

### [HTTP refresh](#) (26/04/2009 )

Le *HTTP refresh* peut s'employer de deux manières, mais seulement une manière est à conseiller si on ne veut pas atterrir sur la liste noire des moteurs de recherche.

Pages qui s'actualise ou complètent dans un rythme plus ou moins régulier comme un journal online peuvent utiliser le *HTTP refresh* dans un rythme logique et en relation avec l'actualisation de la page pour pouvoir mieux informer ses visiteurs.

L'utiliser, par contre, le *HTTP refresh* pour diriger les visiteurs vers une autre page Web est considéré comme de la *Black Hat SEO*. Pour rediriger les visiteurs vers une autre page du même site ou d'un autre site le Webmaster doit obligatoirement utiliser le *redirect 301* ou le *redirect 302*.

### [hyperlien](#) (10/03/2008 )

On parle plus souvent simplement du lien que du hyperlien; mais il s'agit de toute manière de la partie réactive d'une page (un texte ou une image) qui mène l'internaute après un clique soit vers une autre page, soit vers un endroit spécifique de la même page Web ou l'ouvre un fichier autre que HTML.

Les hyperliens doivent donc se distinguer du texte qui les entourent et capter les yeux du visiteur pour qu'il les trouve facilement et les utilisent. Di un lien mène vers d'autres fichiers, le Webmaster doit informer le visiteur de ce fait avant qu'il utilise la clique.

Pour l'optimisation des pages Web le Webmaster ou le référenceur utilise le texte entre les balises des liens. L'utilisation des images les oblige donc à créer un deuxième lien textuel pour que les moteurs de recherche sachent de quoi la page ciblée parle. Le choix du texte est primordiale lors de la création des liens pour que les pages soient indexées correctement dans les moteurs de recherche.

### [image map](#) (18/03/2008 )

Les *image maps* sont des image découpée en plusieurs zones cliquables. Une clique sur ces zones ouvre une page de la même manière qu'un lien textuel et sert donc de navigation.

Tant que les *image maps* sont *embedded* dans le code HTML, la majorité des moteurs de recherche est capable de suivre les liens utilisés, sous condition que la règle des trois niveaux soit respectée.

Dès que les *image maps* sont appelées en tant que fichiers externes, la situation change et les robots ne peuvent plus suivre les liens. Vu que les liens sont des éléments primordiaux en optimisation, une telle *image map* empêche donc l'indexation par les moteurs de recherche et doit être évité.

### [indexation](#) (28/04/2009 )

Si on parle de l'indexation en référencement on parle, en règle générale, de deux faits différents.

D'un côté se passe l'indexation par les moteurs de recherche et désigne le fait qu'ils ajoutent une ou plusieurs pages selon des règles bien définies dans leur base de données.

De l'autre côté parle-t-on aussi de l'indexation si un Webmaster ou référenceur soumet ses pages aux moteurs de recherche. Plusieurs moteurs de recherche, dont Google, n'acceptent seulement une soumission manuelle.

### [introduction en Flash](#) (02/02/2008 )

La création d'une page d'introduction en Flash, donc une page d'accueil en Flash, demande beaucoup de travail et exprime souvent aussi la créativité du programmeur. Pour le référencement toutefois une introduction en Flash est une très mauvaise solution.

La majorité des internautes n'aime pas de telles introductions vu qu'elles n'offrent aucune information et, s'il y a du son ajouté, personne ne peut ouvrir une telle page au lieu de son travail. La perte en visiteurs est donc tellement considérable que, déjà économiquement, la solution Flash comme page d'accueil est un désastre, sauf pour proposer une formation en Flash, évidemment.

La majorité des moteurs de recherche ont en plus besoin des informations pour indexer une page correctement, donc du contenu sur la page d'accueil, ce qui n'est pas le cas dans une introduction Flash. Même les moteurs de recherche capables de déchiffres (dans une certaine limite) une introduction en Flash ne vont pas afficher un telle page dans une bonne position dans les résultats.

### [ITO](#) (26/03/2008 )

Les trois lettres représentent les mots *Internet Traffic Optimizer*. Le *ITO* est la personne qui optimise une page Web pas seulement pour les moteurs de recherche mais aussi pour générer du trafic.

Le *ITO* utilise des moyens comme blogs, RSS feeds, la rédaction des articles online etc. pour attirer des Internautees et fournit donc un élément supplémentaire lors de l'optimisation des pages Web.

Aujourd'hui il devient de plus en plus important d'ajouter les éléments du Web 2.0 dans l'optimisation des pages Web et les Webmasters doivent apprendre (en plus) le métier du *ITO* pour ne pas perdre du terrain.

### [keyword \(mots clés\)](#) (01/12/2007 )

Les keywords, leur emplacement et leur quantité jouent un rôle très important dans l'optimisation des pages Web pour le référencement. La majorité des Webmasters essaient de placer trop de mots clés différents et trop de mots clés identiques dans leurs textes.

Il est quasiment impossible d'utiliser plus que trois mots clés par page sans que le contenu ne devienne pas illisible lors de l'optimisation. Il ne faut pas oublier que les keywords doivent exister dans le titre principal, dans le titre du document, plusieurs fois dans le texte et en plus dans les valeurs de certains attributs.

Il s'y ajoute que les rédacteurs des annuaires n'acceptent pas de pages dont le contenu soit peu lisible et suroptimisé. Il faut donc bien choisir les mots clés pour chaque page et de rédiger un texte qui contient ces mots clés dans de contextes différents. Si un Webmaster surcharge ses pages en mots clés il risque d'inverser l'effet et de se voir banni des moteurs et de ne pas pouvoir entrer dans les annuaires non plus.

### [KISS](#) (19/03/2008 )

Le terme KISS représente l'expressions "**Keep It Simple, Stupid**", donc "fait le facile, pour idiots". Il s'agit de la règle principale du webdesign, de l'optimisation ainsi que du web marketing.

Ce règle se base sur le fait qu'un nombre très élevé des internautes n'ont que des connaissances élémentaires de l'Internet et ne savent donc pas reconnaître une navigation évoluée et ne peuvent pas lire les textes plus développés. On ne parle pas des néophytes mais des personnes qui utilisent Internet régulièrement mais qui ont besoin des solutions très simples.

Vu que ces personnes représentent un marché online très élevé il est obligatoire d'appliquer le *KISS* dans tous les niveaux de la création des sites et surtout dans l'optimisation des pages Web pour le référencement. Il ne faut jamais faire "penser" ou faire "chercher" ses visiteurs si on cible un site qui apporte des résultats.

### [lemmatisation](#) (25/12/2007 )

Plusieurs moteurs de recherche, dont Google, utilisent dans leur algorithme la lemmatisation (aussi appelée stemming), c'est-à-dire la recherche selon la famille de mots. Cette technique se base sur les lemmes, la forme canonique des mots.

Les moteurs de recherche sont donc capables de voir une liaison, par exemple, dans les mots promener, promenade, promenades, se promener etc. en les "réduisant" au lemme "promen". Les robots sont donc, dans une certaine limite, aussi capables à détecter un bourrage de mots clés en utilisant l'ensemble de mots d'une famille en nombre trop élevé.

Pour évaluer la qualité des liens de retour, le lemmatisation joue également un rôle en comparant tous les nom, adjectifs et verbes des mots clés dans les deux pages liées.

### [liens absolus](#) (15/02/2008 )

On parle des liens absolus si le Webmaster utilise l'*URL* complet dans ces liens. Il s'agit donc de l'alternative aux liens relatifs. Mais quel est l'avantage ou le désavantage des liens absolus dans l'optimisation des pages Web ?

Côté moteurs de recherche, l'utilisation des liens absolus est plutôt un désavantage vu que les moteurs traduisent aussi les liens relatives en liens absolus mais le temps de chargement des pages augmente, ce qui est un facteur négative pour le référencement ainsi que pour l'accessibilité. Le ou les mots clés du nom de domaine par contre sont de toute manière lus par les robots.

Il y a toutefois une exception. Si un Webmaster utilise de fichiers *PDF*, il est nécessaire d'y introduire toujours les liens absolus vu que les liens dans un PDF ne réagissent pas avec un lien relative ce qui fâche certainement chaque visiteur.

### [lien cassé](#) (25/01/2008 )

On parle d'un lien cassé (*broken link* en Anglais) si l'internaute, après une clique, n'atterrit pas sur la page désirée mais sur une page d'erreur 404.

L'apparition des liens cassés arrive si le site ou la page ciblé soient *offline* ou n'existe plus, si la page ciblée a changé de nom ou l'emplacement ou tout simplement si le Webmaster a mal structuré son site.

Les robots ne savent évidemment pas la raison pourquoi un lien (*link* en Anglais) soit cassé et "pensent" donc, quelquefois à tort, que le site avec les liens cassés ne soit plus entretenu et n'apporte donc plus rien d'important aux internautes. Un ou très peu de liens cassés vont donc déclasser le site et le PageRank descende. S'il y a des nombreux liens cassés le site peut complètement disparaître des bases de données des moteurs de recherche.

### [lien contextuel](#) (03/01/2008 )

Un lien contextuel est un lien publicitaire qui se base sur l'analyse du contenu de la page d'où le lien part. Un programme analyse le contenu, en règle générale selon les

mots clés principaux, pour y ajouter des liens vers d'autres sites dont le contenu est en relation avec le contenu de la page de base.

L'espace publicitaire est, dans la majorité des cas, facturé en fonction des clics générés. Pour que le système de liens contextuels fonctionne pour les deux parties il est nécessaire que les pages avec les annonces publicitaires soient correctement optimisées.

L'exemple le plus typique des liens contextuels est le système AdSense de Google selon lequel seulement de liens en relation avec les mots clé principaux sont la base pour les liens ajoutés. Un site non optimisé donne donc de résultats insatisfaisants pour les internautes et le site qui a souscrit à ce système de publicité.

### [liens croisés](#) (22/12/2007 )

Un nombre important de Webmasters qui possède plusieurs sites sur de sujets divers lient souvent l'ensemble de leur sites, même si le contenu de ces sites n'a peu en commun. Ce système de liaison s'appelle liens croisés (*cross linking* en Anglais).

