CULTURE DU WEB

Référencement

Ce qu'il faut savoir

Pierre **Barthélemy**
Consultant SEO

Smile
OPEN SOURCE SOLUTIONS

www.smile.fr • +33 (0)1 41 40 11 00 • contact@smile.fr
www.smile-oss.com • blog.smile.fr • twitter: @GroupeSmile

PREAMBULE

Smile

Smile est une **société d'ingénieurs experts** dans la mise en œuvre de **solutions open source** et l'intégration de systèmes appuyés sur l'open source. Smile est membre de l'**APRIL**, l'association pour la promotion et la défense du logiciel libre, du **PLOSS** – le réseau des entreprises du Logiciel Libre en Ile-de-France et du **CNLL – le conseil national du logiciel libre**.

Smile compte plus de 600 collaborateurs dans le monde, dont près de 500 en France (janvier 2012), ce qui en fait *le premier intégrateur français et européen de solutions open source.*

Depuis 2000, environ, **Smile mène une action active de veille technologique** qui lui permet de découvrir les produits les plus prometteurs de l'open source, de les qualifier et de les évaluer, de manière à proposer à ses clients les produits les plus aboutis, les plus robustes et les plus pérennes.

Cette démarche a donné lieu à **toute une gamme de *livres blancs*** couvrant différents domaines d'application. La gestion de contenus (2004), les portails (2005), la business intelligence (2006), la virtualisation (2007), la gestion électronique de documents (2008), les PGIs/ERPs (2008), les VPN open source (2009), les Firewall et Contrôle de flux (2009), les Middleware orientés messages (2009), l'ecommerce et les Réseaux Sociaux d'Entreprise (2010) et plus récemment, le Guide de l'open source et NoSQL (2011). Chacun de **ces ouvrages présente une sélection des meilleures solutions open source** dans le domaine considéré, leurs qualités respectives, ainsi que des retours d'expérience opérationnels.

Au fur et à mesure que des solutions open source solides gagnent de nouveaux domaines, Smile sera présent pour proposer à ses clients d'en bénéficier sans risque. Smile apparaît dans le paysage informatique français comme **le prestataire intégrateur de choix** pour **accompagner** les plus grandes entreprises dans l'adoption des meilleures solutions open source.

Ces dernières années, Smile a également étendu la gamme des services proposés. Depuis 2005, un département consulting accompagne nos clients, tant dans les phases d'avant-projet, en recherche de solutions, qu'en accompagnement de projet. Depuis 2000, Smile dispose d'un studio graphique, devenu en 2007 Smile Digital – agence interactive, proposant outre la création graphique, une expertise e-marketing, référencement, éditoriale, et interfaces riches. Smile dispose aussi d'une agence spécialisée dans la TMA (support et l'exploitation des applications) et d'un centre de formation complet, Smile Training. **Enfin, Smile est implanté à Paris, Lille, Lyon, Grenoble, Nantes, Bordeaux, Marseille, et Montpellier. Et présent également en Espagne, en Suisse, au Benelux, en Ukraine et au Maroc**.

Quelques références de Smile

SMILE est fier d'avoir contribué, au fil des années, aux plus grandes réalisations Web françaises et européennes. Vous trouvez ci-dessous quelques clients nous ayant adressé leur confiance.

Sites Internet

EMI Music, Salon de l'Agriculture, Mazars, Areva, Société Générale, Gîtes de France, Patrice Pichet, Groupama, Eco-Emballage, CFnews, CEA, Prisma Pub, Véolia, NRJ, JCDecaux, 01 Informatique, Spie, PSA, Boiron, Larousse, Dassault Systèmes, Action Contre la Faim, BNP Paribas, Air Pays de Loire, Forum des Images, IFP, BHV, ZeMedical, Gallimard, Cheval Mag, Afssaps, Beneteau, Carrefour, AG2R La Mondiale, Groupe Bayard, Association de la Prévention Routière, Secours Catholique, Canson, Bouygues Telecom, CNIL…

Portails, Intranets et Systèmes d'Information

HEC, Bouygues Telecom, Prisma, Veolia, Arjowiggins, INA, Primagaz, Croix Rouge, Eurosport, Invivo, Faceo, Château de Versailles, Eurosport, Ipsos, VSC Technologies, Sanef, Explorimmo, Bureau Veritas, Région Centre, Dassault Systèmes, Fondation d'Auteuil, INRA, Gaz Electricité de Grenoble, Ville de Niort, Ministère de la Culture, PagesJaunes Annonces…

E-Commerce

Krys, La Halle, Gibert Joseph, De Dietrich, Adenclassifieds, Macif, Furet du Nord, Gîtes de France, Camif Collectivité, GPdis, Projectif, ETS, Bain & Spa, Yves Rocher, Bouygues Immobilier, Nestlé, Stanhome, AVF Périmédical, CCI, Pompiers de France, Commissariat à l'Energie Atomique, Snowleader, Darjeeling…

ERP et Décisionnel

Veolia, La Poste, Christian Louboutin, Eveha, Sun'R, Home Ciné Solutions, Pub Audit, Effia, France 24, Publicis, iCasque, Nomadvantage, Gets, Nouvelles Frontières, Anevia, Jus de Fruits de Mooréa, Espace Loggia, Bureau Veritas, Skyrock, Lafarge, Cadremploi, Meilleurmobile.com, Groupe Vinci, IEDOM (Banque de France), Carrefour, Jardiland, Trésorerie Générale du Maroc, Ville de Genève, ESCP, Sofia, Faiveley Transport, INRA, Deloitte, Yves Rocher, ETS, DGAC, Generalitat de Catalunya, Gilbert Joseph, Perouse Médical…

Gestion documentaire

Primagaz, UCFF, Apave, Géoservices, Renault F1 Team, INRIA, CIDJ, SNCD, Ecureuil Gestion, Région Centre, Serimax, Véolia Propreté, NetasQ, Generali, Bureau Veritas, Alstom Power Services, Mazars, SNCF, HEC…

Infrastructure et Hébergement

Agence Nationale pour les Chèques Vacances, Pierre Audoin Consultants, Rexel, Motor Presse, OSEO, Sport24, Eco-Emballage, Institut Mutualiste Montsouris, ETS, Ionis, Osmoz, SIDEL, Atel Hotels, Cadremploi, SETRAG, Institut Français du Pétrole, Mutualité Française, Ministère de l'écologie et du développement durable…

Consulter nos références, en ligne, à l'adresse : http://www.smile.fr/clients.

Ce livre blanc

C'est à dessein que ce livre blanc ne s'intitule pas « Référencement – secrets d'experts » : son but est bien de présenter les principes fondamentaux du référencement, tant du point de vue des techniques sous-jacentes que des démarches visant à l'optimiser.

Avant de faire appel à un prestataire spécialisé dans l'optimisation du référencement naturel (on parlera de SEO tout au long de ce document), il conviendrait que chaque responsable de site connaisse ce minimum que nous présentons ici.

Il y a beaucoup d'idées fausses concernant le SEO. Par exemple, qu'il suffit de payer un bon prestataire pour que vos sites/portails soient automatiquement dans les premières pages de Google, ou qu'il suffit de travailler son SEO pendant le lancement de son site puis de ne plus rien toucher dans les mois/années qui suivent.

Un accompagnement SEO est un ensemble d'étapes à respecter, avant la mise en ligne (ou refonte), pendant le travail de conception et obligatoirement après tout le travail de mise en place (suivi, optimisations…). Tout cela afin que votre projet puisse récolter le plus de visibilité possible, avec un trafic le plus qualifié possible.

La première chose que nous aimerions transmettre dans ce recueil est que le SEO n'est pas une sorte de sorcellerie aux recettes cryptiques et mystérieuses, mais un processus tout à fait raisonné, qui consiste plutôt à *mettre en avant* la pertinence réelle de votre site plutôt qu'à *faire croire* à une pertinence qu'il n'aurait pas.

N'hésitez pas à nous transmettre votre avis à l'adresse : contact@smile.fr

SOMMAIRE

LES BASES

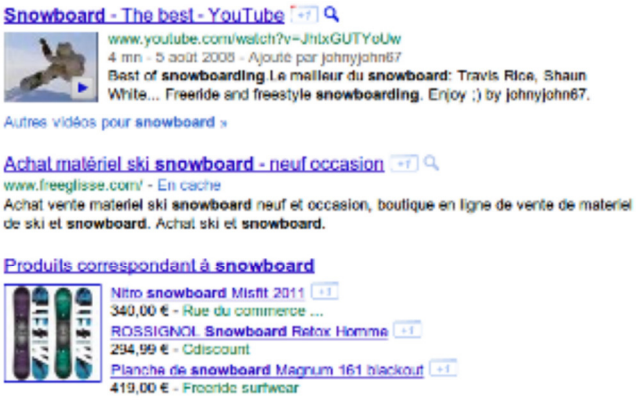
Le service aux internautes

Mettons-nous un peu à la place d'un outil de recherche. Son objectif est de servir ses visiteurs, en les aidants à trouver rapidement l'information qu'ils recherchent. Donc de présenter les milliards de résultats de recherche dans l'ordre de ***pertinence***. Bien sûr la notion de *pertinence* est très subjective, et la tâche du moteur est précisément de quantifier cette pertinence d'une manière qui corresponde *le plus souvent* aux attentes des internautes.

Cette pertinence s'est aujourd'hui étendue. En effet, ce qui est présenté aux internautes n'est plus uniquement des résultats affichant les « meilleurs sites web », mais aussi des résultats liés à la fameuse « recherche universelle [1] ». Vous trouverez donc des résultats de recherche (SERP[2]) complémentaires comme des images, des vidéos, des cartes (adresses), des produits, les réseaux sociaux, des liens sponsorisés...

Par exemple, si vous tapez la requête « SNOWBOARD » dans un moteur de recherche, il vous proposera toutes sortes de résultats en termes de contenus. En effet, s'il existe des sites web pertinents sur cette requête, il existe aussi d'autres types de contenus qui font peut-être partis de votre recherche initiale ? En tapant « snowboard », c'est peut-être le produit que vous cherchez plutôt que des actualités sur le sport ?

[1] **Recherche universelle** : les moteurs de recherche affichent de plus en plus au sein de leurs pages de résultats des éléments qui ne sont pas uniquement des pages Internet standard, mais également des images, des vidéos, plans et fils d'actualité.
[2] **SERP** : Search Engine Result Page. Ce qui signifie littéralement : résultats affichés par les moteurs de recherche.



Il y a quantité de sociétés et sites web qui proposent de vendre des snowboards, mais aussi des blogs, des sites d'infos (produits ou plus généralistes), la fédération de snowboard … et tous ces sites doivent se distinguer parmi les 119 millions de résultats (nous reviendrons plus loin sur cette notion de somme des résultats).

Google a aussi étendu ses SERP en les personnalisant de plus en plus en se basant sur vos habitudes et vos contacts. Si vous êtes connectés à votre compte Google, selon les liens partagés par vos contacts ou vous-même via Google+ ou le bouton +1, vos résultats seront sûrement différents de ceux de votre voisin. Et ce qui est vrai avec Google l'est aussi avec les autres principaux outils de recherche actuels (Bing, Yahoo dans les pays occidentaux, Baidu en Chine, Yandex en Russie…). Même si bien sûr, chaque marché à ses spécificités et un référenceur doit prendre en compte les habitudes de recherche spécifiques des internautes locaux. Sans compter que pour un même outil, les options de recherche seront différentes d'un pays à l'autre. Par exemple, les fonctionnalités sur Google ne sont pas les mêmes en France et en Suisse (pas de Google+1).

Le travail du moteur de recherche est de parvenir à distinguer la Fédération de snowboard qui ne parle que de ça et les pages éventuellement consacrées au snowboard sur un site plus généraliste comme skipass.com.

Ce travail doit obligatoirement être totalement automatisé, puisqu'il porte sur des milliards de pages : il est hors de question qu'un intervenant humain passe 15 secondes à évaluer la pertinence de chaque page.

