

Livre blanc | 2007

Référencement : la revanche du contenu

2007 a marqué une véritable révolution dans le marché du référencement. Les principaux moteurs de recherche placent désormais la qualité du contenu d'un site au cœur de son référencement naturel, mais aussi payant – une évolution qui modifie en profondeur les stratégies de liens sponsorisés.

Face à ces enjeux, les actions de référencement traditionnelles, basées sur une approche par mots-clés et l'optimisation des contenus existants, montrent très vite leurs limites. La solution ? Mettre en œuvre une approche éditoriale globale, visant à exploiter un univers sémantique riche et régulièrement actualisé.

La révolution "contenu"	3
1. La base du référencement naturel	3
Pas de référencement sans contenu texte.....	3
Contenu et popularité	3
2. Un impact sur le référencement payant	3
La prise en compte de la "qualité"	4
Des contenus plus qualitatifs, une réflexion en amont	4
3. Le Web 2.0 carbure au contenu	4
Des médias naturellement référencés	5
Un nouveau champ de veille	5
Les techniques traditionnelles de référencement montrent leurs limites	6
1. Optimisation des pages : « Le feriez-vous si Google n'existait pas ? »	6
Vivre sans les pages satellites.....	6
Back to the basics.....	8
2. Requêtes de « micro-traffic » : les limites de l'approche par mots-clés	8
Un nombre illimité de combinaisons.....	8
Comment ignorer quatre visites sur cinq	9
Des requêtes rares à l'origine de la majorité du trafic	10
3. Le netlinking, beaucoup d'efforts, peu d'effets.....	11
Moins de liens, mais plus de qualité	11
De l'échange au don.....	11
Bon pour le lecteur, bon pour le moteur.....	13
1. Priorité à l'internaute	13
2. Structurer l'information	14
Organiser les contenus.....	14
Multiplier les points d'entrée	15
Limiter la profondeur	16
Respecter la "pyramide inversée"	16
3. Faire court	17
4. Exploiter les niveaux de lecture	17
Titre de page	17
Titres et sous-titres	18
Intertitres, gras.....	19
5. Soigner les liens.....	19
Vers le référencement global.....	21
1. Une approche plus large de l'univers sémantique	21
Puiser à toutes les sources	21
Penser en trois dimensions.....	22
Varier les formats.....	24
2. Une démarche dynamique	25
Se maintenir au cœur du système	25
Soutenir l'attention	25
3. Se fixer des objectifs et vérifier qu'ils sont atteints	26

La révolution "contenu"

Le rôle du contenu en matière de référencement naturel est connu depuis la naissance des moteurs, même s'il était rare de voir un site en tirer pleinement parti. Mais l'indifférence au contenu n'est plus d'actualité : désormais, il s'invite dans le **référencement payant**, et se trouve au cœur du **Web 2.0**.

1. La base du référencement naturel

Le contenu éditorial est indispensable à la **bonne indexation** d'un site et, par sa richesse, sa variété et sa pertinence, il en conditionne la **visibilité**. Il joue également un rôle majeur (et trop souvent sous-estimé) dans la **qualité des liens entrants**.

Pas de référencement sans contenu texte

Quelle est la **première source d'information des moteurs de recherche** ? Le contenu éditorial.

Les *crawlers*, les robots des moteurs de recherche qui parcourent en permanence le Web, ne "voient" ni les images, ni les animations Flash, ni les vidéos... mais sont **friands de contenu texte**, qu'ils indexent intégralement. Et seules les pages correctement indexées peuvent être référencées par les moteurs, et donc apparaître dans leurs **résultats**.

Le contenu d'un site constitue donc la base même de son référencement. Avec ce corollaire : plus l'offre éditoriale d'un site est riche et variée, plus sa **visibilité** sur le Web est grande.

Le périmètre de votre contenu détermine ainsi celui de votre **territoire** sur la toile : devenir un vaste continent, une haute montagne, bref, un **point de repère** dans le monde en expansion constante du Web, nécessite de développer et d'actualiser des contenus pertinents.

Dans le monde réel, la **visibilité** est avant tout question d'images, **sur le Web, c'est une question de mots**.

**Pas de contenu, pas d'indexation.
Pas d'indexation, pas de référencement naturel.**

Contenu et popularité

Une fois les pages d'un site correctement indexées, un autre facteur entre en jeu au moment où le moteur estime leur pertinence par rapport à une requête donnée (et donc leur **positionnement** dans sa page de résultats). Il s'agit de leur "popularité", c'est-à-dire du nombre et surtout de la qualité des **liens** pointant vers ces pages. C'est le principe du célèbre PageRank de Google.