Vu que ces liens se trouvent souvent sur le même serveur et partagent aussi la même IP, les robots se rendent évidemment compte de cette méthode qui tente à accroître artificiellement la popularité d'un site et inversent l'effet souhaité. L'emploi de trop de liens croisés est donc pénalisé par la majorité des moteurs de recherche.

L'algorithme des moteurs de recherche cherchent, en plus, un plus value dans les pages ciblées et cherchent donc une relation entre les sites en question. Si un lien d'une page qui parle de la philosophie propose donc un lien vers une page qui parle de l'animal de compagnie de la famille, la liaison n'a donc aucun sens.

### [lien de retour](#) (20/12/2007 )

Les liens de retour (*backlinks* en Anglais) sont devenus une nécessité pour obtenir un bon *PageRank* et pour apparaître dans les premières pages des moteurs de recherche. Un Webmaster est aujourd'hui donc obligé de faire la quête pour obtenir ces fameux liens de retour.

Un lien de retour est un lien qu'un autre site a inséré dans ses pages qui cible en règle générale la page d'accueil du propre site. Les moteurs de recherche "pensent" que de tels liens ciblent exclusivement de sites de qualité et de sites qui ont une certaine importance. Leur algorithme prend toutefois en considération si les liens de retour ne proviennent pas de fermes de liens et si le contenu ciblé soit en relation avec la page qui a inséré le lien.

Un lien de retour ne doit pas être confondu avec un lien croisé ou un échange de liens qui prennent moins d'importance lors du référencement qu'un lien de retour. Il est donc important de trouver de *backlinks* de qualité ce qui n'est pas toujours évident et pose souvent un problème d'optimisation de sites.

### [liens en footer](#) (30/01/2008 )

Il y a de nombreuses pages Web où un nombre important de liens sont placés en bas des pages, donc dans les *footers*. C'est pourquoi on parle de liens en *footer*.

Quelquefois ces liens sont à cause de la couleur ou de sa taille à peine lisible, ce qui n'est pas le cas pour de liens importants.

Vu que la majorité des référenceurs malhonnêtes utilisaient (et utilisent) cette méthode pour améliorer le positionnement de leurs pages et ces liens mènent souvent vers de *doorway pages* ou sont utilisés pour le *cross-linking* et l'échange des liens, les moteurs de recherche ont déclassé ces liens. Leur valeur est donc en dessous de liens normales et ne peuvent plus servir à l'optimisation des pages Web.

Les liens en *footer* doivent donc, dans l'avenir, seulement être de liens qui sont effectivement de mineure importance pour les internautes et ne plus mener vers des informations importants du site.

### [lien naturel](#) (22/02/2008 )

Un lien naturel est un lien à l'intérieur d'un texte qui mène vers une page qui offre des informations supplémentaires sur le mot ou l'expression en question.

Le texte de ces liens fait partie intégrale du texte de la page et se remarque exclusivement par la technique employé qui permet de distinguer un lien dans un texte. Les liens naturels sont placés sans que le Webmaster prenne la popularité des liens en considération.

Les liens naturels joue donc un rôle très varié dans l'optimisation des pages Web, pour le site qui emploie les liens naturels ainsi que pour les pages ciblées par ces liens.

### [linkbaiting](#) (31/03/2008 )

Les mesures (honnêtes et malhonnêtes) destinées à persuader d'autres sites d'ajouter des liens vers le propre site s'appellent *linkbaiting* qu'on pourrait traduire par "amorcer de liens".

La plupart de temps il s'agit de créer des contenus uniques qui persuadent autres sites d'insérer un lien vers ces pages. Ces pages ont souvent peu ou rien à voir avec le but principal du site et servent donc exclusivement du *linkbaiting*. Cette technique fait donc la plupart de temps partie de la *Black Hat SEO* avec les conséquences adéquates.

Le *linkbaiting*, donc la tentative d'augmenter la popularité d'un site artificiellement, n'a donc peu en commun avec une liaison normale de sites. Plusieurs référenceurs créent même de blogs et forums exclusivement destinés au *linkbaiting*. La méthode fonctionne aussi longtemps jusqu'un concurrent signale l'abus auprès des moteurs de recherche.

### [Lisibilité](#) (23/04/2009 )

Une page Web n'est pas seulement optimisée pour l'indexation et le positionnement, mais aussi pour les internautes. Sans leurs visites répétées et un temps assez important qu'ils passent sur une page, l'optimisation ne sert à rien.

C'est pourquoi il faut créer de pages Web conformes aux exigences de la lisibilité. Lisibilité signifie à créer de pages bien lisibles pour tout le monde, aussi pour ceux qui ont de troubles de vue.

Mais lisibilité signifie aussi à structurer le texte de chaque page de la manière que chaque visiteur peut trouver ce qu'il cherche seulement en survolant le contenu. L'emploi de sous-titres est donc une double obligation.

### [liste noire](#) (17/12/2007 )

Chaque moteur de recherche dispose d'une liste noire (*blacklist* en anglais) qui est établi selon de critères diverses. Aucun moteur ne publie donc ses paramètres ce qui impose aux Webmasters de ne jamais négliger aucune règle de l'optimisation pour le référencement.

L'utilisation de toute technique frauduleuse (*spamdexing*, ferme de liens, pages satellites etc.) peut être détecté par les robots et sera puni d'une manière ou d'une autre. Dès que les moteurs de recherche trouvent donc un site optimisé de la manière que le positionnement soit amélioré par la triche ils le mettent sur la liste noire. Le site sera donc, selon la méthode utilisé par le Webmaster, désindexé pour un certain temps ou pour toujours.

Mais il n'y a pas seulement les robots qui détectent les méthodes malhonnêtes du référencement (*black hat* référencement). Chaque internaute et surtout chaque concurrent peut dénoncer un site qui, par la suite, pourrait être banni des moteurs de recherche.

### [log file](#) (29/02/2008 )

Chaque site dispose d'un fichier *log* (*log file* en Anglais) sur le serveur qui fournit toutes les informations nécessaires pour optimiser ses pages et pour analyser les visites et les visiteurs du site.

Le fichier log peut être brut ou traité de la manière que même le "non initié" puisse le lire sans avoir recours à un logiciel d'analyse. Les *log files* enregistrent des informations comme adresse *IP* du visiteur, le navigateur utilisé, etc. qui permettent de connaître les *hits*, les pages vues, les visiteurs uniques etc. selon le programme d'analyse et la configuration du serveur.

Seulement l'analyse profonde des *log fils* permet d'optimiser les pages Web d'une manière idéale vu qu'ils fournissent toutes les informations qui concernent les visiteurs et leur comportement.

### [mauvais voisinage](#) (08/04/2008 )

Le pire ce qui peut arriver à un Webmaster-Référenceur est d'atterrir dans un mauvais voisinage (*bad neighbourhood* en Anglais). Un problème qui arrive parfois si on ne dispose pas d'un serveur dédié mais est obligé de partager un serveur avec d'autres Webmasters, et si on a la malchance qu'on utilise donc la même *IP* qu'un spammeur ou celle d'un site banni.

Les moteurs considèrent dans ce cas le site qui utilise la même *IP* également comme malhonnête et le site en question disparaît donc de leur index. Mais le mauvais voisinage n'est pas seulement de partager une *IP* banni. Un Webmaster qui soumet son site auprès des fermes des liens ou ajoute sur sa page un lien qui mène vers un site de poker ou un site pornographique se retrouve également en mauvais voisinage et risque des problèmes.

Il est donc très important de choisir un hébergeur sérieux qui choisit bien ses clients en qui surveille l'espace alloué à ses clients. Et il est important de bien choisir les sites qu'on cible par ses liens. Les référenceurs utilisent en règle générale exclusivement de liens qui ciblent des sites qui disposent d'un PageRank.

### [menu déroulant](#) (24/04/2009 )

La majorité des menus déroulants se basent sur JavaScript et posent donc plusieurs problèmes en même temps.

Il y a le problème de compatibilité entre les divers navigateurs, le problème que de tels menus ne fonctionnent pas si l'internaute a bloqué les scripts et, finalement, il y a le problème avec les robots qui ne comprennent pas JavaScript.

L'utilisation de menus déroulantes en JavaScript comme seule moyen de navigation signifie donc que les robots ne peuvent pas suivre les liens du script et, par la suite, ne pas indexer les pages ciblées par ces liens non plus.

### [meta-moteur](#) (02/01/2008 )

Un meta-moteur est un outil de recherche (programme) qui ne cherche pas dans un seul index des moteurs de recherche mais dans plusieurs en même temps. Selon la qualité du meta-moteur, la recherche peut s'éteindre sur plusieurs centaines des moteurs de recherche.

On distingue généralement deux systèmes de meta-moteurs. Ceux qui sont installés dans l'ordinateur de l'internaute et ceux qui fonctionnent online. Dans les meta-moteurs l'internaute peut souvent modifier certains paramètres de recherche pour mieux cibler les résultats. Vu que les meta-moteurs cherchent les résultats dans les bases de données d'autres moteurs, leur vitesse est toutefois souvent assez restreinte.

L'optimisation des pages Web ne peut pas directement influencer les résultats des meta-moteurs, mais leur connaissance aide à savoir quels moteurs de recherche soient important pour la soumission des sites. En ciblant les moteurs de recherche qui font partie du système des meta-moteurs, les résultats sont automatiquement améliorées.

### [meta refresh](#) (23/02/2008 )

On parle du *meta refresh* si un Webmaster utilise la balise <META> pour que la page sera automatiquement actualisé après un temps prédéfini ou pour ouvrir un autre URL que celui que l'internaute a utilisé lors de la recherche, donc pour rediriger l'internaute vers un autre site.

Les moteurs de recherche réagissent très sensible à cette technique et punissent son emploi en règle générale. Si une page change de nom, l'emplacement ou le serveur il est donc impératif d'utiliser la redirection 301 au lieu du *meta refresh*.

Un Webmaster qui veut optimiser ses pages Web pour les moteurs pensent toujours comment les robots voient la page et ne seulement comment les internautes la perçoivent.

### [metajacking](#) (17/02/2008 )

L'algorithme des moteurs de recherche inclue le facteur combien de personnes visitent un site en utilisant le moteur de recherche, cliquent donc sur le lien d'un résultat.