Enfin, la tâche du moteur de recherche est rendue plus difficile encore par le fait que les gestionnaires de sites ont pour objectif avoué de le tromper pour obtenir les meilleurs résultats ! Le moteur veut établir de manière automatique la vraie pertinence de chaque page, le gestionnaire du site veut faire croire que son site est plus pertinent qu'il ne l'est réellement.

On a donc une vraie opposition, une guerre interminable, entre moteurs et webmasters. Si le moteur se laisse tromper par les sites, il perd sa crédibilité. Il lui faut donc trouver toujours plus d'algorithmes qui ne pourront être abusés par les webmasters.

Cela a fait la réussite de Google, mais cela a été aussi un bénéfice pour l'Internet en général, en redonnant sa place à la vraie pertinence.

La domination de Google

Ce n'est un secret pour personne, Google domine largement ses concurrents en Europe dans le domaine de la recherche sur Internet.

71,3% des français se connectent à Internet (Février 2011)
9 sur 10 effectuent des recherches
90% d'entre eux utilisent **Google**

Yahoo et Bing (Microsoft) étaient les principaux moteurs pouvant titiller le géant de Mountain View, ils viennent pourtant de fusionner leurs résultats de recherche (Yahoo intègre depuis août 2011 les résultats de Bing). Et ne parlons pas des outils français comme Exalead, Orange ou autres. Ils sont anecdotiques vis-à-vis de Google mais intéressants. Cependant, il faut garder à l'esprit que vous n'allez pas optimiser un site pour apparaître uniquement sur Google, les bonnes pratiques à respecter étant sensiblement les mêmes pour tous les outils de recherche, vous allez aussi vous positionner sur ces autres outils.

Attention, cette domination n'est pas aussi forte partout dans le monde. Aux US, Bing reste une valeur sure, en Asie Google est largement derrière des outils comme Baidu (en Chine, ce moteur représente 70% de part de marché). Même constat dans les pays de l'Est avec Yandex par exemple. Votre stratégie SEO sera donc différente sur ces pays qui ont des règles différentes d'indexation.

Référencement Pourquoi ?

Les internautes accèdent à un site de trois manières : (a) en tapant directement l'URL ou en la sélectionnant dans un signet (bookmark), (b) en suivant un lien depuis un autre site/blog/forum/réseaux sociaux, et (c) par une recherche sur un outil de recherche.

Pour trouver un site qu'ils ne connaissaient pas auparavant, seules restent les voies (b) et (c), et différentes études estiment que le moteur de recherche est la manière utilisée dans plus de 80% des cas pour découvrir un site que l'on ne connaissait pas.

Lorsqu'ils utilisent un moteur de recherche, il est évident que les internautes ne peuvent parcourir plus de quelques pages de réponse, et qu'en conséquence seuls les sites figurant sur les premières pages seront visitées.

Il est donc d'une importance primordiale de figurer en bonne place dans les résultats de recherches si l'on veut attirer des visiteurs par ce canal. Tout le monde le sait, et c'est la raison pour laquelle le SEO est devenu une spécialité à part entière dans le monde des technologies Internet.

Les anglophones appellent cette activité Search Engine Optimization, c'est-à-dire optimisation pour les moteurs de recherche. Ce qui est plus explicite finalement car il ne s'agit pas d'être référencé naturellement, mais bien d'optimiser au mieux le référencement de votre site en mettant en place tout un panel d'interventions. Il n'y a finalement pas grand-chose de naturel dans cette optique d'optimisation !

Etant donné les milliards de pages indexées par un moteur de recherche, il est naturellement difficile d'espérer figurer sur la première page pour des recherches larges, comme par exemple « télévision » pour un vendeur de télévisions. Les sites de marques les plus influents y sont présents, et travaillent d'arrache pied pour rester sur cette page avec énormément de visibilité. Pour la plupart des sites, il vaut mieux se fixer des objectifs moins ambitieux, et viser un bon rang pour des recherches plus ciblées, sur des couples ou des triplets de mots. Cette méthode a le mérite de cibler une audience plus qualifiée et intéressée par votre contenu. Attention, nous ne parlons pas ici de « longue traine » (nous expliquerons ce concept un peu plus loin) mais bien de requêtes précises et stratégiques. Et la concurrence n'est pas une excuse pour ne pas tenter de se positionner sur des mots clés très concurrentiels, il faut cependant y mettre les moyens pour espérer pouvoir y figurer.

De plus en plus, les internautes chevronnés savent qu'une recherche trop vague ne sera pas utile, et ils saisissent dès le début une petite liste de mots-clés. Ainsi, les recherches portant sur 3 mots seraient passées de 17% en 2005 à 42% en 2010 (source Ad'Oc) !

Un jeu sans fin

Figurer en première page est un peu un jeu de dupe.

Vos 10 principaux concurrents ont payé une agence SEO des milliers d'euros chacun pour être en première page sur quelques requêtes données et ils y sont. Vous payez vous-

mêmes votre dû et vous voilà en première page, éjectant l'un de vos concurrents en deuxième page. Furieux, celui-ci appelle l'agence chargée de son référencement, qui lui refacture un peu, bricole un peu plus ses pages/backlinks[3], et le ramène victorieusement en première page. Vous appelez votre agence, dépensez encore un peu d'argent, et ainsi de suite…

L'amélioration du référencement est ce qu'on appelle un jeu à somme nulle. Plus précisément, c'est la somme des gains de position qui est nulle, pas les sommes dépensées. Mais le retour sur investissement d'un travail rigoureux sur le référencement naturel est encore aujourd'hui imbattable.

Et pour autant, personne ne peut se permettre d'abandonner le combat. Chaque jour les positions changent, chaque jour vos concurrents optimisent leur contenu et leur stratégie. C'est pourquoi un suivi dans le temps rigoureux d'indicateurs clés liés au SEO est indispensable. Nous verrons plus loin que des outils existent pour cela.

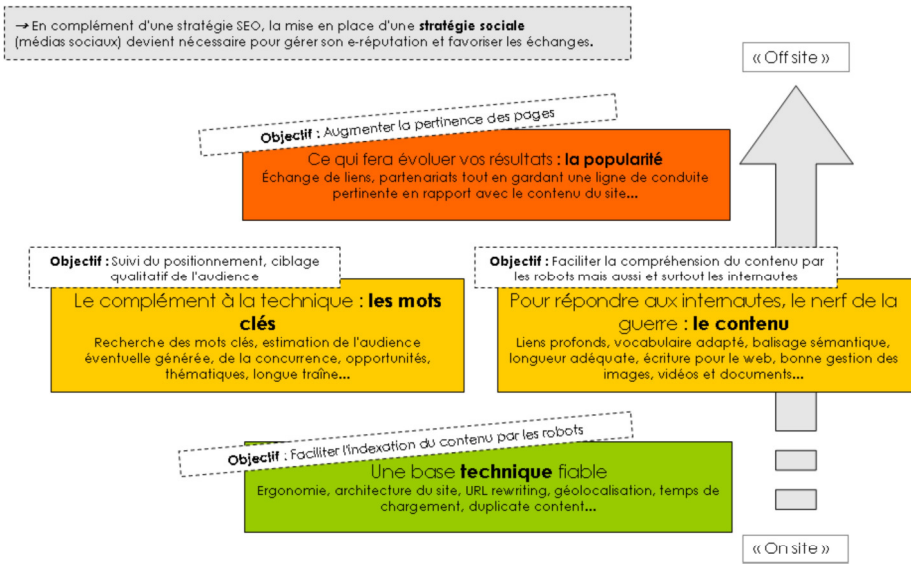
[3] **Backlinks** : liens externes à votre site web, renvoyant vers votre contenu.

La pyramide du SEO

La « pyramide du SEO » est régulièrement évoquée dans les interventions. Cette pyramide représente les différentes étapes nécessaires à un bon référencement naturel.

➔ L'objectif : le renforcement de votre visibilité.

Nous partons d'indicateurs « onsite », c'est-à-dire des leviers qu'il est possible d'activer directement sur les développements. Pour arriver ensuite plus ou moins directement aux indicateurs « offsite », soit les optimisations nécessaires en dehors du travail technique sur votre stratégie Web.



La base est bien sur « l'indexabilité », c'est à dire la compatibilité avec les mécanismes utilisés par le moteur pour parcourir et indexer les sites. Une base technique propre et respectant la plupart des indicateurs vous assurera d'une indexation optimale dans les outils de recherche. Et à terme, de meilleures positions.

Nous pouvons citer ici le choix d'arborescence et de structure de votre site, les optimisations « On page », la gestion de vos liens internes, l'accessibilité (multi-support, balisage sémantique), la confiance dans votre contenu (autorité du domaine, historique des contenus)…

Outre cette base technique, il va vous falloir ensuite travailler votre contenu.

C'est le nerf de la guerre, sans contenu, rien à indexer dans les outils de recherche et aucune pertinence sur votre univers de mots clés.

Que vous ayez un site corporate, d'informations ou de vente en ligne, votre objectif devra rester le même : produire du contenu !

En effet, les outils de recherche et Google en tête n'aiment pas les pages web avec peu de texte ou de contenu pertinent pour l'internaute.

Non seulement ce contenu doit être pertinent et abondant, mais il doit aussi être construit en suivant les bonnes pratiques et en s'assurant que vos titres par exemple soient optimisés pour le référencement naturel. En effet, nous voyons trop souvent des titres d'articles étudiés pour le print, et donc très marketing, mais pas du tout adapté au web et aux recherches des internautes. Nous entrons ici dans la notion parallèle au contenu, les mots clés.

Ce choix de mots clés pour votre contenu est stratégique. Cela va avoir un impact sur l'ensemble de votre site : arborescence, contenu des pages, balises META,… Tout cela en se posant une question simple : qu'est ce que les internautes pourraient utiliser comme requête pour trouver votre site ?

Par exemple :

Préférez : Mode d'emploi <nom de produit> en 5 étapes. A cela : A la découverte d'un outil merveilleux.

Afin de rendre visible tout ce travail préparatoire, voici venu l'étape qui fera la différence par rapport à vos concurrents : **la popularité**.

En effet, si votre contenu n'est pas repris, cité ou rendu populaire par n'importe quel moyen auprès des internautes, celui-ci ne sera pas (ou peu) visible sur les outils de recherche (seule solution alternative, les adwords).

Sur des mots clés peu compétitifs, cette notion de backlinks ne sera pas déterminante, peu de liens pourraient vous permettre d'arriver en tête des résultats. Mais sur des requêtes très populaires (auto, mutuelle, crédit…), vous aurez besoin d'un nombre très important de liens pour pouvoir rendre votre contenu pertinent pour les outils de recherche par rapport à vos concurrents.

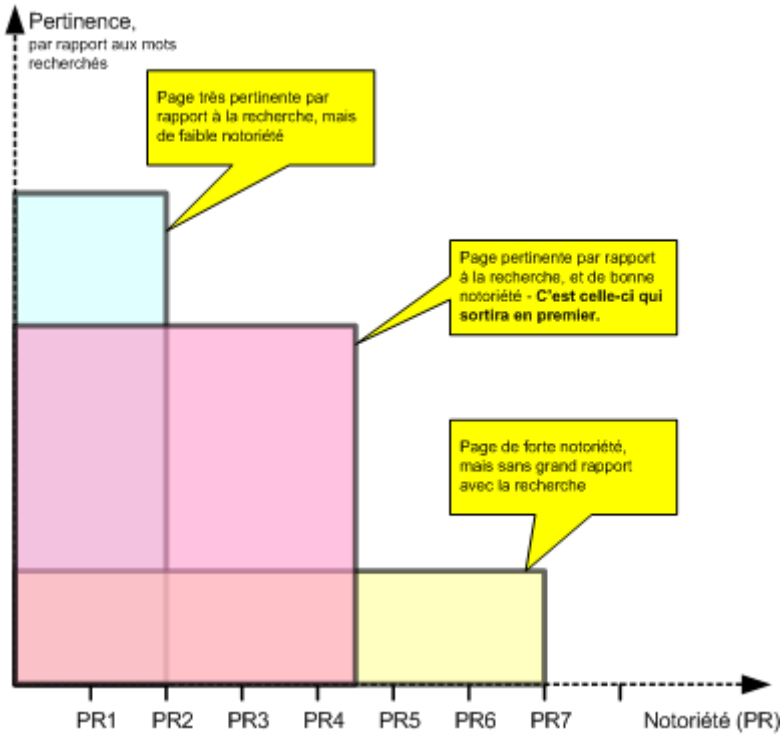
Si nous résumons, l'ordre de présentation des résultats d'une recherche dépend de deux principaux facteurs :

▪ La **notoriété**, qui est une valeur intrinsèque de la page, une mesure qui associe qualité des liens entrants (Backlinks) et leur quantité.