Là encore, le **contenu** des pages a un impact déterminant : il doit être cohérent avec le lien qui pointe vers lui, sans quoi ce dernier ne sera pris en compte que **partiellement** par les moteurs.

Et surtout, un contenu éditorial de qualité va attirer spontanément des liens entrants, provenant de sites à forte notoriété, inaccessibles via une politique traditionnelle d'échange de liens : médias, blogs, sites communautaires... C'est le principe du **contenu viral** (ou *linkbaiting*), qui est en train de remplacer peu à peu les politiques traditionnelles d'échanges de liens.

2. Un impact sur le référencement payant

Jusqu'à début 2006, les démarches de référencement naturel et d'achat de liens sponsorisés étaient considérées, au mieux, comme **complémentaires**. La première permettait une visibilité sur le long terme, la seconde des actions de génération de trafic à court terme, et les deux domaines ne s'interpénétraient pas véritablement.

La prise en compte de la "qualité"

La mise en place par Google de la notion de "Quality Score"¹ a bouleversé la donne. Désormais, **la "qualité" de la page vers laquelle mène un lien sponsorisé peut influencer le positionnement de celui-ci, ou son coût au clic**. L'achat de liens n'est plus déconnecté du contenu du site ; le contenu des pages relayées devient partie intégrante de toute stratégie "payante".

Une relation à sens unique

La qualité d'un site influence sur les liens sponsorisés... Mais la réciproque n'est pas vraie. Jamais une campagne d'achat de liens n'aura d'impact sur le positionnement naturel d'un site, qui dépend essentiellement de deux facteurs : la qualité des liens entrants, et la pertinence de son contenu.

Cette évolution est logique : les moteurs tentent constamment d'améliorer la **pertinence** de leurs résultats naturels ; ils appliquent aujourd'hui ces critères aux résultats sponsorisés. Et elle s'inscrit **dans la durée**. *Panama*, la nouvelle plate-forme publicitaire de Yahoo! dévoilée début 2007, intègre également la notion de qualité des pages.

S'il est toujours possible de bâtir une campagne de liens sponsorisés à partir d'un site au contenu minimaliste, cela va donc devenir **de moins en moins viable économiquement**, dans un contexte où les marges sont faibles et la concurrence, impitoyable.

La qualité des contenus d'un site, qui impactait jusqu'ici essentiellement son référencement naturel, est désormais un facteur essentiel de son référencement payant.

Des contenus plus qualitatifs, une réflexion en amont

Les spécialistes des liens sponsorisés ont très vite pris conscience des enjeux de cette "révolution du contenu". Ils ont analysé la notion de "Quality Score" et en ont déduit la nécessité :

- d'une intégration très en amont des stratégies de référencement naturel et payant,
- d'un développement des **contenus** des sites de leurs clients, en **quantité** comme en **qualité**.

Sur ce deuxième point, les techniques de référencement pratiquées aujourd'hui se révèlent **très largement insuffisantes**. Les attentes qualitatives des moteurs, relayées par les sociétés de *Search Engine Advertising* (SEA), devraient tirer le marché vers le haut. Mais les modifications à apporter au niveau des sites Web sont encore mal estimées et maladroitement mises en œuvre.

3. Le Web 2.0 carburant au contenu

Ces derniers mois ont été le théâtre d'une autre révolution : l'explosion des **blogs**, l'essor des **sites communautaires** et d'échange, la mise en place de **réseaux sociaux** regroupant des milliers, voire des millions d'internautes. La portée réelle de ce "Web 2.0" ne sera mesurable que sur le long terme, mais son développement rapide influence d'ores et déjà le paysage du référencement.

¹ Le Quality Score a été mis en œuvre en deux étapes (11 juillet, puis 6 septembre 2006). Une évolution portant uniquement sur le contenu des textes d'annonces ("Content Quality Score") a été mise en place en novembre 2006. Elle précise notamment : "We [Google] do use a keyword's Quality Score to rank your ads and position them on network pages. Your keyword's Quality Score is based on the relevance of the ad and keywords to a content site, your ad's performance history on the site and similar sites, and the quality of your ad's landing page." (<https://adwords.google.com/>)

Des médias naturellement référencés

Ces nouveaux médias, dits "CtoC" (Consumer to Consumer), disposent en effet de deux atouts majeurs en termes de **référencement naturel** :

- Ils produisent des quantités faramineuses de **contenu éditorial**, d'une qualité qui peut être équivalente (voire supérieure) à celle des sites institutionnels.
- Ils bénéficient d'un réseau de **liens** très développé, même si leur popularité globale reste faible.