Webmasters et référenceurs peu scrupuleux utilisent donc de nouveau la *Black Hat SEO* et emploient le *metajacking* en espérant de récupérer de cette manière les visiteurs d'autres entreprises. Ils incluent donc dans les balises *meta* des noms et slogans d'autres entreprises et parfois même des noms de marques protégés. Cette technique s'appelle *metajacking*.

Il est logique que les algorithmes reconnaissent certaines de ces noms (= mots clés). S'il retrouve donc un *metajacking*, le site en question risque d'être banni au lieu de récupérer de visiteurs. Cette technique cible évidemment exclusivement le positionnement dans les moteurs de recherche vu que les internautes qui atterrissent sur un tel site soient assez déçu de ne pas retrouver les informations qu'ils cherchent. Ils peuvent, par un simple clique, dénoncer un tel site auprès de Google.

### [mot de recherche](#) (13/12/2007 )

Un Webmaster définit pour chaque page Web un ou plusieurs mots clés par lesquels il pense que les internautes la cherchent. Ces mots correctement repartis sur la page augmentent la chance que la page se retrouve dans une bonne position dans les moteurs de recherche.

Les internautes ne suivent pas forcément la logique du Webmaster et cherchent une page avec un ou plusieurs mots de recherche (*search term* en Anglais) dans Google, Yahoo, MSN ou un des autres moteurs de recherche.

Il est donc extrêmement important que les mots clés et les mots de recherche se correspondent. Un Webmaster est donc obligé de surveiller régulièrement les mots de recherche pour adapter ses pages aux besoins des internautes.

### [mots empoisonnés](#) (12/12/2007 )

Chaque page Web doit disposer de quelques mots clés qui se répètent selon un certain algorithme dans le contenu et le code. Une page est donc optimisée de la manière que les moteurs de recherche fournissent un résultat dans les premières pages si l'internaute cherche un de ces mots ou une combinaison de mots clés.

Les moteurs de recherche ne donnent toutefois pas à chaque mot la même valeur. Il y a des mots que certains moteurs ne veulent pas indexer. Si un Webmaster utilise donc un de ces mots pour gagner un PageRank, il peut le même perdre.

Il est donc important d'éviter les mots empoisonnés (*poison words* en Anglais). Il n'est toutefois pas toujours facile à savoir quel mot clé soit empoisonné. Les mots qui sont en relation avec de partenariats, même s'il s'agit de Amazon, peuvent donc entrer dans ce catégorie de mots empoisonnés.

### [mots vides ou mots d'arrêt](#) (15/12/2007 )

Les mots sans contenu comme le, la, du etc., de mots alors qui ne disent rien en soi, sont considérés comme mots vides (*stop words* en Anglais) et ne comptent donc pas pour les moteurs de recherche.

Les moteurs de recherche utilisent les mots vides à deux buts différents. Ils sont "écartés" lors de la recherche et ne sont pas considérés comme contenu de la page, mais ils servent aussi à détecter le *keyword stuffing*, le bourrage de mots clés.

Il est donc nécessaire d'utiliser les mots vides pour ne pas atterrir sur les listes noires des moteurs de recherche mais déduire leur nombre du nombre entier des mots qui constituent le contenu d'une page Web.

### [navigation](#) (09/02/2008 )

La navigation est la partie de la page du laquelle l'internaute trouve les autres pages du site. La navigation se base donc sur la structuration du site et suit des règles de la logique ainsi que de la règle des trois cliques.

Même si la navigation "classique" soit graphiquement pas la plus jolie, en optimisation des pages pour le référencement c'est la seule qui soit vraiment valable et efficace. Les robots ont besoin de texte pour comprendre les liens, ils ne sont pas capables à déchiffrer une image et encore moins d'analyser un *JavaScript* ou une navigation en Flash. De telles techniques empêchent donc que les robots trouvent les pages à l'intérieur du site ce qui peut être catastrophique sur le plan économique.

Si un Webmaster ne veut pas se contenter de la simplicité d'une navigation en HTML/xhtml et optimisé pour les moteurs, il n'a qu'un seul choix, travailler avec deux navigations parallèles ce qui, en règle générale, n'est pas trop idéale pour garder une optique agréable du site.

### [narrowcasting](#) (19/02/2008 )

On parle du *narrowcasting* si un Webmaster crée un site qui cible un publique très restreinte mais très ciblé qui n'utilise que de mots clé très spécifiques.

Un site de *narrowcasting* n'a donc très peu de sites concurrentiels et atterrit donc facilement dans la meilleur position possible des moteurs de recherche, même sans employer toutes les techniques de l'optimisation de pages Web. Le site se trouve généralement dans une niche d'audience, ce qui est souvent le cas pour des sites publicitaires.

*Narrowcasting* ne signifie toutefois pas que le site en question n'aurait que peu de visites ou exclusivement les visiteurs ciblés. Décisif est plutôt la stratégie marketing/optimisation employé en ce qui concerne le succès d'un site de *narrowcasting*.

### [néophobie](#) (03/04/2008 )

La néophobie est l'effet de, consciemment ou inconsciemment, refuser les nouveautés ou changements d'un site. Elle joue donc un rôle assez important pour le webdesign ou le redesign d'un site.

Un site qui parle des techniques récentes ou un site pour les *gamers* demande l'emploi des techniques modernes et permet aussi le design créatif. Il s'adresse donc aux néophiles. Mas la majorité des sites touche justement les personnes qui ont la néophobie et qui ferment une page seulement à cause de l'effet que le logo a changé ou que la navigation se ne trouve plus au même endroit.

Il est donc nécessaire d'adapter les techniques de l'optimisation au public ciblé et cela concerne chaque élément d'une page et s'étend même au choix des mots employés dans le texte et le style grammatical.

### [netlinking](#) (07/01/2008 )

Le mot *netlinking* signifie basically rien d'autre que l'échange des liens mais changeait dernièrement, à cause des méthodes du référencement et du *web-marketing* la signification.

Le *netlinking* se base donc maintenant sur le fait de chercher des annuaires, portails et sites dont le contenu est en relation avec le propre contenu, d'y insérer ses liens ou de les échanger, pour favoriser le positionnement du propre site.

Vu que les liens n'ont pas tous la même valeur pour les moteurs de recherche et les liens des fermes de liens peuvent même être nuisibles, il est devenu nécessaire d'appliquer une stratégie, quelquefois pas plus propre que l'utilisation des fermes des liens, pour persuader les moteurs de recherche d'afficher un site en position privilégiée.

### [nofollow](#) (17/01/2008 )

Pour indexer un site, les robots suivent en règle générale les liens qu'ils trouvent dans une page d'accueil jusqu'au troisième niveau. Ils ne suivent pas seulement les liens internes mais aussi les externes.

Cet effet est également utilisé pour comptabiliser les liens de retour qui jouent un rôle important pour le *PageRank*. Pour éviter que les robots suivent les liens, le Webmaster a deux possibilités. Utiliser la balise <META> du *head* ou d'ajouter aux liens la valeur *nofollow*. De cette manière les robots ne prennent les liens concernés pas en compte.

Il est donc, de l'autre côté, important de savoir si un lien de retour auprès d'une annuaire ou d'un autre site ne soit pas marqué avec la valeur *nofollow* (ne pas suivre). Il y a malheureusement de Webmasters qui utilisent *nofollow* aussi lors d'un soi-disant échange de liens.

### occurrence (21/12/2007 )

L'occurrence et le nombre de fois qu'un mot, une combinaison de mots ou une phrase apparaissent dans le contenu, le texte, d'une page Web.

L'occurrence est utilisée pour calculer, selon un certain algorithme, l'indice de densité de mots clés. Elle fait toutefois seulement partie des valeurs décisives pour le positionnement d'une page vu que les mots clés n'ont pas tous la même importance. En vérifiant l'occurrence le Webmaster a désormais une idée à partir de quel moment il commence à suroptimiser une page.

La faute la plus importante de nombreux Webmasters est de cibler les mots clés individuellement au lieu de calculer l'occurrence des expressions complètes selon lesquels les internautes cherchent une page.

### optimisation (01/01/2008 )

Optimisation désigne la technique de préparer les pages Web de la manière qu'ils obtiennent le meilleur *PageRank* possible et soient donc affichés dans la première page des résultats des moteurs de recherche.

Chaque page doit, en règle générale, être optimisée selon les règles dictées par les moteurs de recherche. L'optimisation utilise une centaine d'éléments, dont les titres, le contenu, les metas, les liens, les attributs etc. pour mettre en évidence un ou trois mots clés par page. Dans l'avenir chaque Webmaster professionnel doit maîtriser au moins les bases de l'optimisation pour que ces clients atterrissent encore dans les moteurs de recherche dans une position "acceptable".

L'optimisation porte seulement des résultats sous la condition que la majorité des pages Web ne soient pas optimisées, ce qui est actuellement le cas. Mais vu que l'optimisation demande de connaissances approfondies sur les algorithmes employés par les moteurs et des heures de travail, il est peu probable que cette situation changera dans les années à venir.

### off the page (20/03/2008 )

Pas seulement les éléments qui se trouvent sur une page décident du *PageRank* et de l'emplacement des pages dans un moteur de recherche. Ils s'y ajoutent aussi des éléments parfois difficilement à gérer qui influencent l'optimisation. Nous parlons ici des éléments *off the page*, en dehors de la page.

Les éléments *off the page* que les algorithmes prennent en compte sont multiples. Il s'agit des textes des liens qui mènent vers nos pages, de la qualité des liens qui ciblent notre site mais aussi du nom de domaine, de l'hébergeur ou du taux de clics par jour.

Il est donc primordial en optimisation de planifier un site même avant d'acheter un nom de domaine ou de choisir un hébergeur. Seulement un planning précis permet ensuite d'optimiser une page à 100 pour cent.

### [page dynamique](#) (06/05/2009 )

Une page dynamique se base, en règle générale, sur une base de donnée qui permet l'affichage des pages selon la souhaite et le besoin des internautes.

Les pages dynamiques fonctionnent à l'aide de scripts PHP, JSP, ASP, Perl ou d'une autre langage de programmation qui permet l'échange des informations avec une base de donnée et qui permet une programmation avec des variables.