▪ La **pertinence** de la page **pour les mots recherchés**, c'est à dire sa plus ou moins grande correspondance avec ce que recherche l'internaute.

Ces deux notions sont totalement transverses, indépendantes l'une de l'autre. Ce sont les deux notions clés qui interviennent dans le référencement, et un chapitre spécifique est consacré à chacune d'elles.

La manière dont ces deux facteurs sont combinés pour produire l'ordre de tri des résultats n'est pas vraiment connue. Ces algorithmes sont protégés comme des secrets d'états par ces sociétés et les référenceurs ne peuvent se baser que sur leur vécu et sur la mise en place de tests tout au long de l'année.



La figure ci-dessus traduit la formule : ordre de sortie = pertinence x notoriété

Cependant, nous pouvons estimer ceci :

▪ A contenu égal, les pages de notoriété plus élevée viennent en tête ;

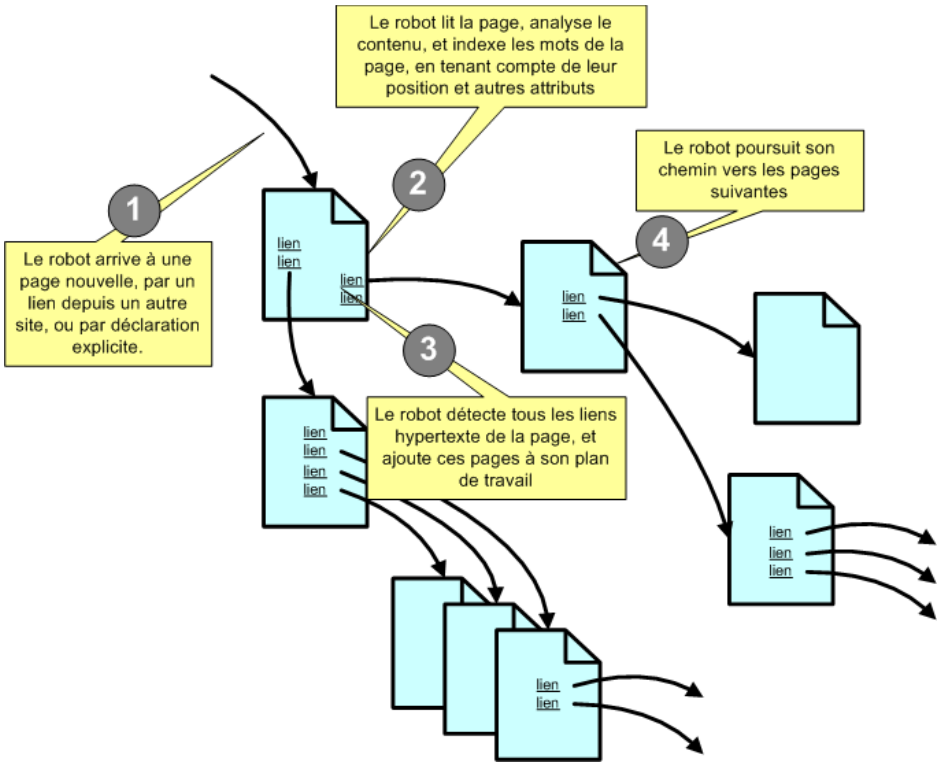
▪ A notoriété égale, les pages les plus pertinentes viennent en tête.

INDEXATION

Le Crawler

A la base du référencement il y a le *robot d'indexation,* appelé encore *Crawler* (« celui qui avance en rampant »). C'est un programme, qui se comporte comme un internaute qui suivrait tous les liens qu'il rencontre. Il lit une page, analyse le contenu et indexe les mots rencontrés, puis suit tous les liens de cette page pour lire d'autres pages. Et ainsi de suite.

Normalement, le crawler devrait découvrir ainsi pratiquement tous les sites, puisqu'il suffit d'un lien vers une page de votre site pour qu'il « entre » et parcourt alors l'ensemble des pages en suivant ainsi tous les liens.



Principe du Crawler

Mais si votre site est tout neuf, aucun autre site n'a encore de lien vers le vôtre. Il est toujours possible de signaler explicitement l'existence d'un nouveau site sur Google. Cependant, il est plus efficace de signaler cette existence par la publication d'articles ou de créer ces liens vers votre site grâce à vos partenaires/votre groupe. Le moteur ne garantit pas qu'il viendra le visiter et l'indexer rapidement, mais sous quelques jours voire maintenant heures, il le fera. **C'est donc bien sûr la première étape pour indexer un site que de signaler son existence aux principaux moteurs de recherche du web.**

La fréquence de visite du crawler n'obéit pas à des règles publiées. Elle dépend du moins du taux de mise à jour du site : si le crawler voit que les contenus du site sont modifiés fréquemment, il revient fréquemment. La fréquence de visite dépend certainement aussi du *Page Rank*[4], de la notoriété du site : le site de Microsoft sera indexé plus fréquemment que celui des pêcheurs de la Marne. Pour autant, il serait tout à fait inutile de chercher à être indexé plus souvent, puisque cela ne donnerait en rien un meilleur référencement.

Les limites du Crawler

Le minimum requis pour que toutes les pages d'un site soient référencées est qu'il soit *crawlable*, c'est-à-dire qu'il ne présente pas d'impasse pour le fonctionnement du *crawler*.

Il faut donc bien comprendre ce que le crawler peut et ne peut pas faire.

Il suit très facilement les liens hypertextes standards (balise <a>). **Le crawler suit aussi maintenant les liens qui résultent de l'exécution d'instructions Javascript ainsi que l'Ajax (source : WebRankInfo : http://www.webrankinfo.com/dossiers/indexation/crawl-javascript-post). Mais attention, les liens inclus dans un programme Flash ne sont toujours pas suivis, évitez les sites « full flash » si vous souhaitez être visible. En effet, le Flash n'étant pas visible pour le Crawler, vous ne donnerez donc pas de contenu pour être positionné.**

Comme vous pouvez le voir sur le lien précédent, le crawler peut maintenant franchir certains formulaires. Mais il est fortement recommandé de faciliter l'indexation des contenus et donc de ne pas placer de formulaire avant des pages qui vous semblent importantes. Il faut donc prévoir d'interdire l'indexation de certains contenus de type formulaire afin de ne pas voir de contenus de mauvaise qualité sur les outils de recherche.

Au strict minimum, un site doit pouvoir être visité de manière complète par le crawler.

[4] **PageRank ou PR** : Algorithme d'analyse des liens concourant au système de classement des pages Web utilisé par le moteur de recherche Google pour déterminer l'ordre dans les résultats de recherche qu'il fournit.

Pour cela, il faut privilégier les liens html naturels, interdire (ou réduire) les liens résultant de javascript ou de Flash, et interdire les formulaires qui seraient le point de passage obligé vers certaines branches du site.

Tendez vers un développement de bonne qualité

Lors du développement d'un site web, et même si les principaux CMS du marché respectent les bases du référencement naturel, il faut s'assurer de maximiser cette compatibilité avec les guidelines des outils de recherche. En effet, la base technique doit pouvoir faciliter le passage des robots de « crawling », mais aussi faciliter leur compréhension du contenu et de l'environnement web de votre site web.

Cette optimisation technique aura aussi pour but d'aller le plus loin possible dans l'accélération du temps de chargement du portail. En effet, les outils de recherche (et Google en particulier) suivent ce paramètre de plus en plus précisément.

Pour cela, voici quelques points à garder à l'esprit lors du développement.

- Séparer le contenu (code html) du code de présentation (css) et du code interactif (JavaScript).
- Ne pas utiliser de code css ou JavaScript directement dans le code source.
- Optimiser la compression des fichiers.
- Regrouper les fichiers .css dans une feuille de style externe unique.
- Regrouper les fichiers .js dans un fichier JavaScript externe unique.
- Optimiser l'ordre du code pour un téléchargement parallèle des ressources, appeler les css avant le JavaScript.
- Optimiser la gestion du cache (système ou navigateur)
- Respecter les normes W3C.
- ...

Afin de vérifier la note que Google donne au temps de chargement de votre page, il existe un plugin « Page Speed » pour Firebug sur Firefox ou directement en ligne :

http://pagespeed.googlelabs.com/http://pagespeed.googlelabs.com/ (qui fonctionne aussi pour les mobiles).

Google y donne aussi quelques pistes d'améliorations.

Ce ne sont bien sûr que des propositions, d'autres bonnes pratiques peuvent être employées par les développeurs pour aller plus loin dans le développement d'un site web.

Attention aux liens cassés

Imaginons que votre site ait conquis une petite notoriété. Vous aviez une page passionnante sur les tapis persans du XVIIème siècle et plusieurs sites spécialisés y ont fait référence, apportant ainsi un peu de leur propre notoriété à cette page. Et de là, cette notoriété se propage, comme on le verra, à l'ensemble de votre site.

Mais un jour, vous réorganisez tout cela et ladite page change d'URL. Ou bien pire, vous changez de technologie et ce sont *toutes vos pages* qui changent d'URL. Les liens entrants tombent alors en erreur (NOT FOUND !) et n'apportent plus leur poids à votre site. Après quelques passes, votre *ranking* s'effondre.

Il est fondamental d'analyser toutes les erreurs NOT FOUND (404) générées par votre site et d'en faire une chasse implacable.

Cela à la fois pour le confort de vos visiteurs – concernant les liens internes – et pour la qualité de votre référencement, concernant les liens entrants.

Il faut conserver la plus grande stabilité dans l'organisation de votre site et ses URLs. Si vous modifiez une page, elle doit conserver la même URL. Si une page est supprimée, elle doit être remplacée systématiquement par une instruction de redirection 301 vers une autre page de votre site, par exemple l'accueil mais plutôt vers le contenu le plus proche existant.

Redirection 301

Si les URLs ont changé, alors la seule bonne pratique est de retourner un code « HTTP 301 : moved permanently » signifiant le changement d'adresse définitif de la page.

C'est une intervention fortement recommandée que de rediriger un ancien contenu lors d'une refonte par exemple vers ses nouveaux contenus. C'est un point à préparer en amont du développement du site, lors de la validation de l'arborescence finale, et à mettre en place lors de la mise en ligne.

Ces redirections de type 301 (définitives) sont nécessaires afin de conserver les acquis en terme de positionnement sur les outils de recherche. Nous insistons sur la méthode, c'est obligatoirement des redirections 301 qui doivent être mises en place, c'est la seule recommandée afin de conserver son trafic provenant du SEO.

Les autres méthodes (JavaScript par exemple) vont provoquer la disparition des pages indexées et donc une remise à zéro du référencement naturel du portail. L'historique et l'ancienneté des contenus étant un point clé des calculs des outils de recherche, cela pourrait donc avoir un fort impact sur le référencement naturel de votre site.

Par cette méthode, les outils de recherche seront prévenus de la mise à jour du portail et de ses contenus, ils transféreront alors le « score SEO » de l'ancienne page vers la nouvelle.

Lorsqu'un ancien contenu n'existe plus sur le nouveau portail, il suffira de le rediriger sur ce qui sera considéré comme une page proche ou générique (sur la page d'accueil si nécessaire).