Ils disposent donc des qualités requises pour se positionner très efficacement dans les résultats naturels des moteurs – un phénomène de plus en plus visible depuis mi-2006.

La compétition pour les meilleures places dans les pages de résultats n'oppose plus seulement des sites traditionnels, acteurs majeurs de leur secteur. Les médias CtoC sont **de nouveaux concurrents**, souvent très actifs. Il convient d'en tenir compte dans le cadre du référencement naturel d'un site.

Un nouveau champ de veille

Il est désormais nécessaire d'étudier avec soin les médias issus du "Web 2.0". Et il peut être intéressant d'en **développer** certains, en parallèle de la communication institutionnelle.

En se limitant au domaine du référencement, l'étude des médias CtoC doit intervenir :

- Lors de la détermination du contexte concurrentiel (ils peuvent constituer une source d'enrichissement des **termes** à employer, des **thématiques** à aborder...).
- Lors de la mise en place de la stratégie de **liens entrants**, en planifiant des interventions sur certains blogs ou sites communautaires soigneusement sélectionnés.
- Lors de la production des **contenus** du site, en intégrant éventuellement un blog, un forum, un espace communautaire... au sein même du site.

Points-clés

- Le visage du référencement naturel a **profondément changé** en quelques mois.
- Le rôle du **contenu éditorial**, reconnu mais jusqu'ici sous-exploité, s'est considérablement renforcé.
- Principales raisons : le développement des **médias CtoC**, la notion de **qualité des liens entrants**, la prise en compte du contenu des sites dans l'efficacité de leurs campagnes de liens sponsorisés.

Les techniques traditionnelles de référencement montrent leurs limites

A l'heure actuelle, le référencement naturel d'un site recouvre des démarches distinctes :

L'optimisation du contenu des pages en fonction de mots-clés considérés comme stratégiques

L'obtention de liens entrants, par un système d'échange ou – plus rarement – d'achat.

Ces techniques actionnent les deux leviers fondamentaux du référencement naturel (**qualité du contenu** et **popularité**), mais de manière très insuffisante. La raison ? Elles restent centrées sur les **moteurs**, et privilégient la dimension **technique**.

A l'heure du Web 2.0, le référencement doit être centré sur les lecteurs, et privilégier la dimension éditoriale.

1. Optimisation des pages : « Le feriez-vous si Google n'existait pas ? »

Traditionnellement, l'optimisation des pages d'un site en vue d'améliorer son référencement naturel se déroule en deux étapes :

- La détermination des mots et expressions-clés stratégiques.
- La création ou la réécriture de pages Web "optimisées", afin d'améliorer le positionnement du site dans les résultats des moteurs pour ces mots et expressions.

Vivre sans les pages satellites

Le principe de cette "optimisation" est de réécrire les pages du site (ou, si nécessaire, d'en créer) en respectant un certain nombre de **paramètres techniques**, le principal étant **la densité élevée en mots-clés**. Dans le texte apparent, mais aussi dans le titre des pages, dans certaines balises HTML "invisibles", dans les liens...

Le principal problème posé par cette méthode est qu'une page "parfaitement" optimisée pour les moteurs (c'est-à-dire truffée de mots et d'expressions-clés, sans souci de logique) risque de devenir rapidement **illisible** pour un internaute humain.

Densité ou quantité ?

Pourquoi cette obsession de la densité ? Parce qu'elle est prise en compte par les moteurs pour déterminer la pertinence d'un mot-clé.

La densité d'un mot-clé désigne le **rapport** entre le nombre d'occurrences de ce mot dans un contenu (paragraphe, lien, titre de page) et le nombre de mots total de ce contenu. Dans le titre "Vivre sans les pages satellites", la densité du mot "pages" est de 0,25 (le terme "les" n'est pas pris en compte par les moteurs).

Cette densité n'est donc **pas synonyme de quantité** : un mot-clé qui apparaît cinquante fois dans un texte très long sera moins "dense" que s'il apparaissait cinq fois dans un texte beaucoup plus court.