L'avantage des pages dynamiques est la vitesse avec laquelle une information spécifique apparaît sur l'écran. Le désavantage est le fait que les pages dynamiques pose souvent de problèmes de référencement et ont donc besoin d'une structuration absolument logique et selon les règles de l'optimisation très précis.

### [page jacking](#) (31/12/2007 )

*Page jacking* signifie de copier le code source (ou d'écrire une code très proche de l'original) d'une page qui est en première position dans les moteurs de recherche. Il s'agit donc des titres, du texte, des balises meta et d'autres éléments utilisés pour l'optimisation des pages Web.

Il est évident que cette méthode fait partie de la *Black Hat SEO* et n'enfreint pas seulement les règles des moteurs de recherche mais est en même temps une violation de propriété intellectuelle et peut avoir de suites juridiques.

La méthode du *page jacking* est sanctionné par les moteurs de recherche et ne produit donc plus l'effet souhaité. Dans le meilleur des cas, les moteurs mettent la page dans la catégorie "pages similaires" et ne les affichent pas.

### [page orpheline](#) (27/12/2007 )

On parle d'une page orpheline si cette page est seulement accessible par l'utilisation de l'adresse URL parce qu'il n'y a aucun lien qui la cible.

Une telle page n'a que peu de chance d'être référencé par un moteur de recherche vu que leurs algorithmes se basent sur une grande partie sur l'emploi de liens de retour. Si une page orphelin soit toutefois assez connue et si elle dispose de nombreux liens vers d'autres pages, elle peut servir pour que d'autres pages du même site ou d'autres sites ciblés obtiennent un meilleur emplacement dans la base de données des moteurs de recherche.

Les pages orphelins apparaissent souvent après l'actualisation d'un site et restent parfois assez longtemps accessibles, surtout si le Webmaster en question a négligé d'utiliser une bonne structuration.

### [page statique](#) (16/01/2008 )

Une page statique (*static page* en Anglais) ne contient aucun contenu dynamique et ne comporte donc aucun code en PHP, ASP ou JAVA. Elle est donc rédigée seulement en xhtml en utilisant les feuilles de style.

Une page statique est l'idéal pour l'optimisation des pages Web vu que les robots

rencontrent ici seulement des éléments claires, généralement identiques et facilement à évaluer à l'aide de leur algorithme. Un Webmaster essaie donc toujours d'utiliser un maximum des pages statiques à l'intérieur de leurs sites.

Il est logique qu'il n'y a presque plus de site qui se base exclusivement sur les pages statiques. C'est pourquoi le développement d'une bonne structuration et l'emploi du fichier robots.txt deviennent nécessaires.

### [PageRank \(PR\)](#) (29/11/2007 )

Le nom PageRank, une marque déposée par Google, fait allusion à Larry Page qui était cofondateur de Google et indique la valeur que Google distribue à certaines pages d'un site, le plus souvent il s'agit de la page d'accueil. Le PageRank est décisif pour l'ordre de liens dans les résultats de recherche.

Google utilise une centaine de paramètres pour définir le "vrai" PageRank d'une page Web dont la quantité et la qualité des liens internes et des liens externes joue actuellement un rôle très important. Le PageRank n'est donc pas forcément un critère de qualité d'un site et favorise la triche par les Webmasters malhonnêtes.

Il y a un nombre important des outils sur Internet qui permettent de connaître le PageRank approximative d'un site et chaque Internaute qui installe la Google Toolbar est informé du PageRank de la page ouverte dans son navigateur. Il s'agit d'une valeur entre 0 et 10. Une barre grise indique les sites bannés et les sites trop récents pour déjà avoir une PageRank.

Google attribue les PageRanks dans un rythme qui varie. Un site doit en règle générale exister un certain temps avant que l'algorithme de Google montre un résultat. Les PageRanks seront ensuite réévalué varie deux à trois mois.

Un des outils qui informent sur le PageRank et l'explication du PageRank se trouvent sous les adresses suivantes :

[Google Dance et PageRank](#)  
[Page Rank \(PR\), un indicateur de Google](#)

### [PageRank hoarding](#) (01/03/2008)

On parle du *PageRank hoarding* si un Webmaster ou un référenceur utilise de techniques ou méthodes qui lui permettent à ramasser un maximum de liens vers "son" site mais en utilisant seulement un minimum de liens externes qui mènent vers d'autres sites.

Le *PageRank hoarding* fait toutefois partie de la *Black Hat SEO*. Il a pour but à augmenter le *PageRank* de certaines pages ou de la page d'accueil du site et fait donc, jusqu'à une limite, partie des techniques d'optimisation des pages Web.

Cette méthode est assez facilement à détecter par les robots des moteurs de recherche et, s'il s'agit d'un abus selon l'algorithme employé, le site sera donc puni et peut même perdre son *PageRank*.

### [piège à robots](#) (08/03/2008 )

L'expression piège à robots (*spider trap* en Anglais) a plusieurs significations. Nous parlons donc ici seulement de sa signification dans l'optimisation des pages Web.

Les pages dynamiques mal structurées et mal employées sont des pièges à robots par excellence. Vu que les robots rencontrent dans ce cas chaque fois une "autre page" lors d'une visite ou même pendant une seule visite (parfois même avec d'autres noms, également géré dynamiquement), ils doivent scruter des milliers de pages qui se basent, peut-être seulement sur une vingtaine de pages vraiment différentes.

Il est donc logique que la majorité des moteurs n'apprécie pas l'emploi exagéré des pages dynamiques et "bloquent" l'abus par l'algorithme employé. Les pages trop dynamiques ne sont donc pas acceptées par les robots et chaque abus est puni. Il est donc nécessaire de bien réfléchir avant de dynamiser chaque page de son site.

### [polices en gras](#) (06/02/2008 )

Les robots des moteurs de recherche donnent aux mots qui apparaissent en gras plus de valeur qu'aux autres mots. Il faut donc correctement employer les balises `<b>` et `<strong>` qui, toutefois n'ont pas la même signification.

Correctement et astucieusement employé, les mots en gras attirent les yeux des internautes et sont donc lus avant l'autre texte. Cet effet joue donc aussi un rôle dans l'algorithme utilisé par les moteurs de recherche. Mais il ne sert logiquement pas à grande chose si trop de mots ou des paragraphes entières sont en gras. Cela empêche plutôt la lecture normale et produit l'effet inverse.

L'optimisation "joue" donc sur la valeur des mots et surtout des mots clés lors de l'utilisation du style en gras. L'abus est toutefois, comme toujours, puni par les moteurs de recherche. Il vaut donc mieux de bien choisir les mots en gras.

### [popularité des liens](#) (13/02/2008 )

La popularité des liens est un facteur non négligeable dans l'optimisation des pages Web pour le référencement. Elle est la quantité des bons liens qui ciblent un site. Chaque moteur de recherche utilise aujourd'hui ce nombre dans l'algorithme employé lors de l'indexation des sites.

La popularité des liens est toutefois un facteur difficile à évaluer et ne pas entièrement à gérer. Les moteurs ne comptent pas chaque lien, mais seulement les liens de qualité qui ont un rapport avec le site en question. Il ne sert donc à rien d'atterrir dans un mauvais voisinage ou d'utiliser les fermes à liens. Chaque lien qui cible un site a une autre valeur qui dépend de nombreux facteurs différents.

Pour obtenir un bon emplacement dans les moteurs de recherche il est donc nécessaire et obligatoire d'avoir des liens qui ciblent le site. Selon la qualité des liens, la qualité du site et la concurrence existante il suffit parfois de 10 à 20 liens pour être indexé et pour obtenir un bon emplacement, mais un site avec beaucoup de concurrence de bonne qualité demande parfois de centaines ou même de milliers de liens de qualité qui pointent sur le site pour pouvoir être affiché sur la

première page des moteurs. Il est donc nécessaire d'étudier cette question avant de créer un site.

### [popularité des mots clés](#) (14/02/2008 )

La popularité des mots clé désigne les mots le plus habituels que les Internautes utilisent lors de leur recherche dans les moteurs de recherche. Il est donc important d'utiliser ces mots pour l'optimisation ses pages et ne pas les mots que personne n'utilise.

Il est donc important de savoir quels mots la majorité des internautes utilisent et d'adapter ses textes aux habitudes des internautes. Un thésaurus aide souvent déjà lors de la préparation du texte. Il faut, par exemple savoir, si la majorité des internautes qui cherchent une formation utilisent le mot formation, ou stage, ou éducation, ou cours, ou apprentissage ou ... lors de la recherche pour que le suite puisse être trouvé.

Les référenceurs utilisent en règle générale des logiciels spécialisés pour tracer la popularité des mots clé avant que la première ligne du texte soit rédigée. Mais un Webmaster peu également utiliser un moteur de recherche comme Google pour se donner une impression de la popularité des mots clé. Parfois il est aussi nécessaire d'utiliser plusieurs synonymes d'un mot donné dans le contenu et non pas seulement le mot clé le plus populaire.

### [positionnement web](#) (07/04/2008 )

Le positionnement Web (*Web positioning* en Anglais) est la technique qui permet à référencer un site Web par l'analyse de la syntaxe du contenu. Le but est donc de promouvoir le site pour la meilleure visibilité dans les moteurs de recherche.

La base est donc l'analyse profonde de chaque page d'un site pour adapter le contenu (le texte) de la manière qu'une machine, les robots, soient capable d'évaluer le contenu réel, de reconnaître les mots les plus importants et à "comprendre" que le site doit être affiché dans la première page des résultats si l'internaute utilise les mots les plus vraisemblables lors de la recherche pour trouver, par ces mots, le contenu qu'ils cherchent.

Ce travail demande souvent de spécialistes en web copywriting, e-marketing et référencement. Dans les petites et moyens entreprises, c'est le Webmster qui doit occuper ces professions en même temps que celle du programmeur et webdesigner.

### [promotion d'un site](#) (11/02/2008 )

La promotion d'un site est chaque action qui cible d'attirer plus de visiteurs et de faire connaître un site. La promotion d'un site commence donc avec le développement du contenu, passe par l'optimisation et la soumission ainsi que toute autre technique apte à attirer de visiteurs intéressé par le sujet et le contenu du site.