En outre, il faudra prévenir certains sites qui possédaient des liens vers le votre afin qu'ils reprennent une forme correcte, dans l'objectif de garder une forte pertinence. Cela demande de savoir les identifier, ce que nous verrons plus loin.

Comme pour les redirections de contenus, les noms de domaines doivent aussi être gérés via des redirections 301 pour éviter que les outils de recherche en déduisent des sites différents.

Principalement entre l'URL http://www.monurl.fr/ et l'URL http://monurl.fr/.

Outre ce cas de contenu différent entre site hébergé sous WWW et sans, il faut idéalement rediriger l'URL sans le www vers votre URL officielle afin de ne pas perdre de visiteurs.

Le fichier Robots.txt

Les robots crawlers sont bien élevés. Surtout ceux des grands moteurs de recherche.

D'une part, ils se signalent au site c'est-à-dire qu'ils ne se font pas passer pour un utilisateur normal utilisant un navigateur normal. Ils se font connaître en renseignant dans leurs requêtes un champ particulier (user-agent), qui permet de les reconnaître. Ainsi, un site peut analyser ce champ, identifier le crawler, et présenter le cas échéant des pages différentes de celles que voient les visiteurs normaux. Nous verrons que cela peut faire partie des techniques visant à optimiser le référencement (mais que nous ne recommanderons pas spécialement pour ne pas faire prendre de risques à nos clients…).

D'autre part, les robots respectent scrupuleusement les consignes qui leurs sont données par le site visité. Avant de visiter un site, le crawler demande à lire un fichier situé à la racine du site, et nommé *robots.txt*. Ce petit fichier, lorsqu'il existe, donne des instructions au robot, en particulier pour lui préciser le rythme d'indexation qu'il doit respecter, afin de ne pas submerger le serveur, ainsi que les 'branches' du site qu'il ne doit pas indexer.

Les indications peuvent distinguer l'un et l'autre des robots visiteurs.

Par exemple les lignes suivantes interdisent les répertoires /cgi-bin/ et /images/ aux robots.

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

Tandis que la ligne suivante interdit seulement le document email.htm au robot de Google :

```
User-agent: googlebot
Disallow: email.htm
```

Un autre exemple, pour un site voulant rester « secret » :

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

Autrement dit : « à tous les robots : n'indexez rien ! ». Cette configuration n'est bien sûr pas recommandée !

Pour plus d'informations :
http://www.searchengineworld.com/robots/robots_tutorial.htm, par exemple.

Google SiteMap.xml

Google propose depuis mi-2005 un nouveau procédé d'interaction de son référencement avec les sites internet, appellé « SiteMap.xml ». Google SiteMaps présente un nouveau moyen de demander l'indexation des URLs, puis d'obtenir des rapports détaillés sur la visibilité des pages sur Google.

Son utilisation repose sur la mise à disposition, par les webmasters, d'un fichier XML contenant les adresses des pages du site à référencer, ainsi que quelques infos complémentaires comme la date de dernière mise à jour.

Exemple :

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.google.com/schemas/sitemap/0.84">
 <url>
  <loc>http://www.example.com/</loc>
  <lastmod>2005-01-01</lastmod>
  <changefreq>monthly</changefreq>
  <priority>0.8</priority>
 </url>
</urlset>
```

Le bénéfice pour les webmasters est une meilleure maîtrise des pages référencées grâce aux nombreux outils proposés sur l'interface Webmaster Tools (requêtes, erreurs, sitelinks….)

Ce format de fichier est aussi reconnu par les autres outils de recherche. Bing proposant lui aussi des outils comparables pour les webmasters.

PERTINENCE

Le poids des mots

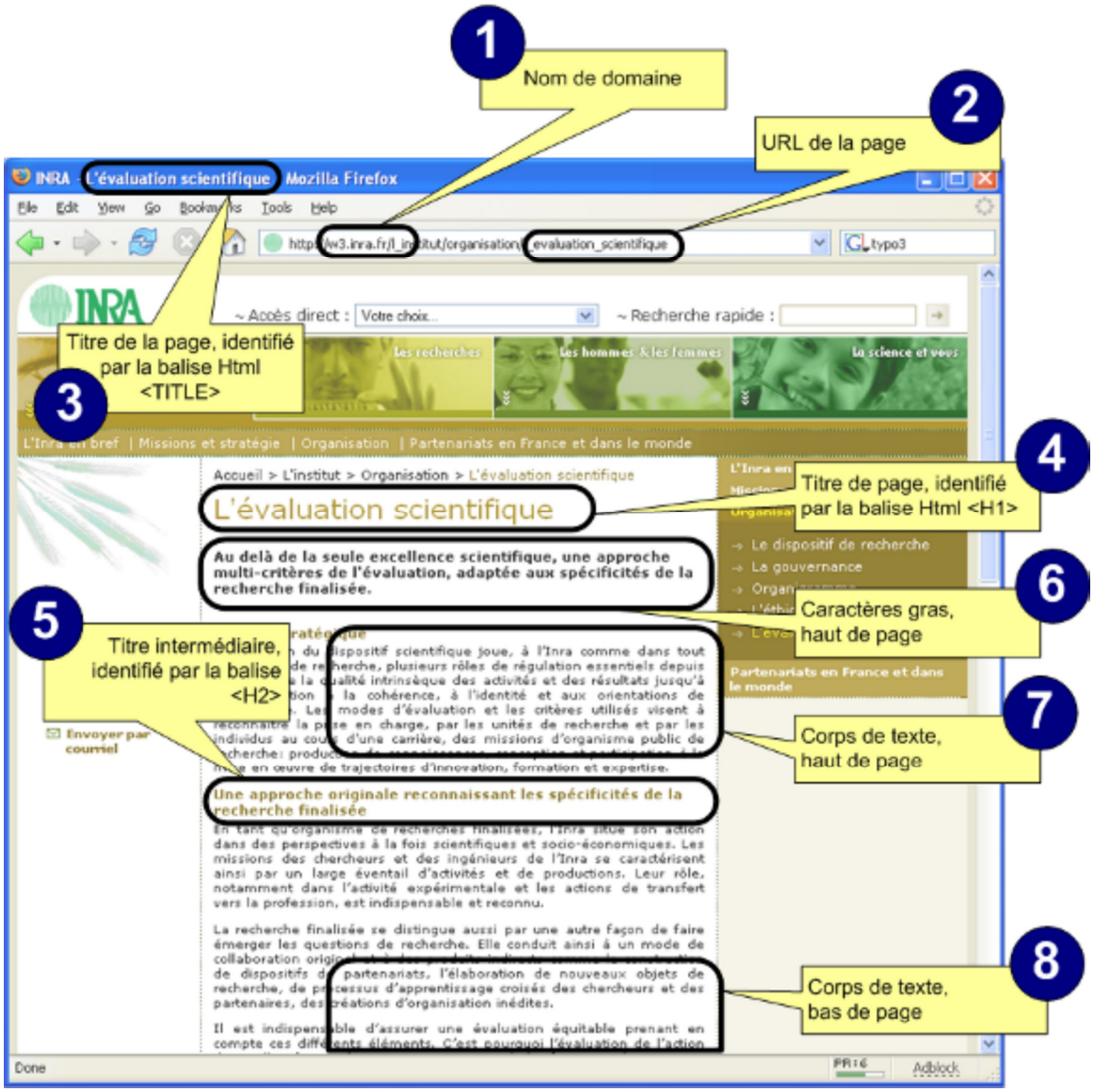
Comme on l'a indiqué précédemment, deux mécanismes se combinent pour déterminer l'ordre des résultats d'une recherche : la pertinence *par rapport aux mots recherchés*, et la notoriété des pages.

Un premier principe, fondamental, est que les mots n'ont pas le même poids selon qu'ils apparaissent dans le titre d'une page, dans un entête, ou dans le corps d'un article.

L'ordre de présence estimé au sein de la page est le suivant :

1. Dans le nom de domaine. Ce n'est pas le plus facile à « travailler » lors d'une refonte par exemple, mais certains s'appliquent à définir des sous-domaines portant des mots clés choisis.
2. Dans l'URL (rewriting)
3. Dans le titre de la page, au sens html (balise TITLE)
4. Dans des titres intermédiaires, selon leur importance (balisage sémantique H1, H2, …)
5. En caractères accentués (gras, ou « strong »).
6. Les mots du haut de page ont un poids plus important que les mots du bas de page.

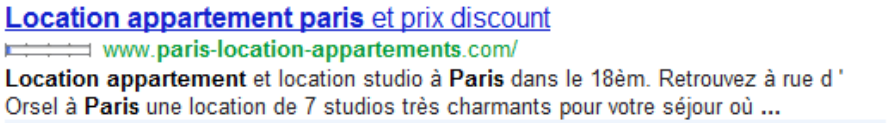
Ce qui est représenté sur la figure suivante.



Les éléments importants d'une page bien référencée

Les URLs

Pour ce qui est du cas des URLs, considérons quelques cas d'école :



Celui-ci, par exemple, a choisi un nom de domaine correspondant directement aux critères de recherche ciblés. Il n'est pas rare de voir ce genre de site créer des sous-domaines cherchant à se positionner sur d'autres requêtes :

Exemple : achat.paris-location-appartements.com
ou vacances.paris-location-appartements.com

(Ce sont des exemples, ces sous-domaines n'existent pas)

L'exemple suivant a une URL beaucoup moins parlante, on en conviendra… Un exemple à ne pas suivre.

L'URL rewriting est une méthode de réécriture des URL techniques générées par les outils de backoffice (CMS par exemple). Cela consiste à transformer une URL technique comportant plus ou moins de paramètres dynamiques par sa traduction en mots clés compréhensibles par un internaute.

Par exemple, plutôt que de voir une URL de type :

http://www.monsite.fr/pid=1234gb43

Nous pourrions réécrire cette URL :

http://www.monsite.fr/avancement-reconstruction-maison.html

Les outils de recherche peuvent cependant suivre et prendre en compte une URL dynamique, et comme les outils de recherche ne donnent pas beaucoup d'importance aux mots clés présents, nous ne sommes pas totalement partisans d'une réécriture d'URL trop poussée. Ce choix doit aussi être fait pour aider l'internaute à naviguer sur votre portail. Comme pour un fil d'Ariane.

Attention cependant, pour un choix portant sur une optimisation de ce type, l'URL rewriting doit respecter certaines règles et ne pas trop en faire. En effet, il ne sert à rien de multiplier les mots clés dans ces URL (que ce soit pour le nom des répertoires ou le nom de la page). Les outils de recherche ne prendront pas en compte l'ensemble des mots clés et iront même parfois jusqu'à pénaliser votre page ou votre portail en entier. C'est un des défauts des CMS qui parfois incluent tous les mots du titre d'une page séparés par des tirets dans l'URL. Ce n'est clairement pas une bonne solution bien que nous ne pensions pas que l'impact sur votre positionnement soit évident. Il vaut mieux se limiter aux mots clés principaux définissant votre contenu.

Nous recommandons de ne pas inclure dans ces URL des mots clés considérés comme « bloquants » comme par exemple : le, la, et, … ce sont des mots clés qui sont généralement appelés « stopword ».

Une bonne URL sera donc la plus courte possible, réécrite si possible, classée dans un répertoire, avec des séparateurs valides (le tiret est recommandé) et comportant des mots clés appartenant au champ sémantique de la page.

Titres

Souvent présentée comme une balise META, c'est en fait une balise HTML (<title>) spécifique qui sert à afficher un texte en haut de la fenêtre de votre navigateur. Mais son optimisation doit bien être une priorité dans votre stratégie SEO.

Le texte indiqué dans la balise Title est un texte qui n'apparait pas dans le contenu d'une page web. En revanche, il est possible de la voir dans la barre supérieure de la plupart des navigateurs web. C'est également le texte qui est repris dans les SERP des outils de recherche, c'est donc votre premier contact avec les internautes.

Considérons un exemple – à ne pas suivre :

1. En 2005

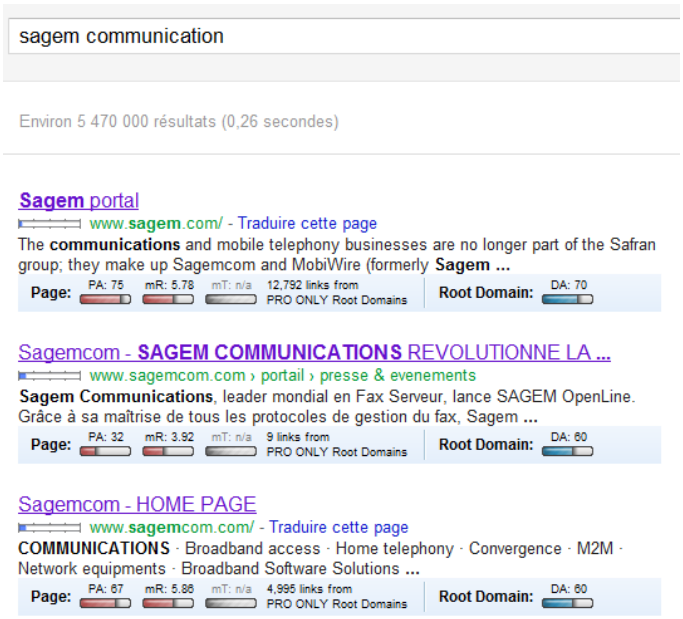
2. En 2012

Ici le site de Sagem Communication (en 2005). Et ensuite la version actuelle.

Le titre (<TITLE>) de la page était « Homepage » en 2005 : non seulement il ne portait pas de mots-clés, mais il ne mentionnait pas même le nom de l'entreprise. Une occasion manquée pour un bon référencement naturel. De plus, dans une liste de favoris, ce lien apparaîtra comme « Homepage », sans plus d'information.

En 2012, une petite amélioration, Sagemcom apparait dans le titre de la page. Cependant, accolé à « -PORTAIL » (encore HOME PAGE sur Google). Bien sur, ces mots clés n'ont pas vraiment d'intérêt en termes de SEO. Le résultat ? La page d'accueil du site Sagem

Communication n'apparait qu'en 3ème position dans les résultats de recherche, et seulement après leur page « presse & événement » mieux optimisée.



Balises META