Pour contourner cet écueil, il existe deux solutions :

- La première est **éditoriale**, et consiste à développer un contenu qui réponde à la fois aux attentes des moteurs... et à celles des lecteurs.
- La seconde est **technique**, et c'est – de très loin – la plus répandue. Elle repose sur un raisonnement simple : puisque ces pages sont illisibles, il suffit de les rendre... invisibles. Et de créer des pages de contenus destinées **aux seuls moteurs**, qui renvoient automatiquement les lecteurs humains vers une version non optimisée – dont le contenu peut être très différent. Ce sont les fameuses « **pages satellites** », qui ont constitué pendant des années la technique d'optimisation la plus prisee².

Que pense Google des pages satellites ?

C'est l'un des secrets les moins bien gardés du Web. La page qui réunit l'ensemble des conseils techniques et éditoriaux édictés par le célèbre moteur pour faciliter l'indexation et le référencement des pages Web, est librement accessible : <http://books.google.com/webmasters/guidelines.html>.

Sobrement intitulée "Instructions pour les webmasters", elle a été mise en place dès 2001, et a assez peu évolué depuis. On y trouve notamment, dans la section "**Principes de base**" de la rubrique "**Qualité**", deux paragraphes qui devraient normalement servir de socle à toute action de référencement :

"Créez des pages pour les utilisateurs, pas pour les moteurs. Ne trompez pas vos utilisateurs, et ne présentez pas un contenu différent aux moteurs de recherche et aux utilisateurs [...]"

"Chaque fois que vous créez une page, posez-vous la question : "Est-elle utile pour mes utilisateurs ? Créerais-je cette page si les moteurs de recherche n'existaient pas ?"

Depuis cinq ans, le fait que les "pages satellites" sont contraires aux instructions des moteurs est donc inscrit noir sur blanc. Ce qui n'a pas empêché de fonder de nombreuses solutions de référencement naturel sur cette technique.

Mais cette optimisation "technique" n'est plus d'actualité : **2006 a marqué l'arrêt de mort des pages satellites**. Après quelques années de tolérance, les principaux moteurs (dont Google) ont déclaré la guerre à ces contenus cachés³. Pour une raison simple : ils sont contraires au principe même des moteurs, qui tentent d'offrir aux utilisateurs les résultats les plus **pertinents** possibles.

² La notion de "pages satellites" recouvre plusieurs pratiques : les pages peuvent être hébergées par le référencieur ou sur le site de l'éditeur ; la redirection peut prendre place lors de la visite de l'internaute ou du robot (on parle alors de cloaking)... Toutes sont à proscrire.

³ La sanction la plus connue est la suppression de l'index de Google des sites allemands de BMW et de Ricoh. Elle a été opérée en février 2006 par la "webspam team" du moteur. (<http://www.matcutts.com/blog/ramping-up-on-international-webspam/>)

Back to the basics

La mort des pages satellites (effective dans la plupart des pays européens, mais qui se traduit en France par une lente agonie) a l'immense mérite de remettre à l'honneur quelques principes simples :

1. Un site n'est pas destiné aux **moteurs**, mais aux **lecteurs**.
2. **Tout doit être visible**.
3. Bref, il faut faire comme si Google n'existait pas...

Les actions de référencement se sont donc recentrées sur le contenu "visible" du site, et sur une optimisation moins mécanique des pages – même si la qualité éditoriale des pages optimisées peut encore être **nettement améliorée**.

Cependant, le référencement traditionnel continue à ne s'intéresser qu'à **une poignée de mots-clés**. Face à la complexité croissante des requêtes saisies par les internautes, c'est se consacrer à l'arbre – et oublier qu'il cache une forêt.

2. Requêtes de « micro-traffic » : les limites de l'approche par mots-clés

Pages satellites ou contenus visibles, l'optimisation traditionnelle repose sur une **liste limitée de mots et d'expressions-clés** considérés comme stratégiques. Leur nombre est en général compris entre dix et cinquante, en fonction des méthodes de référencement – et des budgets.

Cette approche est **indispensable** (il est nécessaire de se positionner sur ce cœur de mots-clés), mais elle se révèle très largement **insuffisante**. Il suffit de se pencher sur les statistiques de fréquentation de votre site pour réaliser que chaque mois, **des milliers de requêtes différentes** participent au trafic global. Ces "requêtes de micro-traffic" représentent un gisement de visites potentielles qui échappe à une démarche par mots-clés.

limiter le travail de référencement à 25 mots-clés, c'est ignorer 80 % du trafic.