Vu que les résultats de l'optimisation permettent seulement d'obtenir une meilleure position dans les moteurs de recherche pour que les internautes trouvent le site plus facilement, il y a d'autres techniques jointes qui sont le lien entre optimisation et le e-marketing. La majorité des référenceurs utilisent donc les deux techniques

ensembles.

La promotion d'un site demande donc, selon le contenu du site, un travail étroite avec la presse spécialisé, l'emploi des listes de diffusion (*mailing-lists* en Anglais), la création de *blogs*, la participation dans des forum spécialisés et même la rédaction de textes pour les revues et autres publications *online*. Le Webmaster doit donc développer un plan de promotion précis pour chaque site qu'il crée s'il veut obtenir des résultats.

### [redirect](#) (23/01/2008 )

La technique du *redirect* est utilisée pour afficher une autre page que celle que l'internaute a choisie soit dans la fenêtre du navigateur ou en utilisant un lien ou simplement, soit immédiatement après que la page choisie commence à s'afficher.

Même si le *redirect* puisse être, au premier vue, une solution pour un Webmaster qui a changé de nom de domaine, le provider, partagé le contenu d'un site etc., dans la majorité des cas le *redirect* sert à attirer les internautes vers un site qu'il ne soit pas indexé ou un site commercial qui propose de produits qu'ils ne cherchent pas.

C'est aussi la raison pourquoi les moteurs de recherche en règle générale ne suivent pas les *redirects* et cette solution est donc exclue si on veut optimiser un site ou si on change le nom de domaine.

### [réécriture d'URL](#) (30/12/2007 )

L'emploi des pages dynamiques en *PHP*, *ASP* ou *JAVA* pose souvent un problème de référencement vu que ces pages prennent parfois des extensions assez longues en ajoutant des caractères non-standard selon les moteurs de recherche.

Pour éviter ces problèmes un Webmaster utilise donc le réécriture d'URL (*URL rewriting en Anglais*), c'est-à-dire il "programme" le serveur de la manière que ces adresses "illisibles" soient transformées aux adresses lisibles par les robots et les internautes. Chaque page aura de cette manière seulement une extension standard, mais unique, par exemple produit-livre.php.

Utiliser de pages dynamiques sans employer le *URL rewriting* signifie donc négliger l'optimisation des pages Web et négliger le référencement auprès des moteurs de recherche.

### [référencement automatique](#) (08/01/2008 )

Lors d'un référencement automatique un site est soumis à de nombreux moteurs de recherche par le biais d'un logiciel. Il s'agit parfois de centaines de moteurs à la fois.

Comme cette technique se base sur la quantité et ne pas sur la qualité, le référencement automatique n'est pas adapté aux besoins des moteurs et ne rapporte donc certainement pas le résultat espéré. Un Webmaster ne gagne donc pas du temps en utilisant cette méthode mais risque, au contraire, qu'un nombre important de moteurs refuse d'indexer le site soumis.

De plus en plus de moteurs regardent cette technique comme "malhonnête", avec le résultat qu'un site référencé automatiquement peut atterrir dans les listes noires des moteurs. Il est donc absolument déconseillé d'utiliser le référencement automatique.

### [referrer](#) (21/03/2008 )

Dans l'optimisation des pages Web il est nécessaire de savoir de quelle page un internaute vient sur une certaine page à l'intérieur de notre site. Cette page qui précède de la page ouverte s'appelle *referrer*.

*Referrer* peut être une page à l'intérieur du même site, une page située dans un autre site ou la page d'un moteur de recherche ou d'une annuaire. Seulement cette information permet au référenceur d'adapter la page aux espérances du publique.

Le *referrer* est toujours une *URL*, donc un lien absolu qui est inscrit dans le log file sur le serveur. Pour savoir de quelle page les Internautes proviennent, il est donc nécessaire d'analyser ces *log files*.

### réinclusion (26/01/2008 )

Si un site a été déclassé au PageRank 0 ou s'il a été effacé de la base de données des moteurs de recherche à cause du *Black HAT SEO* comme l'utilisation de texte invisible ou du spamming, le Webmaster peut, si les problèmes n'existent plus, demander la réinclusion.

Avant la demande d'une réinclusion chaque problème comme le mauvais voisinage ou d'autres problèmes possibles doit être résolu. Selon la gravité de l'infraction, il suffit parfois de laisser jouer le temps ou de re-soumettre le sitemap.xml. Si le spamming était responsable du "punissement", ces actes ne suffisent en règle générale pas et il faut adresser un mail aux moteurs concernés, s'expliquer et de demander par cette manière la réinclusion.

Il n'y a toutefois aucune garantie que la demande soit traitée favorablement et en cas de "fautes graves" il ne faut pas compter d'être accepté de nouveau de si tôt.

### resoumission (03/02/2008 )

On parle d'une resoumission si un Webmaster essaye de soumettre un site plusieurs fois aux mêmes moteurs de recherche en croyant que cela augmentera sa chance d'obtenir un bon emplacement dans les résultats.

La resoumission semble aux Webmasters souvent important vu que leurs pages n'apparaissent pas dans l'espace d'une semaine dans les résultats ou n'apparaissent pas du tout. Mais il y a aussi de Webmasters qui pensent sur l'effet "de plus que je soumetts de plus j'ai la chance que les robots trouvent les pages". C'est le contraire !

Les robots n'aiment pas de travail supplémentaire et s'il n'accepte pas un site ou certaines pages, il y a un problème sur ces pages ou le site en question ou, tout simplement, le robot n'avait pas encore le temps de passer. La resoumission a de toute manière l'effet que le site soit "puni" est descendu dans l'échelle ou, s'il y a une resoumission répétitive, disparaît complètement de l'index.

### résultat naturel (12/01/2008)

Les résultats naturels sont les sites affichés par un moteur de recherche ou d'un annuaire qui suivent exclusivement des règles de l'algorithme, on parle ici aussi du classement naturel.

Un résultat naturel exclut donc chaque lien sponsorisé, promotionnel ou contextuel. Le placement d'un site dans un résultat naturel dépend énormément de l'optimisation du site, de l'ancienneté du site, du nom de domaine et des pages et du sujet. En

règle générale seulement un professionnel du référencement est capable de lancer un site dans une première page d'un moteur.

En réalité, les résultats naturels s'affichent qu'après les liens payants dans les moteurs de recherche et des annuaires. La quantité des liens payants dépend du moteur de recherche ou de l'annuaire et peut s'étendre sur plusieurs pages.

#### [résultats par pays](#) (16/03/2008 )

Beaucoup de Webmasters optimisent leurs pages en relation avec les résultats de recherche qu'ils obtiennent dans leur propre pays. Cela donne le résultat souhaité si le public ciblé se trouve également dans le même pays et que le site n'existe pas dans d'autres langues que celle du pays des visiteurs.

Dès qu'un site cible les internautes d'autres pays ou qu'il existe en plusieurs langues, cette méthode d'optimisation ne donnera jamais les résultats souhaités vu que les moteurs affichent dans chaque pays seulement les résultats liés aux pays et à la langue. Il y a même une différence si on cherche, (dans Google, par exemple) les pages du pays, les pages dans la langue spécifique ou les pages Web en général.

Si une page existe donc en Français, Anglais et Allemand, il est nécessaire de vérifier les résultats dans [google.fr](http://google.fr), [google.com](http://google.com) et [google.de](http://google.de). Chaque requête doit également être faite dans les trois possibilités que Google propose dans les pages de recherche respectives.

#### [robot \(spider en anglais\)](#) (27/11/2007 )

Il y a un nombre étonnant de **robots** (plus que mille) qui sont chaque jour à la recherche de nouveaux sites et de pages qui ont été modifiées. Il est donc logique que ces robots tombent, plus ou tard, sur un nouveau site, sous condition qu'un autre site a créé un lien qui pointe vers ce nouveau site. Mais cela peut prendre des jours ou des mois. Seulement la gestion de son site et la soumission correcte aide donc aux robots à trouver un site dans la ou les semaines qui suivent.

Chaque Webmaster a besoin de ces robots pour que son site soit retrouvé par les internautes, mais, de l'autre côté il y a aussi une centaine de robots qui n'ont rien à voir avec le référencement. Ces robots scrutent le net pour trouver des adresses mail ou de failles de la programmation pour s'introduire, d'envoyer du spam ou commettre une autre sorte de l'e-criminalité. Le problème d'un Webmaster est donc aussi de bloquer certains robots. La méthode la plus facile est l'utilisation d'un fichier `.htaccess` ou de modifier certains paramètres du serveur. Mais nous parlerons du fichier `.htaccess` séparément.

### [score](#) (02/03/2008 )

Les moteurs de recherche affichent les sites selon leur importance qui est calculé selon leur algorithme. Ils considèrent donc une page qui sort en première position dans leur résultat comme la plus importante sur le sujet en question.

Ce calcul des moteurs s'appelle *score*. La page qui est affichée en première position des résultats dispose donc du *score* le plus élevé. Le but de l'optimisation des pages Web est donc de "transformer" les pages de la manière que les moteurs comprennent que la page qu'ils trouvent chez nous soit la plus importante sur le sujet que nous avons indiqué dans le titre de la page.

Comme le *score* est calculé selon l'algorithme individuel des moteurs, le *score* n'est pas forcément identique chez les divers moteurs. Il en a qui prennent la page avec le meilleur *score* comme base pour attribuer un *score* aux autres documents, d'autres qui calculent un document idéale imaginaire et encore d'autres se fient aux mots clés.

### [script furtif](#) (04/04/2008 )

Un script furtif (*stealth script* en Anglais) est un script type *CGI* qui adapte le contenu de la page au "visiteur". Les internautes reçoivent dans ce cas pas la même page que les moteurs de recherche. Cette méthode fait évidemment partie de la *Black Hat SEO*.

Comme chaque nom des robots est connu, il est facile d'écrire un script furtif qui livre à chaque moteur de recherche une page optimisée selon les besoins du site ou du propriétaire du site. Les pages livrées aux internautes n'ont souvent rien en commun avec celles que "lisent" les robots et trompent donc les deux parties.

La majorité des robots est aujourd'hui capable de reconnaître cette technique et n'insèrent de telles pages plus dans leur index. Un site qui emploie cette technique risque même d'être banni des moteurs de recherche.