```
<meta name="description" content="votre contenu" />
```

Cette balise n'est pas prise en compte par les outils de recherche pour positionner votre site. Cependant c'est l'un des éléments qui renforcera le taux de clic sur vos résultats de recherche en incitant les internautes à cliquer sur vos résultats. En effet, une description correctement rédigée fera peut être la différence avec les concurrents présents sur une même requête.

```
<meta name="keywords" content="a oublier" />
```

Cette balise n'aura **aucun impact** sur votre référencement naturel. En effet, elle n'est plus prise en compte par les outils de recherche depuis quelques années. Inutile donc de perdre du temps à optimiser cette balise, sans compter que vous allez donner ici des informations à vos concurrents sur votre stratégie de positionnement facilement et rapidement récupérées.

```
<meta name="robots" content="index,follow" />
```

Cette balise va servir aux outils de recherche pour connaitre vos préférences sur l'indexation (ou non) de votre page, et ainsi de savoir s'il faut continuer à suivre les liens présents dans la page.

Elle n'est pas vraiment indispensable, les robots de crawling vont forcément prendre en compte votre page web. Nous considérons l'utilisation du robots.txt plus efficace. Cependant, le respect des guidelines nous impose sa présence. Il existe bien d'autres valeurs possibles dans cette balise, nous citerons uniquement « noodp » qui permet de prévenir le robot de ne pas utiliser les données venant de l'annuaire DMOZ.

Autres balises nécessaires ou recommandées :

La balise rel="**canonical**" : cette balise est maintenant prise en compte par Google. Elle permet de donner aux outils de recherche l'URL du contenu original si jamais il devait être reproduit ailleurs sur votre site (spécialement efficace pour un site marchand par exemple). La page dupliquée doit donc comporter cette balise. Cependant, nous recommandons bien sûr d'éviter à tout prix le *duplicate content* sur votre site web. Une rubrique spécifique du livre blanc en parlera plus loin.

Open Graph Protocol

L'OPEN GRAPH PROTOCOL a été lancé par Facebook mais est maintenant aussi reconnu par le réseau social concurrent Google+. Ce nouveau protocole permet d'ajouter du sens à vos contenus pour le partage sur les réseaux sociaux principalement. C'est un pas de plus vers le web sémantique. Cela rend le partage de vos pages web sur les réseaux sociaux plus efficace et améliore par la même occasion leur référencement naturel.

Exemple de quelques balises OG :

```
<meta property="og:title" content="Titre de la page et du contenu"/>
<meta property="og:type" content="article"/>
<meta property="og:url" content="http://monsite.fr/article.html"/>
<meta property="og:site_name" content="texte a insérer ici"/>
<meta property="fb:page_id" content="on"/>
```

Ces balises vont permettre à vos pages web d'avoir des liens générés sur Facebook (via les boutons « j'aime ») beaucoup plus parlant et percutant. En effet, l'information est mise en valeur et vous permettra de récupérer du trafic sur votre site web. Vu la puissance du partage des contenus sur les réseaux sociaux, présenter vos contenus de façon ludique et efficace (image/vidéo + description + URL) semble être aujourd'hui indispensable à tout développement.

Ces balises ne sont cependant pas forcément nécessaires sur tous les contenus de votre portail, mais vos pages web comportant des vidéos, des galeries photos ou articles pourraient bénéficier de ces balises spécifiques comme des microformats dont nous allons parler dans le point suivant.

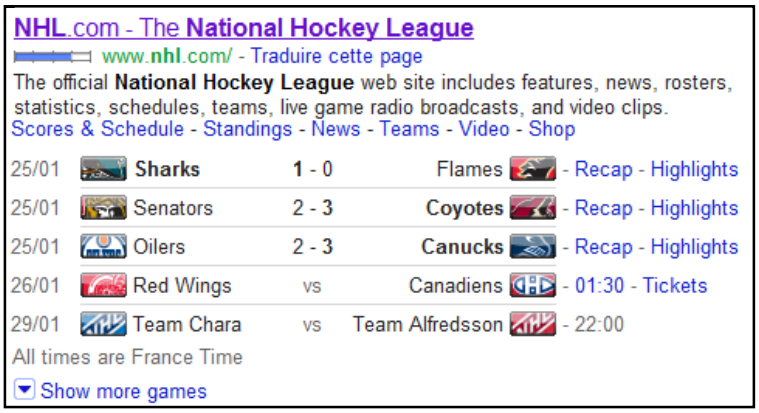
Plus d'informations sur ce protocole : http://ogp.me/

Microformats

Comme pour l'Open Graph protocol, il est possible depuis quelques temps de rajouter des balises spécifiques sur votre contenu pour améliorer les résultats de recherche. Ici nous parlons d'impact sur la recherche universelle de Google.

Ce balisage supplémentaire vous permet d'afficher des résultats de recherche différents. Par exemple dans le cas d'un produit, d'afficher une image, les votes des internautes, une description…

Un bel exemple de ce que peut vous permettre les microformats peut être vu en action sur une recherche sur la requête « NHL ».



En effet, le site officiel de la NHL aux Etats-Unis a utilisé les microformats pour afficher directement dans les outils de recherche le calendrier des matchs passés (avec un lien vers un récapitulatif, les highlights ainsi que le logo de chaque équipe) ou matchs à venir.

L'affichage des résultats de recherche via des microformats n'est cependant pas systématique. C'est même plutôt aléatoire encore pour le moment. Mais les outils de recherche les prennent de plus en plus en compte. C'est un balisage d'avenir.

Plus d'informations sur les microformats : http://schema.org/

Balisage sémantique

Une autre conséquence importante de cette pondération des mots dans la page est la suivante :

Il faut utiliser les vraies indications de titres du html (H1, H2, …) plutôt que des styles spécifiques.

Les outils de recherche parcourent et analysent le code HTML de votre portail avant de générer leurs classements. Ces moteurs doivent comprendre le sens d'un document afin de le classer convenablement et de proposer ainsi des résultats pertinents à leurs utilisateurs. C'est pourquoi ils donnent une grande importance à la structuration des contenus via ce type de balises.

Les balises du langage HTML sont les alliées des moteurs dans leur quête de pertinence. Lorsque ces balises sont utilisées judicieusement, elles permettent d'analyser plus finement la structure d'un document ainsi que de pondérer plus facilement l'importance d'une information ou d'un niveau de lecture.

Mais déployer un balisage sémantique présente aussi de multiples intérêts notamment pour **l'accessibilité**.

Par exemple, voici une arborescence Hn type :

<h1> TITRE DE L'ARTICLE/DU CONTENU </ h1>

<h2> Introduction éventuelle de l'article</ h2>

<h3> sous titre 1</ h3>

<p> Paragraphe et contenu de l'article</ p>

<h3> sous titre 2</ h3>

<p> Paragraphe et contenu de l'article</ p>

…

Attention, en général les CMS gèrent cette arborescence Hn (et souvent H1 est lié au logo). Il est donc nécessaire de revoir le code sur l'ensemble du portail et de garder la même structure Hn sur l'ensemble du contenu.

Des styles spécifiques auront peut-être un rendu de titres, mais ne pourront pas être compris comme des titres par le robot d'indexation.

C'est-à-dire qu'il faut définir <H1>Le Référencement</H1> plutôt que <font face=…>, ou encore <p style=…>. Dans le premier cas on énonce clairement que l'expression « le référencement » a un rôle de titre de chapitre de premier niveau, un rôle important donc. Dans les cas de mise en forme directe, ce n'est pas aussi clair pour le robot.

Bien entendu, on utilisera une feuille de style pour définir la mise en forme associée aux titres H1, H2, H3…

Texte des liens

Les mots intervenant *dans les liens qui pointent vers cette page* ont également une forte pondération. C'est un point souvent méconnu, qu'il est important de souligner car c'est la seule information *extérieure à la page* elle-même, qui influence fortement son référencement.

On suppose que si une page B comporte un lien vers la page A et que ce lien mentionne « le framework Symfony », cela signifie que pour l'éditeur de la page B, la page A était particulièrement pertinente en rapport avec ce thème.

On aurait pu dire que ce jugement est d'autant plus valable que la page B appartiendrait à un autre site, ou un autre nom de domaine, car l'appréciation de pertinence serait plus objective. Même si c'est une chose sur laquelle il est facile de tricher (spamblog par exemple) et dont les moteurs de recherche font une chasse impitoyable, les liens entrants venant de l'extérieur restent l'un des principaux moyens de juger de la pertinence d'une page web.

Ainsi, au sein même de votre site, il est important de choisir vos mots pour créer des liens internes.

Le texte des liens pointant vers une page est considéré comme partie intégrante de la page, avec une pondération importante.

Il faut donc éviter les liens de type générique tels que « voir l'article » ou « cliquez ici ».

Par exemple :

En savoir plus sur les lentilles vertes du Puys et la santé [http://monsite.com/lentilles.html]

Associe le mot *« santé » aux « lentilles vertes du Puy »,* apportant ce mot comme contenu complémentaire à la page.

Tandis que

...Les lentilles vertes du Puys sont un trésor de santé, (voir l'article)

N'apporte que les mots « voir » et « article » dans l'indexation de la page citée.

Et les images ?

Avant toutes choses, il faut éviter de gérer les liens de votre site web uniquement sur des images, cela est probablement moins pertinent et efficace qu'un simple lien texte.

Il faut systématiquement accompagner vos images d'une balise « ALT » qui décrit l'image. Cette optimisation des images vous permettra de rajouter de la pertinence à votre page. Mais aussi de positionner vos images dans les recherches spécifiques images et donc de récolter un trafic non négligeable. N'oubliez pas que le nom de votre image est elle aussi importante ! Par exemple, préférez une image nommée logo-masociete.jpg plutôt que 159GF93.jpg

Les outils de gestion de contenu

Les sites web modernes s'appuient généralement sur des outils de gestion de contenus, ou *content management systems (CMS)*, et il est donc naturel de s'interroger sur la compatibilité de ces outils avec un bon référencement.

Si vous n'êtes pas déjà familiers des principes de la gestion de contenu et des meilleurs outils en la matière, nous vous recommandons **les livres blancs de Smile intitulés *« Gestion de contenus : le meilleur des solutions open source » ou « 200 questions pour choisir un CMS »*.**

Dans un site statique, les pages que voit l'internaute sont des fichiers placés dans une arborescence de répertoires. Le chemin d'accès indiqué dans l'URL est le reflet fidèle des répertoires conduisant au fichier.

Dans un site dynamique, et en particulier un site construit au moyen d'un CMS, les pages n'existent pas sur le serveur, elles sont construites au fur et à mesure qu'elles sont demandées. Les « contenus », c'est-à-dire les textes, images ou documents composant le site, sont placés en général dans une base de données, d'où ils sont obtenus pour fabriquer les pages.

Cela étant, le crawler lui ne s'intéresse pas à la manière dont les pages sont fabriquées : il les demande par une requête http, comme le ferait un simple internaute, les obtient et les lit. Bien sûr dans certains cas, en regardant la forme d'une URL on peut deviner de quelle manière la page a été produite.

Mais il faut bien se souvenir du point suivant :

Le crawler ne fait pas de discrimination, les pages dynamiques ne sont pas moins précieuses à ses yeux que les pages statiques.

Il reste malgré tout quelques différences dont il faut se préoccuper :

- l'URL générée doit permettre d'identifier chaque contenu ; certains CMS utilisent dans ce but une technique appelée *URL rewriting* (ré-écriture d'adresse) permettant d'utiliser le titre des articles et de leur rubrique, comme adresse de la page ;

- On entend dire aussi qu'il faut éviter les paramètres dynamiques dans l'URL, que Google n'apprécierait pas, car ils sont souvent utilisés pour passer des variables de sessions. Cependant, il n'y a aucun problème allant contre ces URL. Elles sont bien prises en compte par Google même si ce n'est pas la meilleure des solutions.

- le nombre de paramètres figurant dans l'adresse doit être le plus petit possible (il est conseillé de ne pas dépasser 3 paramètres) ;
- les balises META (Titre, description) doivent être rendues variables en fonction de chaque article ; dans le cas contraire, les moteurs de recherche pourraient considérer toutes les pages générées comme étant trop similaires et en conséquence n'en conserver qu'une.

Les contraintes qu'impose l'utilisation d'un CMS peuvent alors être transformées en avantages, comme par exemple l'augmentation de la variance du contenu des articles.

URL stables, signifiantes et uniques

Au delà même de la problématique de référencement, la stabilité des URLs est un principe de base du web, mais un principe que certains outils ne respectent pas.

A une URL doit correspondre une page donnée de contenu. La même URL utilisée le lendemain doit fournir la même page.

L'outil de CMS, ou l'application servant les pages, ne doit pas insérer dans l'URL des données techniques variables qui ne sont pas pertinentes pour référencer la page concernée : ni jeton de session, ni information de contexte.

A l'inverse, le CMS ne doit pas non plus utiliser d'information de contexte implicite (i.e. ne figurant *pas* dans l'URL) pour déterminer la page à présenter.

Une autre exigence simple à satisfaire par le CMS est qu'il doit permettre de définir des URLs signifiantes, c'est-à-dire intelligibles, du type : /www.monsite.com/societe/resultats.html et non /www.monsite.com/cmstool ?Id=1294.

Certains CMS sauront utiliser directement le *titre* de la page pour constituer l'URL, d'autres permettront d'indiquer soi-même l'URL désirée. Mais ceux qui n'ont que des URLs reprenant des paramètres dynamiques sont à écarter si possible même si ce n'est pas bloquant pour votre site et son positionnement dans les outils de recherche.

Une autre considération, moins connue, est la réciproque de la précédente : **une même page ne doit pas correspondre à plusieurs URLs différentes.** Car dans ce cas, Google flaire la multiplication artificielle des pages. On a vu ainsi des sites qui utilisaient plusieurs noms de domaine, par exemple www.monsite.com et www.monsite.fr, en servant les mêmes pages sous l'un et l'autre. **C'est une chose à ne pas faire, il faut plutôt mettre une instruction de 301 REDIRECT de l'un vers l'autre.**

Interdiction du duplicate content

C'est sûrement l'un des points les plus importants à suivre tout au long de la « vie » de votre portail : NE PAS DUPLIQUER VOTRE CONTENU ! (mais aussi les titles, descriptions…)

Google et les autres outils de recherche font une chasse impitoyable au contenu dupliqué. Surtout depuis l'année 2011 et la publication de la série d'algorithmes nommés PANDA chez Google. C'est un enjeu considérable lorsque l'on indexe des milliards d'URL mais que l'on doit faire le tri entre ce qui est pertinent et le spam. Les internautes attendent des résultats pertinents en quelques secondes, et vous n'avez que 10 résultats en première page…

Les outils de recherche n'indexent donc pas toutes les URL qu'ils trouvent, tout simplement parce que beaucoup de contenus n'ont aucun intérêt ou sont détectés comme contenu déjà existant (dupliqué).

Les outils de recherche peuvent décider de plusieurs pénalités lorsqu'ils détectent du contenu dupliqué. Selon la gravité de l'erreur ou du spam réalisé, la sanction pourrait être :

- Le pire : être désindexé (disparaitre) des résultats de recherche
- Etre moins souvent crawlé → vos mises à jour ne seront pas détectées rapidement.

- Perdre des positions dans les résultats et être placé dans un index secondaire plus rarement interrogé et donc plus rarement consulté par les internautes.

Dans tous les cas, vous risquez donc d'être moins visible sur Internet…

Sur quoi les outils de recherche se basent-ils pour décider d'un contenu dupliqué ? Et avec quoi comparent-ils ces résultats ?

- Un même contenu sur plusieurs URL différentes.
- Comparaisons des contenus dupliqués avec la popularité de la page et l'autorité du site. Qui aurait copié qui ?

- Un contenu identique, ok. Mais est ce qu'il existe la présence d'un lien vers la source (citation) ?
- La date de publication des contenus, la source la plus récente est forcément celle en qui les outils de recherche auront le plus confiance.
- La date de la première indexation, la date de publication étant aisément falsifiable, les outils de recherche vérifient depuis quand ce contenu est présent dans leur index.

La règle est simple : une page web = un contenu unique = une seule URL

Pour vous assurer contre le contenu dupliqué, voici quelques conseils à suivre. Tout d'abord les erreurs les plus fréquentes en termes de SEO :