### [searchjacking](#) (24/02/2008 )

*Searchjacking* fait partie du *spamdexing* et signifie optimiser une page de la manière qu'elle soit trouvée aussi en utilisant des mots clés qui n'ont aucune relation avec le contenu de la page. La technique s'étend aussi dans l'optimisation des metas.

Le *searchjacking* fait partie du *Black Hat SEO* et l'emploi sera donc puni par les moteurs de recherche dès qu'ils découvrent l'utilisation de cette technique.

Mais le *searchjacking* est aussi une mauvaise idée du Webmaster vu qu'il attire des visiteurs non ciblés et non concernés par le contenu. Le trafic augmente, mais les internautes qui atterrissent sur la page la ferment aussitôt ou bloquent même l'*URL* et ne reviennent même plus pour un contenu qui les concerne.

### [site fantôme](#) (08/12/2007 )

Un site fantôme (*ghost site* en Anglais) est un site qui n'est plus entretenu depuis un certain temps mais qui reste visible aux internautes qui ne se rendent souvent même pas compte qu'ils visitent un site abandonné.

Contrairement aux internautes, les robots regardent la date de la mise à jour sur le serveur et découvrent donc les sites fantômes sans problème et les font lentement disparaître des premières pages de recherche, sauf, s'il y a de nombreuses *backlinks* (liens de retour) et si un nombre important des internautes continue à visiter ces site abandonné.

Si les Webmasters peu conscientes n'enlèvent donc pas les *backlinks* et les internautes se laissent tromper par un contenu dépassé, il est possible qu'un site fantôme garde parfois la première page dans les moteurs de recherche pendant des années et Internet perde une partie de sa réputation.

Plus d'informations sur les sites fantômes se trouvent sous l'adresse : <http://www.mariposa.eu.com/lexique/lex/ghosit.html>

### [site miroir](#) (28/12/2007 )

Un site miroir (*mirror site* en Anglais) est la copie d'un site qui est généralement hébergé sur un autre serveur. La raison de créer de sites doubles varie. Elle peut être de "rapprocher" un site géographiquement au publique cible, à équilibrer la bande passante ou simplement à "persuader" les moteurs de recherche d'indexer deux sites du même contenu.

Utiliser des sites miroirs, même si cela soit parfois nécessaire, oblige toutefois le Webmaster d'éviter, par un fichier robots.txt, d'empêcher les robots à visiter le double. Les moteurs regardent de tels sites comme *spamdexing* et duplication de contenu et pénalisent donc l'essaie de les faire référencer. Dans le meilleur de cas, ils n'acceptent pas l'indexation du site miroir.

Un Webmaster qui veut que ses sites soient correctement enregistrés par les moteurs de recherche, respectent les règles des moteurs et n'essaient jamais à soumettre un site miroir.

### [sitemap](#) (14/03/2008 )

Le mot *sitemap* s'applique à deux fichiers différents utilisés lors de l'optimisation des pages Web pour le référencement. Il y a le *sitemap* destinés exclusivement aux moteurs de recherche et invisibles pour les internautes et le *sitemap* destiné aux internautes et aux moteurs de recherche en même temps.

Le *sitemap* "classique" est celui qui a l'extension .html, .php, .asp etc. et qui est destinée aux moteurs de recherche ainsi qu'aux visiteurs. Ce *sitemap* contient, dans un ordre logique, tous les pages du site avec des liens vers chaque page. Un lien vers ce *sitemap* se trouve aussi dans chaque page du site pour que les internautes le trouvent sans aucun problème. Les visiteurs peuvent donc à partir de ce *sitemap* visiter chaque page que le site contient.

Les moteurs de recherche utilisent le même *sitemap* pour indexer les pages du site. L'avantage pour les moteurs est double. D'un côté ils "comprennent" ainsi la structure du site et de l'autre côté ils trouvent chaque page sans être obligé de suivre chaque lien sur chaque page. Ce *sitemap* devait toutefois être accompagné du *sitemap* exclusivement destiné aux moteurs avec l'extension .xml.

#### [snippet](#) (03/05/2009 )

Le *snippet* et le bout du texte que les moteurs de recherche, dont Google, affiche dans la page des résultats pour informer les internautes du contenu de la page.

Si la page Web est correctement optimisée, Google affiche comme *snippet* la valeur de l'attribut *description* de la balise META, mais se limite toutefois aux premières 160 caractères de la description.

Si Google ne trouve pas de description de la page, le moteur reprend un bout du texte de la page et le *snippet* ne sert donc pas beaucoup aux internautes qui espèrent à trouver une information précises qui puisse les guider lors de leur recherche.

Pour savoir plus sur les snippets, je conseil la visite de la page

[Les snippets et comment les optimiser](#)

#### [soumission d'articles](#) (27/03/2008 )

Il devient de plus en plus important d'utiliser la soumission des articles pour favoriser l'optimisation des pages Web et pour recevoir de liens de retour de qualité et pour acquérir un meilleur *PageRank*.

La soumission des articles signifie d'écrire des articles en relation avec le sujet du site et de les envoyer aux sites spécialisés dans la publication online. Ces articles offrent en règle générale une information supplémentaire sur le sujet en question et se dirigent à un group très ciblé.

Les articles soumis ne sont généralement pas payés et d'autres Webmasters peuvent les reproduire ou citer dans leurs sites. L'auteur obtient en contrepartie un lien vers son site. Les Webmasters qui publient ensuite le même article sont également obligés d'ajouter un tel lien. Vu que ces liens proviennent souvent des sites d'une bonne réputation, elles servent outre mesure dans l'optimisation des pages Web pour le référencement.

#### [soumission payante](#) (13/01/2008 )

On parle d'une soumission payante si quelqu'un doit payer une certaine somme pour figurer dans un moteur de recherche ou d'un annuaire. Cette politique concerne surtout les annuaires.

Pour qu'une soumission payante ait des résultats il est important de choisir la bonne catégorie et d'être sûr que l'annuaire placera l'annonce sur la première page. Il y a malheureusement des annuaires qui acceptent tellement d'annonces qu'ils ne peuvent plus garantir un bon positionnement.

En règle générale seulement un référenceur peut bien choisir les annuaires ou moteurs qui garantissent des résultats et lui seule est souvent capable de rédiger les descriptions nécessaires.

### [sous-titres](#) (18/02/2008 )

Les algorithmes des moteurs de recherche ont besoin des informations pour faire la différence entre un texte "normal", un titre et la partie du texte qui traite un sujet spécifique.

La meilleure solution est donc de structurer le contenu de la manière qu'on puisse utiliser des sous-titres. La valeur des mots clés des sous-titres se situe entre les titres de la page et les parties du texte marquées par `<b>`, `<strong>`, `<i>` ou `<em>`. Par cette technique on peut fournir aux moteurs des mots clés supplémentaires qui définissent seulement une certaine partie du texte d'une page.

Il est conseillé d'utiliser, si possible, le mot clé dominant aussi pour la rédaction des sous-titres mais de ne pas répéter les titres dominants. Les moteurs aiment des informations précises et supplémentaires mais ne pas le *keyword spamming*.

### [spam report](#) (05/05/2009 )

Le *spam report* est le formulaire de dénonciation mis en place par Google par lequel chaque personne peut dénoncer un site qui ne suit pas les consignes de Google.

Le *spam report* sert surtout à éliminer les sites qui emploient le *spamdexing* de la base de données de Google, qui ne s'occupe donc pas du besoin des internautes mais ciblent seulement leur positionnement avec des pages minimalistes.

Les algorithmes de Google ne peuvent pas faire la relation entre, par exemple, l'abus des référenceurs à l'aide de blogs et forums qu'ils ont créés avec le même contenu ou un contenu similaire pour occuper plusieurs places dans les premières pages des moteurs. Une dénonciation de cette *Black Hat SEO*, par contre, peut augmenter la fiabilité de Google.

### [spamdexing](#) (14/12/2007 )

Le mot *spamdexing* est une combinaison des mots *spam* et *indexing*. Ce terme, souvent appelé référencement abusive en français, désigne chaque technique et chaque essai de tromper les moteurs de recherche.

Parmi les méthodes de *spamdexing* compte l'emploi de mots invisibles, l'utilisation de mots sans relation avec le contenu (contenu web ou balise meta), le bourrage de mots clés, le *pagejacking* (détournement de pages), l'indexation excessive auprès de moteurs de recherche ou l'emploi abusive d'alias de domaine ou de sous domaine, pour nommer les méthodes les plus courantes.

Les robots sont capable à découvrir ces méthodes "interdits" et les punissent: Le *spamdexing* peut même bannir un site des moteurs de recherche et donc anéantir chaque essai d'optimisation.

### [splash page](#) (05/12/2007 )

Une *splash page*, également appelée *splash screen*, est une page d'entrée qui, au lieu du contenu propose qu'un lien d'entrée, la sélection de langues ou une animation Flash.

L'utilisation de *splash pages* fait partie des fautes les plus importantes qu'un Webmaster puisse commettre. De telles pages n'ont que des inconvénients et n'apporte rien aux internautes.

Les moteurs de recherche n'acceptent plus les splash pages, la règle des trois cliques n'est souvent plus applicable, le chargement de la page retarde l'accès au contenu et, statistiquement, les *splash pages* font perdre un nombre très important de visiteurs ou de clients.

Plus d'informations sur les *splash pages* se trouvent sous l'adresse : <http://www.mariposa.eu.com/lexique/lex/splpag.html>

### [splog](#) (10/02/2008 )

*Splog* se compose des mots *spam* et *blog*. Il s'agit de *blogs* qui sont exclusivement créés pour augmenter le *PageRank* est les résultats dans les moteurs de recherche de certaines pages en utilisant un nombre importants de liens qui mènent vers la page en question.

Vu que seulement les liens important, le contenu de ces *blogs* est assez restrictif, souvent gérer automatiquement par de scripts et, parfois avec un contenu volé. Il est donc logique que cette technique fait partie du *Black Hat SEO* et peut mener à un bannissement par les moteurs de recherche.

Il ne faut toutefois pas confondre les *splogs* avec des *blogs* de haute qualité qui peuvent effectivement favoriser le positionnement d'un site dans les moteurs de recherche.