- Contenu accessible avec et sans www. Attention à vos redirections 301…
- Des liens internes différents vers un même contenu (attention à vos fiches produits présentes dans plusieurs catégories !)
- Vos pages avec une faible qualité. Par exemple 2 fiches produits très proches, et se retrouvant avec des descriptions identiques et aucun contenu différentiant
- Attention à vos liens entrants contenant des paramètres (de tracking de campagnes par exemple)
- Utilisez vos fichiers robots.txt et sitemap.xml pour cacher les pages à risques (peu de qualité) et surtout celles que vous ne voulez pas voir indexées (par exemple votre backoffice).

- Mise en place d'un « meta robot noindex » sur ces pages ou d'attribut « nofollow » sur les liens
- Suppressions systématiques de vos URL périmées (404) ou indexées par erreur. L'outil Google Webmaster Tools vous aidera pour cela (mais d'autres outils existent).

Bien sûr, la cause du contenu dupliqué n'est pas toujours interne, elle peut être externe avec des « voleurs de contenu ». Ces contenus externes dupliqués pourraient aussi vous pénaliser même si vous avez comme avantage l'ancienneté de leur mise en ligne. Mais pour éviter tout problème, il est nécessaire de vérifier de temps en temps si vos contenus ne seraient pas présents sur d'autres sites web.

Mais surtout n'interdisez pas à vos visiteurs de se servir du clic droit en pensant par cela, empêcher la fonction du copier/coller. La seule conséquence à cela sera de perdre vos visiteurs, mais pas d'empêcher le plagiat éventuel.

NOTORIETE

Backlinks

La gestion des liens entrants vers votre contenu sera un **point clé de votre positionnement**. En effet, depuis les débuts du référencement naturel, ce point est tout particulièrement pris en compte par les outils de recherche. C'est sûrement le travail qui vous permettra de récolter le plus de résultats, mais qui vous prendra aussi le plus de temps.

Obtenir de nouveaux liens pointant vers son site web consiste à augmenter artificiellement sa popularité. Google n'apprécie pas spécialement ce type d'activité, comme tout ce qui est artificiel... comme tout ce qui s'éloigne du naturel...

Cette gestion des backlinks doit être continue. Les outils de recherche prenant en compte l'ancienneté des liens, si votre travail date de quelques mois voir années, il perdra de son influence. Surtout depuis les dernières mises à jour de Google qui donnent beaucoup d'importance au contenu récent.

Ces backlinks doivent comporter des mots clés pertinents liés à votre activité et au contenu présent. Par exemple (lorsque c'est possible de le négocier), préférez un lien avec un texte de ce type : « Le leader de la vente du produit X » (ainsi qu'avec la balise ALT remplie que l'on appelle aussi « ancre ») plutôt qu'uniquement : « Cliquez ici ».

Ne créez pas trop de backlinks en même temps. En effet, si les outils de recherche détectent une vague de liens vers votre site trop importante sur un laps de temps très court, alors que la moyenne est très basse d'habitude, ils pourraient en déduire une tentative de spam. Etalez votre stratégie dans le temps afin d'éviter ce problème. Même si cette masse de liens entrants peut aussi être lié à une actualité reprise largement sur d'autres sites, il faut rester prudent.

Les stratégies de réseaux de sites sont très puissantes et présentent de nombreux avantages :

- Maitrise de votre environnement et des sites web sources
- Choix des ancres textes (quel texte renvoie vers votre contenu)

- Pérennité des supports, vous savez si un site va disparaitre, mais vous savez aussi si vous devez y modifier vos liens lors d'une refonte.

Votre réseau autour de vous est un terrain idéal pour avoir un nombre de liens entrants assez efficace. C'est ici que vous devez commencer le travail en négociant des liens depuis ces sites web.

Attention, ici aussi la duplication de contenu est fortement déconseillée ! Travaillez le texte lié à vos liens. La notion de qualité est importante ici aussi. Le lien doit être présent sur un site pertinent par rapport à votre contenu, il doit lui-même être considéré comme important par Google (Page Rank)…

L'idéal selon votre secteur étant de réussir à trouver des possibilités de liens depuis des sites du gouvernement ou avec un nom de domaine en .gov, .edu... Ces sites ont un poids en termes de référencement naturel très important.

Rappel historique : le PageRank

En 1998, Larry Page et Sergey Brin, étudiants à Stanford University, créent le moteur de recherche Google sur la base de l'algorithme qu'ils ont mis au point : *Page Rank* (*PR*).

Le principe du Page Rank, est le suivant. On considère que lorsqu'une page du web contient un lien vers une autre page, cela signifie que l'auteur de la première accordait un peu de valeur à l'auteur de la seconde puisqu'il jugeait pertinent d'y faire référence. Ainsi, si des milliers de sites de l'Internet contiennent des liens vers la page du site drupal.org consacrée au CMS Drupal, c'est que cette page a quelque intérêt aux yeux de tous ceux qui y ont fait référence.

C'est donc cela qui fait que le site Drupal.org arrivera en tête de votre recherche : des milliers de sites y font référence tandis qu'une plus petite partie ferait référence à une page du site Smile traitant du même sujet, alors que Smile a aussi sa part de pertinence sur cette recherche.

De manière plus précise donc :

- L'internet, *« la toile »*, constitue un immense réseau de pages, reliées entre elles par des liens hypertexte.
- Chaque page P_1 qui contient un lien hypertexte vers une page P apporte une voix, un vote, en faveur de cette page.

- Chaque page *répartit* ses votes entre toutes les pages vers lesquelles elle pointe. Si une page porte 10 liens vers 10 autres pages, alors chacun de ces liens n'apporte qu'un dixième du vote de la page.

- Les votes d'une page sont *pondérés* par le *Page Rank* de cette page. Un lien depuis le site www.cnn.com (PR9) vers votre site lui apporte beaucoup plus qu'un lien depuis le site lalentillevertedupuy.com (PR3).

Revenons sur ce dernier point. Les *Page Rank* de Google sont restitués sur une échelle de 0 à 10. Mais ce *PR* affiché est une représentation logarithmique du *PR* calculé. La base du logarithme n'est pas connue, et varie dans le temps, puisque c'est par définition celle qui permet à la page la plus référencée d'être à la valeur 10. Imaginons que le logarithme soit en base 10. Cela signifie qu'un lien venant d'une page notée *PR5* vaut autant que 10 liens venant d'une page *PR4*, et autant que 100 liens de pages *PR3.*

Une autre manière d'exprimer cela est qu'il faudrait 10^{10} liens de pages sans valeur *(PR0)* pour apporter autant qu'un seul lien depuis la page d'accueil du site W3C (l'un des quelques *happy few* qui avaient des pages *PR10*).

Il faut savoir que toute cette mécanique porte sur des *pages* et non des *sites*. Ce n'est pas un site dans sa globalité qui est plus ou moins bien noté, **c'est chacune de ses pages**. Il peut y avoir une importante disparité de notes entre les pages d'un même site.

Il faut comprendre également que les liens internes à un site sont pris en compte, au même titre que les liens externes. Cela étant, les mécanismes de pondération et de répartition des votes font que les liens internes ne peuvent seuls remonter la notation d'un site dans son ensemble – ou très peu. En revanche, ils ont pour effet soit de concentrer la note sur certaines pages, soit au contraire de répartir la note. Schématiquement, un site comportant beaucoup de liens internes aura tendance à propager et moyenner ses notes vers l'ensemble de ses pages.

Un critère de plus en plus difficile à tromper

L'un des effets de cette évaluation par vote (on peut estimer qu'un backlink est un vote positif pour votre site) est qu'elle est de plus en plus difficile à tromper. Certes il est toujours possible de créer des tas de pages qui pointeront vers votre site, mais Google (suite à sa mise à jour nommée PANDA) fait une chasse impitoyable à ce genre de liens estimés être du spam. N'oublions pas que la création de liens pour améliorer le positionnement de son site Internet est formellement interdit par Google. Même entre sites du même groupe.

Cette voie de tricherie reste ouverte par rapport aux algorithmes de vote : en construisant des dizaines de milliers de pages pointant vers votre accueil, vous apportez effectivement autant de micro-votes, qui finissent par peser. C'était la technique utilisée par la plupart des comparateurs de prix et ses semblables, qui bien souvent polluent les résultats de vos recherches en multipliant les noms de domaines pour le même contenu et les liens croisés entre eux.

La mise a jour PANDA est principalement destinée a ce genre de sites web, mais ne soyons pas naïf, destiné aussi à protéger les propres outils de comparaison de Google.

Le partage, nouvel eldorado

Une traduction simple de l'algorithme *PageRank* est qu'il est bon que d'autres sites pointent vers votre site, c'est-à-dire contiennent un ou plusieurs liens hypertexte en direction de vos pages. Et cela d'autant plus que ces sites sont eux-mêmes connus.

Encore une fois, avant d'essayer de *tromper* ce mécanisme en construisant des liens trompeurs, il est largement préférable d'essayer de jouer le jeu, et d'obtenir de vrais liens, partant de vrais sites.

Si le contenu de votre site est intéressant et que vous faites l'effort de le partager autour de vous, alors vous verrez que les liens viendront tous seuls, car d'autres trouveront opportun de faire référence à votre site. Si votre site contient un contenu unique sur l'histoire du stylo à bille, alors tous les sites évoquant ce sujet voudront faire référence à cette page.

Ensuite, vous pouvez bien sûr demander à vos partenaires de tous ordres de bien vouloir placer des liens vers votre site. Si vous commercialisez des produits, alors ce pourra être les sites de vos distributeurs.

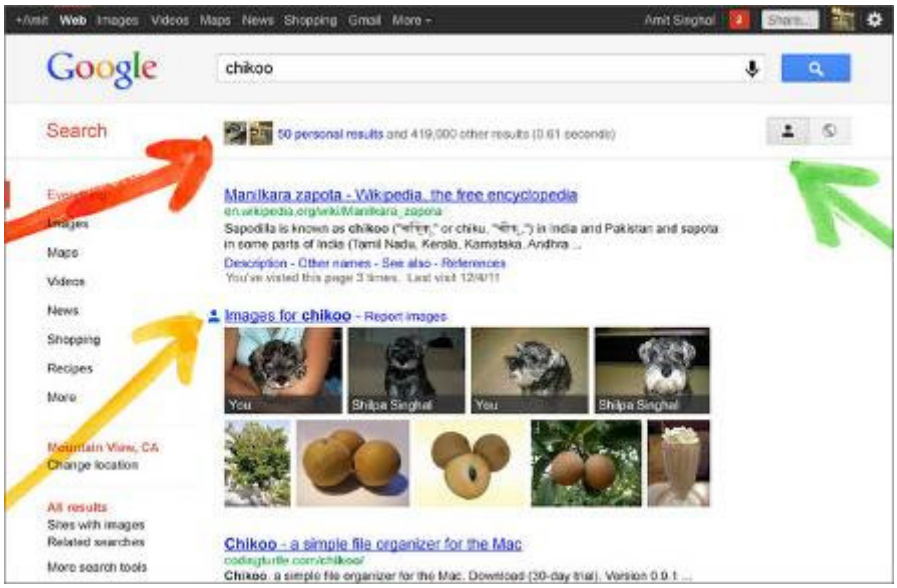
Si votre entreprise appartient à un groupe, alors il est intéressant que les sites du groupe placent des liens croisés vers les autres sites du groupe. Ce n'est pas une façon naturelle de générer du backlink bien sûr, mais cela serait dommage de ne pas se servir de cette base pour créer du lien entrant vers votre contenu. Surtout que les sites de votre groupe sont pertinents pour vous citer.

La limite de cette technique est dans le nombre : trop de liens dilue l'apport de chacun. Aussi là encore la qualité prime sur la quantité : privilégiez ceux avec vos partenaires et/ou des acteurs pertinents de votre domaine.

Depuis quelques temps, nous voyons apparaitre un nouveau type de partage, le social. C'est particulièrement vrai depuis le début de l'année 2012 avec l'apparition de l'option de recherche sociale chez Google. Ce nouveau type de résultats de recherche appelée « Search, + your world ». Littéralement, « la recherche, + votre monde » a pour objectif de vous proposer les contenus partagés par vos contacts sur Google+ liés à votre recherche. De nombreuses critiques ont suivies cette mise en ligne (uniquement aux US pour le moment), en effet, est-ce que vos contacts sont les plus pertinents pour vous proposer des contenus en relation avec votre recherche ?

Cette nouvelle option a aussi pour but de mettre en avant et d'imposer le réseau social de Google, en concurrence avec Facebook ou Twitter. D'où les nombreuses critiques actuelles puisque Google ne se base que sur son propre réseau.

Cependant, l'apparition de ce type d'option doit vous faire prendre conscience qu'une stratégie de partage de vos contenus (et d'animation) sur les réseaux sociaux devient indispensable.



Nouvelle présentation des résultats par Google.com (Search+ your world). En rouge, les résultats par vos contacts comparés au nombre de résultat global. En jaune, les images liées à votre compte Google+ et vos cercles (contacts). En vert, l'option activée ou désactivée.

LA DEMARCHE

La vraie pertinence

L'une des premières choses à retenir pour un bon référencement est la suivante : **avant d'essayer de tromper le moteur, essayez de le satisfaire**. Considérez un peu le référencement comme de la séduction : avant d'essayer d'avoir l'air subtil, spirituel et attentionné, soyez-le vraiment !

Ce sera peut-être la meilleure des recettes, et cela pour deux raisons : la première, c'est que vous obtiendrez un bon référencement sans faire des choses compliquées ou tordues, et la seconde c'est que vos visiteurs en profiteront directement puisqu'ils trouveront des informations plus pertinentes sur votre site.

Facile à dire ? Certes, mais pas impossible à faire. La vraie recette tient en peu de mots : placez sur votre site de l'information intéressante et abondante traitant des thèmes correspondant à l'indexation souhaitée. Votre site vend des fournitures de bureau ? Et bien trouvez des choses intelligentes à dire sur les fournitures de bureau. **Vous devez en être capables, c'est votre métier après tout, les fournitures !** Citez des marques, des modèles, des catégories, l'histoire du stylo à travers les âges, les qualités de papier, tout est bon. Attention, pas des listes de mots placés côte à côte : non, du contenu, du vrai, non seulement *intelligible*, mais même *intelligent* si possible.

Ensuite, organisez tout cela en sections, sous-sections, ajoutez des liens internes de navigation, et voilà. Sans même tricher, vous avez fait la moitié du travail, et votre référencement est déjà assez bon. Alors imaginez en optimisant un peu !

Il vaut parfois mieux payer quelqu'un à créer du contenu intelligent pour votre site que payer quelqu'un à faire croire que ce contenu est intelligent.

Quels mots pour arriver à mon site ?

C'est toujours l'une des premières questions à se poser : **pour quels ensembles de mots clés est-ce que je souhaite être bien positionné ?** Si j'ai des choses à vendre, alors que recherchent mes clients ? Et plus précisément, comment mes clients exprimeront-ils leur recherche ?

C'est la première question qu'il faut se poser, et il faut se la poser avant de commencer à écrire pour son site : Comment mes visiteurs exprimeront-ils leur recherche ? Quels mots utiliseront-ils ?

Comme on l'a vu, les internautes savent de plus en plus qu'il leur faut cibler leur recherche en combinant plusieurs mots. C'est donc aussi pour différents *groupes de mots* qu'il conviendra d'apparaître en bonne place.

Le premier exercice est donc de lister ces mots et groupes de mots par écrit, à l'occasion d'une séance de réflexion de type *brainstorming*.

Ensuite, on s'assurera que ces mots sont bien présents dans vos pages. Il arrive couramment que rédaction et référencement soient deux processus disjoints : on essaye a posteriori d'associer des mots-clés à des articles déjà écrits. Mais il est largement préférable que les textes du site utilisent effectivement les ensembles de mots choisis.

Attention également aux synonymes ou variantes. Dans le cas du site du CNLL (Conseil National du Logiciel Libre) par exemple, les visiteurs peuvent saisir « opensource » ou bien « open source » ou encore « logiciel libre », et d'autres équivalents encore. Il est difficile d'utiliser systématiquement tous ces mots dans un article, et le souci d'un style clair amènerait plutôt à choisir une formulation unique. Mais pour la qualité du référencement, il pourra être préférable au contraire de varier les expressions. Varier les expressions à dessein, certes, mais tout en évitant les variantes de pur style, qui au contraire pollueraient la perception.

Soyons clairs toutefois : si le vocabulaire, *pour les thèmes fondamentaux*, doit être étudié avec soin, il ne s'agit surtout pas d'*écrire pour le référencement*, c'est-à-dire de faire des phrases qui n'auraient pas d'autre finalité que le référencement. Elles gêneraient le lecteur, sans apporter le bénéfice attendu. La notion de **qualité** doit rester l'une des plus importantes lors de la construction de votre contenu. En effet, un contenu de qualité sera partagé et donc… mieux référencé !

Quels mots recherchaient mes visiteurs ?

La réflexion amont, évoquée ci-avant, doit être validée par une analyse en aval : quels mots avaient saisi mes visiteurs lorsqu'ils sont parvenus sur mon site par un moteur de recherche ?

Les outils de suivi d'audience tels que Google Analytics, Analyser (AT Internet, anciennement XiTi)… permettent de connaître les mots-clés qu'avaient saisi les visiteurs de

votre site, si c'est au moyen d'un tel moteur que l'internaute est arrivé. En effet, les mots-clés recherchés sont inscrits dans l'URL appelante, ou « referer ».

Il est important de consulter régulièrement cette liste des mots-clés ayant conduit à votre site, pratiquement dans toute son étendue.

C'est ce qui permettra de valider ou d'ajuster les mots que vous-même vous utilisez pour votre référencement. Peut-être que vos visiteurs avaient une manière de formuler leur recherche qui n'était pas ce à quoi vous vous attendiez. Peut-être aussi que certains visiteurs parviennent à votre site par erreur, avec des mots-clés qui ne correspondent pas à la finalité de votre site. A moins que vous ne recherchiez l'audience à tout prix, ces erreurs de routages impliqueront également un réajustement des mots utilisés pour le référencement.

Les mêmes outils, de suivi d'audience, vous donneront une autre information essentielle : la part de vos visiteurs qui sont arrivés sur votre site par l'intermédiaire d'un moteur de recherche. Il est essentiel de la connaître et de la suivre.

Si votre site connaît une chute d'audience par exemple, est-ce dû à un problème dans son référencement ou une autre source de visites ? Il est fondamental de pouvoir répondre à cette question. Bien d'autres facteurs peuvent être considérés : un site concurrent draine du trafic, un problème en hébergement a ralenti votre site et fait fuir des visiteurs, un site partenaire a retiré un lien qui amenait des visiteurs, ou tout simplement l'intérêt de vos informations a baissé.

Attention aussi à cette notion de mots clés saisis par les internautes. Google cache cette information pour ses inscrits connectés à leur compte depuis mars 2012 (https). C'est-à-dire que lorsque vous serez connecté à votre compte Google, votre navigation sera chiffrée. Et les sites web que vous visiterez via une recherche sur son moteur web ne pourront plus enregistrer quelle recherche vous avez faite pour arriver chez eux.

Ce qui va faire apparaitre une nouvelle donnée dans vos listes de mots clés, cette ligne s'appellera par exemple sur Google Analytics : « **not provided** ». Elle est d'ailleurs déjà présente dans vos résultats aujourd'hui, mais dans une petite proportion (environ 8% constaté sur certains sites).

Cela implique donc qu'une partie de votre analyse de trafic ne sera plus exploitable pour votre stratégie SEO.

Google estime à 15/20% de trafic « caché » dans l'avenir sur vos statistiques. Mais c'est une donnée importante à suivre dans les mois à venir.

Quels liens pointent vers mon site ?

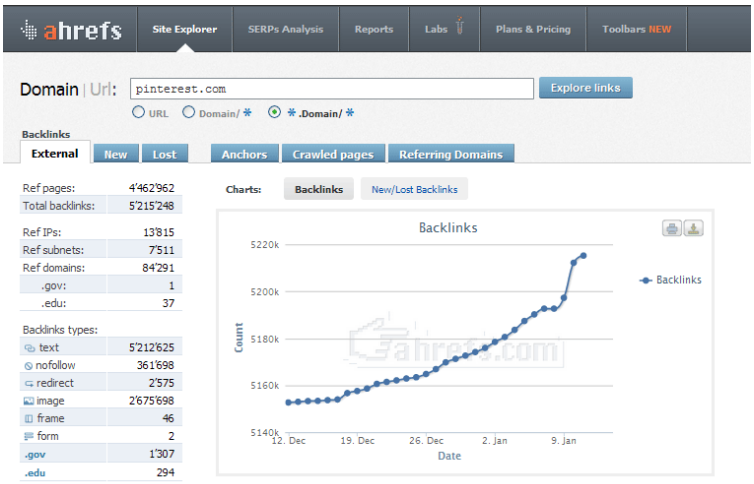
On a vu toute l'importance des liens entrants vers votre site, surtout en provenance de sites eux-mêmes à forte notoriété. Il est donc bien sûr intéressant de connaître ces liens que d'autres ont définis vers vos pages.

Il existe une fonction de Google qui répondra à cette question : il suffit de saisir dans la commande de recherche :

Link : www.monsite.com

Et Google listera les pages, dont il a connaissance, référençant votre site.

Cependant nous vous conseillons plutôt de passer à des outils en ligne comme ahrefs.com ou majesticseo.com par exemple qui vous donneront énormément d'informations sur vos liens entrants. En effet, non seulement Google ne vous donne pas d'informations qualitatives à propos de ces liens, mais ce n'est plus une commande de confiance aujourd'hui.



Le volume compte

Le nombre de pages d'un site est, en soi, un facteur de bon référencement.

On a coutume de privilégier la qualité sur la quantité, et il est clair que pour un visiteur, il serait préférable d'avoir 20 pages synthétiques et pertinentes plutôt que 200 pages diluées et redondantes.

Le service du visiteur et les besoins du référencement seraient-ils, pour une fois, contradictoires ? Pas vraiment : il suffit de mettre en lignes 200 pages toutes synthétiques et pertinentes !

Non, ce n'est pas si simple bien sûr. Mais retenons juste ce principe : le volume compte.

Pour les sites qui présentent, au moyen d'applications spécifiques, des contenus issus d'une base de données, par exemple des petites annonces d'emploi ou d'immobilier, ou bien des produits issus d'un catalogue, il y a une conséquence toute simple : la totalité des pages de contenus doit être référencée. C'est à dire qu'il faut faire en sorte d'aménager un chemin pour le crawler qui mène vers chacune des pages de détail.

Lorsqu'on est un site d'annonce tel que Cadremploi.fr par exemple, avec 15 000 offres d'emploi en base de données, donner accès à ces 15 000 pages de contenus pertinents pour l'indexation, par rapport aux quelques centaines de pages de contenus éditoriaux, peut faire une énorme différence.

LES RUSES

Des réseaux de pages creuses

Le summum du détournement de pertinence est peut être atteint avec un site de type comparateur en ligne, qui fabrique des milliers de pages vides de sens, correspondant aux paires de mots-clés recherchées par les visiteurs. Il suffit qu'un internaute tape « vol tourisme Italie » pour que le site fabrique une page vol_tourisme_Italie.html. Cette page contient le résultat d'une recherche sur ces mots-clés, c'est-à-dire un contenu qui semble pertinent, mais n'a en fait aucune valeur ajoutée vraie. Les comparateurs de prix, comme Twenga et ses semblables procèdent de manière identique : quels que soient les mots, ils ont toujours des pages à mettre en face. Ainsi, le site soumet à Google des milliers de pages vides, dont le seul contenu est lui-même issu d'une recherche, peut-être sur Google soi-même ! A quoi sert tout ce vide ? Sans doute à créer de l'audience en se servant à outrance de la longue traîne[5], puisque ce type de pages a provisoirement réussi à tromper le moteur de pertinence de Google, et sortent donc fréquemment en haut de classement. Et un peu d'audience, permet un peu de pub et de juteux bénéfices. Mais même les publicitaires devraient se méfier de telles pratiques, qui associent leurs marques à une tromperie.

La technique est donc clairement à déconseiller : à la fois très lourde à mettre en place, et assez risquée. Surtout depuis l'année 2011 et la publication de l'algorithme Panda, justement fait pour traquer ce genre de résultats.

Sans compter que fabriquer une telle pollution à grande échelle sur le web est profondément incivique.

[5] **Longue traîne** : fait référence aux mots clés qui attirent séparément peu de visiteurs sur un site Internet. Le cumul de ces mots clés à faible trafic peut alors représenter une part non négligeable du trafic total d'un site.

Les pages spéciales moteur

Comme on l'a dit, les robots indexeurs sont bien élevés : d'une part ils respectent les instructions du fichier *robots.txt* et d'autre part, ils ne cherchent pas à se faire passer pour un internaute quelconque, ils s'identifient clairement, au moyen du paramètre *user-agent* qui est défini dans chacune des requêtes http.

User-agent permet généralement d'identifier le *navigateur*, et certains sites l'utilisent pour adresser des pages différentes selon les possibilités du navigateur cible.

Ainsi, le robot Google s'identifie en indiquant « user-agent=googlebot » dans chacune de ses requêtes.

Il est donc possible d'utiliser ce paramètre pour servir à Google des pages spéciales, différentes de celles qui seront servies aux internautes.

Cette technique a été beaucoup utilisée aux débuts du référencement, pour servir à chaque moteur d'indexation des pages correspondant à ses caractéristiques. Yahoo aimait les keywords, on lui en donnait, … Altavista voulait des <H1> mais ne supportait pas le « bourrage » de keywords, on lui donnait satisfaction aussi.

C'est une technique complexe, qui demande un travail considérable, pour des résultats aujourd'hui assez faibles.

Cependant elle a encore ses adeptes aujourd'hui, spécialement pour la sur-optimisation des pages satellites. Pour voir un exemple, tapez par exemple « louer appartement » sur Google, et regardez les premiers résultats. Dont celui-ci : http://www.acheter-louer.fr/location_appartement.html

Comme vous pouvez le voir, le mot clé recherché est présent dans l'URL mais aussi partout sur la page. Sous la forme de liens, du mot clé en gras, d'images…

Cette technique de pages satellites fonctionne encore donc très bien. Cependant, ce site a tout de même fait les choses bien en intégrant de vraies annonces dans ces pages qui peuvent être définies comme des pages de résultats de recherches internes au site. Mais cela redirige aussi vers un tri par arrondissement, etc. Un maillage interne qui semble efficace au vu de sa place sur Google.

La punition des fraudeurs

On l'a dit, le référencement est une guerre sans merci. Mais dans cette guerre, les moteurs disposent de l'arme atomique et pas vous : **le déréférencement ou *blacklisting***. Si le moteur de recherche décèle une tentative de tricherie, il peut *black-lister* le site dans son ensemble, c'est-à-dire que plus aucune recherche ne restituera des pages de ce site, pas même en 1000ème position. Le site n'existe plus pour Google.

C'est une punition sévère, qui peut durer plusieurs mois. Et comme tout cela est régi par des algorithmes, sans intervention humaine, il est très difficile d'aller supplier un retour en grâce. Le cas n'est pas théorique et nombre de prestataires en référencement un peu trop inventifs s'y sont déjà brûlé les doigts. BMW, Castorama, Ricoh, ou bien même Netbooster, en savent quelque chose.

Bien que depuis quelques temps il semble que Google ne « supprime » plus les sites qui trichent de son index, mais dévalue simplement leur *PR*, cela reste une raison suffisante pour ne pas essayer de s'y risquer.

Mais comme nous l'avons souligné plus haut, la principale raison est ailleurs : viser un meilleur référencement sans tricher, c'est aussi mieux servir vos visiteurs, en leur offrant une vraie pertinence des contenus.

EN CONCLUSION

Après plusieurs années d'expérience des acteurs de ce domaine, et l'observation de l'évolution des moteurs, il apparaît que la **qualité du fond** (richesse de contenu, pertinence, organisation, spécialisation des pages) et **de la forme** (simplicité, respect des normes, application de règles simples d'organisation du contenu) **restent les valeurs sûres** : un site **bien pensé**, **bien réalisé**, et **bien suivi**, devrait dans la grande majorité des cas obtenir et conserver un bon positionnement.

Du côté des moteurs, l'hégémonie de Google a permis de stimuler le web pour en augmenter la qualité. L'internaute doit toutefois rester vigilant et critique car cela pourrait entraîner des dérives et excès, et après tout, les résultats d'une recherche ne constituent qu'un seul point de vue.

En somme, que l'on soit du côté des webmasters ou du côté des internautes, le plus sûr est de conserver son bon sens.

Si vous avez des besoins en référencement, vous pouvez contacter Smile Digital, l'agence numérique de Smile, spécialiste en stratégie on-line, SEO, conception graphique et ergonomique... à l'adresse : contact@smile.fr