### [statistiques](#) (05/02/2008 )

Seulement la consultation et l'analyse des statistiques permettent aux Webmaster et aux référenceurs de savoir si un site soit optimisé ou non. Seulement l'optimisation de la page d'accueil permet d'attirer un internaute et le faire entrer dans le site et l'optimisation du contenu permet finalement de le garder sur le site.

Il faut donc savoir analyser les statistiques correctement et parfois installer ses propres scripts de statistique pour tirer de conclusions. Les *hits* nous disent seulement combien d'éléments ont été vu sur une page ou un site, ce qui ne permet pas de savoir combien d'internautes ont visité le site. Le nombre des **visiteurs** nous informe combien d'ordinateurs ont consulté nos pages, mais ne nous dit pas s'il s'agit de la même personne qui revenait 30 ou 40 fois par jour. Le nombre des **visiteurs uniques** nous informe du nombre de différentes personnes qui ont, dans un certain espace de temps, consulté nos pages.

Mais il y a d'autres facteurs qui peuvent jouer un rôle, par exemple la provenance (le pays) des visiteurs, le temps de séjour sur certaines pages ou la connaissance de

quels autres sites ils sont venus. Seulement l'analyse complet des statistiques nous permet de savoir quelles parties d'un site doivent être modifié et donc optimisé

[stop character](#) (21/01/2008 )

On parle de *stop characters* si l'adresse d'une page HTML contient les caractères "&", "=" ou "?", qui se trouvent généralement dans de pages dynamiques gérées par PHP, ASP ou JAVA.

De telles caractères posent plusieurs problèmes pour les robots et donc pour l'optimisation des pages Web. Quelques moteurs refusent catégoriquement de référencer une telle page, d'autres n'acceptent que le premier niveau et s'arrêtent là. De tels *URL* font donc souvent partie du "Web invisible" et n'existent donc pas pour la majorité des internautes. Toutefois, si un lien cible une page qui utilise dans son adresse les *stop characters*, il y a forte chance qu'elle soit indexé mais sans que les moteurs acceptent les lies placés dans la page ciblée.

Il est donc important de ne pas forcer avec le contenu dynamique mais de l'utiliser très ponctuelle et à prendre les mesures, par exemple exploiter les liens de retour, pour que les robots soient capable d'indexer les pages dynamiques, ce qui est quelquefois un défi pour le Webmaster en question et nécessite des connaissances approfondies en référencement.

### [taille du texte](#) (27/01/2008 )

Le taille "standard" d'un texte correspond environ à la taille définit par la balise <font size="3">. La modification de cette taille permet donc de mettre en valeur une partie du texte ou d'exprimer qu'un texte soit de moindre importance.

Le choix de la taille joue donc un rôle important pour la mise en page du texte mais elle joue également un rôle dans l'optimisation des pages Web. Un robot qui détecte donc un nombre important de mots clé dans un texte définit, par exemple par <font size="1">, considère cette technique comme *spam* et l'effet souhaité est envers. A la limite, de telles pages ne seront même pas indexées par les moteurs de recherche.

Il est donc important de faire une différence entre mise en page et optimisation et d'éviter d'insérer les mots clés dans un texte qui s'affiche en miniature et presque illisible.

### [taux de rebond](#) (08/05/2009 )

Les algorithmes des moteurs de recherche incluent le taux de rebond dans leur positionnement des pages Web.

Le taux de rebond est le pourcentage de visiteurs qui quittent le site après la consultation d'une seule page. Cela signifie pour les moteurs de recherche qu'ils ne s'intéressent pas au contenu du site mais qu'ils atterrissent par mégarde au à cause de fausses promesses de la description sur elle. Le taux de rebond doit donc être le plus bas possible pour que la page puisse occuper une bonne position dans les moteurs.

Il est évident que l'algorithme n'évalue le taux de rebond pas sans regarder un paramètre supplémentaire. Un visiteur qui reste cinq minutes sur la page a donc une autre valeur qu'un autre qui quitte la page déjà après trois secondes.

### [texte des liens](#) (15/01/2008 )

Les moteurs de recherche cherchent dans le texte du lien (*link text* ou *anchor text* en Anglais) le ou les mots clés décisif pour la page ciblée. Ils utilisent donc ces mots pour savoir s'il y a une relation entre le lien et la page ciblée.

Il est donc primordiale en optimisation pour le référencement des pages Web de bien choisir ses textes des liens qui se trouvent entre les balises <a href=""> et </a>, même s'il s'agit de liens internes. Il est donc exclu d'utiliser de mots comme "ici", "cliquez ici" ou similaire pour inciter un internaute à utiliser un certain lien.

Les publicités qui se basent sur les textes des liens sont donc très recherchées et souvent aussi bien payé. La question reste donc seulement si les pages ciblées ne font pas partie d'une mauvais voisinage.

### [texte invisible](#) (30/11/2007 )

Pendant des années de nombreuses Webmasters ont essayé d'optimiser leurs pages en utilisant des textes invisibles, plus exactement en plaçant un nombre important de mots clés, invisible aux internautes, dans leurs pages.

Les méthodes varient et commencent avec un texte qui prend la même couleur que l'arrière-plan, de textes placés hors de l'écran ou en utilisant les CSS (feuilles de style) qui permettent, par exemple, de placer un texte en dessous des images.

Cette triche fonctionnait pendant plusieurs années. Mais les robots d'aujourd'hui sont capables de reconnaître ces textes, même si l'internaute ne se rend pas compte de ces textes. Comme les moteurs de recherche "punissent" chaque essai de triche il est donc logique qu'un site qui utilise des textes invisibles sera banni et n'entrera dans aucune base de données des moteurs de recherche.

### [thésaurus](#) (28/01/2008 )

Un thésaurus permet de proposer les synonymes pour les mots clés qui servent à l'optimisation des pages Web et qui sont utilisées dans les titres des pages Internet.

Vu que les internautes ne cherchent pas tous les mêmes mots, il est nécessaire de leur proposer dans la même page les alternatives possibles. Au lieu d'utiliser, par exemple, seulement le mot maison, il peut être utile d'utiliser également les mots bungalow, chalet, résidence, pavillon etc. pour augmenter la chance que les internautes trouvent une certaine page.

L'optimisation des pages signifie aussi de travailler le contenu et de lier les besoins et limites des moteurs de recherche avec le besoin des internautes.

### [titre de la page](#) (03/12/2007 )

Chaque page Internet a besoin d'un titre qui se trouve entre les balises `<title>` et `</title>`. Ce titre, qui s'affiche en haut du navigateur est le premier élément qu'un robot lit lorsqu'il scrute la page.

Ce titre doit être unique pour chaque page et se compose d'un minimum de 30 caractères et un maximum de 80 caractères. Comme les moteurs de recherche n'aiment pas des mots consécutifs, séparé par les commas, il est logique que chaque titre soit une phrase courte.

Le titre du *head* contient chaque mot clé qui définit la page, vu que les mots clé dans le titre compte en règle générale comme 4 ou 5 mots dans le texte. Un mot clé qui n'existe pas dans le titre ne peut donc pas être considéré comme mot clé. La rédaction du titre de la page est donc une tâche très important du Webmaster.

### [top-level page](#) (26/02/2008 )

La page d'accueil de chaque site s'appelle en règle générale `index.html`, `index.htm`, `index.php`, `default.asp`, etc. selon la configuration du serveur et/ou l'habitude du Webmaster. Cette première page d'un site est la *top-level page*.

De nombreux moteurs de recherche demande de soumettre seulement la *top-level page*, vu qu'avec une structuration correcte de trois niveaux les robots sont capables à trouver chaque autre page du site à partir de cette page. Si un Webmaster soumet dans un tel cas plusieurs pages de son site, il risque d'être puni par les moteurs et risque même de ne pas être indexé avec son site.

Pour favoriser le référencement il est donc conseillé de respecter les règles de la structuration, de créer une page *sitemap* et d'ajouter un lien vers le *sitemap* à partir de la *top-level page*.

### [traffic-death redesign](#) (07/03/2008 )

On parle d'un *traffic-death redesign* si un Webmaster remplace les pages d'un site existant par d'autres pages qui dispose d'un nouveau design mais qui ne disposent plus des mêmes noms, la page d'accueil à part.

Vu que des nombreuses pages du site avec les anciens noms ont été indexés par les moteurs de recherche, chaque internaute qui clique sur un des anciens *URL* reçoit donc une page d'erreur 404.

Cette action, sans précautions, du *traffic-death redesign* est fatale pour l'optimisation des pages Web et le référencement. Si le Webmaster garde les anciennes pages sur le serveur, les robots détectent du contenu double et refusent d'indexer les nouvelles pages, mais si le Webmaster ne garde que les nouvelles pages, l'erreur 404 ne fera pas seulement baisser le trafic mais fera en même temps baisser le *PageRank* et, sous certaines conditions peut faire disparaître le site des moteurs de recherche. Une des solutions que le Webmaster puisse prendre pour éviter le *traffic-death redesign* est donc la redirection 301.

### [TrustRank](#) (05/01/2008 )

Google a déposé ce nom en 2005, sans toutefois fournir une explication claire sur l'algorithme employé. La vraie signification de *TrustRank* est donc encore moins connue que le calcul du fameux *PageRank*.

Selon les divers auteurs et les expériences des référenceurs le *TrustRank* pourrait être définie comme un "indice de confiance" de la part des moteurs de recherche et surtout de la part de Google. Il s'agit probablement de l'ancienneté du site, de la qualité du contenu, de la quantité en pages etc. qui influencent le *TrustRank* de la même manière que les liens de retour et leur qualité influence le *PageRank*.

Selon certaines sources, le *TrustRank* serait aussi un algorithme qui détecterait plus facilement le *spamdexing* et serait donc destinée à afficher de résultats de confiance.

### [trusted feed](#) (01/05/2009 )

Le *trusted feed* est un référencement automatisé à l'aide du langage *xml* et nécessite un contrat avec le moteur de recherche en question.

Le *trusted feed* a l'avantage qu'un nombre très importantes de pages, surtout de pages difficiles à référencer comme les bases de données ou les sites an flash puisse être indexés.

Le désavantage du *trusted feed* est le fait qu'il s'agit d'un service payante des moteurs de recherche et que Google ne l'accepte pas mais considère cette technique du référencement comme *spamdexing*.

Avant de souscrire un service de *trusted feed* il est donc important à évaluer les avantages et les désavantages du service.

### [typosquattage](#) (26/12/2007 )

Le typosquattage (*typosquatting* en Anglais) se base sur le fait qu'un nombre assez élevé des internautes fait de temps en temps de fautes de frappe lorsqu'ils tapent une adresse Internet dans le navigateur. Le typosquatteur utilise cette faille pour attirer ces internautes vers son propre site.

Typosquattage est donc le fait d'acheter les domaines qui se distinguent que légèrement d'un domaine connu qui vend en règle générale le même produit. L'internaute ne se rend souvent même pas compte qu'il est atterrit dans un site similaire qui copie souvent même la mise en page du site cherché. Presque chaque domaine d'une réputation élevé est, d'une manière ou d'une autre victime de typosquatteurs.

La seule possibilité d'éviter de tels abus est donc d'acheter dès le début tous les noms de domaines similaires et de rediriger les Internauteurs vers le site originaire.

Plus d'information sur le typosquattage se trouvent sous l'adresse <http://www.mariposa.eu.com/lexique/lex/typsqu.html>

### Valdateur (28/11/2007 )

Une page Web s'affiche seulement correctement dans les divers navigateurs et les divers systèmes d'exploitation si le code des pages est conforme à la norme du W3C. Même les robots des moteurs de recherche acceptent des pages valides plus facilement dans leurs bases de données que les pages dont le code est "bricolé".

Pour vérifier son code chaque Webmaster peut utiliser un des validateurs qui se trouvent sur Internet. La majorité de ces validateurs permet la vérification des pages offline pour que le Webmaster puisse corriger ses fautes avant de mettre son site online.

La majorité des validateurs affiche chaque erreur du code avec un petit commentaire qui facilite la correction et aide à comprendre le problème. Un validateur en français se trouve à l'adresse suivante:

<http://www.validome.org/lang/fr>

Un validateur aide aussi aux autodidactes de se perfectionner dans la rédaction de codes et garantit finalement de publier de sites professionnels.

### veille concurrentielle (04/05/2009 )

Le terme veille concurrentielle s'applique surtout au web marketing et désigne la surveillance des actions, des méthodes publicitaires ainsi que de produits et services de la concurrence.

Dans l'optimisation des pages Web pour le référencement la veille concurrentielle est surtout la surveillance de sites concurrentiels et de leur positionnement dans les moteurs de recherche.

Comme Google et les autres moteurs de recherche changent leurs algorithmes dans un rythme irrégulier la veille concurrentielle est souvent la seule possibilité pour découvrir ces changements et pour pouvoir adapter ses propres méthodes de l'optimisation aux besoins des moteurs de recherche.

### VEO (Visitor Enhanced Optimization) (05/03/2008 )

En suivant les règles de la *SEO (Search Engin Optimization)* un Webmaster ou un référenceur optimise les pages pour qu'elles obtiennent la meilleur position dans les moteurs de recherche pour que les internautes les trouvent facilement.

Mais trouver un site ne signifie pas d'apprécier le site, de le mettre dans les favoris ou de revenir. Pour satisfaire les visiteurs et pour que le site aura vraiment du succès, les pages doivent également être adapté et optimisées pour les visiteurs. On parle ici de la *VEO*, de la *Visitor Enhanced Optimization*.

La *VEO* utilise d'abord le contenu pour que les pages soient correctement indexées dans les moteurs et non pas avec des mots clés qui n'ont pas beaucoup à faire avec le vrai contenu du site. Mais la *VEO* utilise aussi les *RSS feeds*, les *Podcasts*, les *blogs*, les *vidéos* etc., donc des techniques qui attirent le public bien ciblé.

### visiteur (25/02/2008 )

Le **visiteur** est souvent confondu avec le **visiteur unique** même si leur nombre n'a rien en commun et le nombre des visiteurs ne dit pas beaucoup sur la quantité des internautes qui arrivent la première fois sur un site.

Le mot visiteur est utilisé pour chaque entrée d'un visiteur dans le *logfile*, n'importe si la même personne visite une page une ou 30 fois pendant, par exemple, une journée. Le chiffre augmente donc avec chaque visite et fournit, sans analyse approfondie, aucune information au Webmaster.

Seulement ensemble avec le nombre des visiteurs uniques, des internautes donc, qui apparaissent dans une certaine espace de temps la première fois, un Webmaster puisse tirer certaines conclusions qui peuvent, ensuite, jouer un rôle pour l'optimisation des pages Web.

### visiteur unique (07/12/2007 )

Pour connaître la réussite d'un site on distingue les visiteurs uniques des visiteurs. Seulement le chiffre des visiteurs uniques, en règle générale comptés par mois, est un chiffre valable vu qu'ici chaque visiteur qui dispose de le même *IP* n'est compté qu'une seule fois.

Même si un internaute revient alors dix fois pour regarder, peut-être, d'autres pages ou regarde la même page encore une fois, il ne compte qu'une seule fois. La vérification se passe par le *IP* et un cookie qui sera enregistré dans l'ordinateur du visiteur.

Les compteurs simples fournissent dans la plupart de temps que les visiteurs et enregistrent la même personne qui vient chaque jour dix fois aussi dix fois ce qui falsifie un résultat. Une personne qui, par exemple, regarde le même livre dans la même page 30 fois ne l'achète qu'une seule fois et seulement la première visite comme visiteur unique peut donc fournir un résultat comparatif.

### [web copywriting](#) (11/03/2008 )

Avant que l'analyse du contenu fasse partie des algorithmes des moteurs de recherche, le texte des pages était rédigé soit par l'entreprise qui employait le Webmaster, soit par le Webmaster lui-même.

Pour actuellement acquérir une bonne position dans les moteurs de recherche "écrire" un texte ne suffit plus. Les textes doivent être rédigés pour les internautes mais en même temps aussi pour les moteurs de recherche. La rédaction de ces textes optimisés s'appelle *web copywriting*.

Le *web copywriter* analyse donc le texte de l'entreprise pour le modifier de la manière qu'il contient les mots clés nécessaires pour le référencement et pour attirer en même temps les internautes pour augmenter le chiffre d'affaires. Le *web copywriting* fait aujourd'hui partie des références.

### [web profond](#) (13/03/2008 )

L'expression web profond est le synonyme de web invisible (invisible web en Anglais) et désigne tous les fichiers que les robots ne peuvent pas scruter et que les moteurs de recherche n'ajoutent par conséquent pas à leurs bases de données.

Selon les moteurs de recherche le web invisible comporte des fichiers en flash, des pages dynamiques non conformes aux exigences des moteurs, des fichiers son, des vidéos etc. qui, toutefois, peuvent jouer un rôle important pour certains sites et comportent parfois des informations essentielles.

Même si plusieurs moteurs de recherche commencent à adapter leurs algorithmes pour pouvoir indexer ces fichiers, leur optimisation pose souvent de nombreux problèmes. Mais une page non indexée est souvent introuvable par les internautes et les Webmasters doivent donc être conscients que le web invisible peut bloquer l'évolution normale d'un site et qu'il faut penser à cet effet lors de la structuration du site.

Plus de détails se trouvent dans l'article sur le [web invisible](#)

### [Web invisible](#) (30/04/2009 )

Le Web invisible désigne toutes les pages Web qui ne se trouvent pas dans l'index des moteurs de recherche. Il y a trois groupes principaux de pages qui font partie du Web invisible.

Il s'agit d'abord de toutes les pages qui ne sont jamais soumises aux moteurs de recherche vu qu'il s'agit des pages à accès limité, destinées à un réseau Intranet ou tout simplement bloquées par le fichier robots.txt.

Le deuxième groupe se compose des pages que les robots ne peuvent pas accéder à cause de scripts dynamiques, d'une navigation par JavaScript, l'emploi non optimisé du Flash et d'autres fautes de programmation.

Le Web invisible se compose finalement aussi de toutes les pages que les robots refusent à indexer et de sites qui se trouvent sur la liste noire des moteurs de recherche.

Le Web invisible est traité dans la page  
[Le Web invisible ou le Web profond](#)

[White Hat SEO](#) (29/01/2008 )

On parle du *White Hat SEO* (*Search Engine Optimization*) si le référenceur utilise exclusivement de méthodes honnêtes de l'optimisation et accepte les directives des moteurs de recherche sans enfreindre la moindre de leurs règles.

Une page selon la *White Hat SEO* propose donc exclusivement du contenu valable et complet et les pages sont seulement optimisées pour de mots clés en relation avec le contenu de la page sans que la densité de ces mots soit forcé et exagérée.

Utiliser la *White Hat SEO* est donc la seule possibilité d'acquérir un bon positionnement dans les moteurs de recherche et de ne pas risquer le bannissement par les moteurs de recherche.

[Xenu](#) (17/03/2008 )

Xenu est un logiciel gratuit installé dans l'ordinateur qui cherche les liens cassés d'un site et établit une analyse qui aide aux référenceurs et Webmasters de corriger ces problèmes.

Chaque lien cassé, même s'il s'agit des images ou des fichiers autre que html et n'importe s'il s'agit des liens internes ou des liens externes, produit l'erreur 404 pour les moteurs de recherche et peut donc poser des problèmes de référencement.

Il est donc obligatoire de tester les liens de chaque site avant la publication et ensuite encore une fois les liens du site publié. Xenu fait partie des logiciels qui permettent de tester des sites (*link-checking program* en Anglais) assez complexes et comporte que peu de bogues.

Le logiciel peut être chargé à partir du lien <http://home.snafu.de/tilman/xenulink.html>.

La liste présentée ici est quotidiennement actualisée et nous conseillons donc à visiter le blog aussi souvent que possible pour apprendre la terminologie des Webmasters et référenceurs.

<http://seo-optimisation.over-blog.net/>

Notre documentation en PDF sera mise à jour dans un rythme irrégulier et la version la plus récente sera toujours accessible par les liens dans les pages du blog ainsi que les pages du site

<http://www.seo-reference.net/>

Si vous avez des questions ou commentaires, n'hésitez jamais à laisser un commentaire dans les pages du blog.