Un nombre illimité de combinaisons

Car désormais les internautes saisissent majoritairement **deux ou trois mots** dans les moteurs de recherche, et les requêtes à quatre ou cinq mots, quasi-inexistantes il y a encore deux ans, gagnent du terrain⁴. Une conséquence logique de la croissance exponentielle du nombre de pages Web. Et d'une meilleure maîtrise des outils de recherche par les utilisateurs.

Résultat : le fossé se creuse chaque jour davantage entre **le nombre limité de mots et expressions déterminés lors du référencement**, et le nombre illimité de combinaisons saisies par les internautes.

⁴ Fin 2005, [une étude réalisée par Wyszistat](#) montrait déjà que les requêtes à deux et trois termes représentaient plus de la moitié des requêtes totales, et que les requêtes à quatre termes dépassaient les 10 %.

Analyse d'audience : distinguer trafic de notoriété et trafic naturel

Lorsque vous analysez le trafic issu des moteurs de recherche, il est indispensable de distinguer deux sources :

- **La notoriété.** La requête saisie correspond au nom de votre société ou d'une de vos marques. L'efficacité de cette source est liée à votre politique de communication globale, pas à votre référencement naturel.
- **Le trafic naturel.** Toutes les autres requêtes, non spécifiques à vos marques.

Pour mesurer précisément l'efficacité de votre référencement naturel, vous devez donc **"soustraire" le trafic de notoriété**. Certains outils d'analyse vous permettent de le faire automatiquement.

Comment ignorer quatre visites sur cinq

Une expérience simple consiste à utiliser un outil d'analyse d'audience pour afficher, sur une période donnée (un mois, par exemple), **les visites issues des moteurs de recherche** – et les **requêtes** correspondantes.

Il suffit ensuite de répondre à deux questions :

- 1. Quelle est la part des vingt-cinq premières requêtes (hors notoriété, Cf. encadré) dans le nombre total de visites sur le site ?**
- 2. Combien de requêtes différentes ont-elles été saisies par les internautes au cours de la période, et ont généré au moins une visite ?**

Réponse à la première question ? En fonction des sites, les vingt-cinq premières requêtes représentent entre **5 et 15 % des visites**.

Réponse à la deuxième question ? Chaque mois, **plusieurs milliers de requêtes différentes** sont à l'origine d'au moins une visite⁵.

Optimiser un site à partir d'une liste de vingt-cinq expressions-clés revient donc à laisser de côté 85 % de son trafic. Au minimum.

⁵ Bien sûr, sur ces milliers de requêtes, certaines ne sont que des combinaisons des 25 expressions-clés stratégiques. Mais la plupart sont totalement inédites.

Diversité des requêtes : trois exemples (janvier 2007)

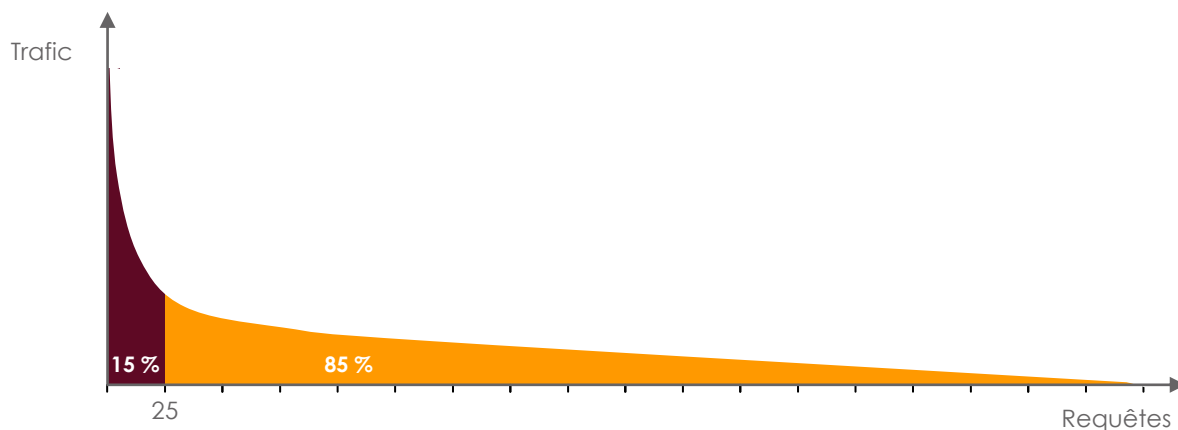
	Cybermarchand	Site généraliste	Site spécifique
Type de site	Site de e-commerce, matériel informatique	Portail d'information multithématique	Site BtoB spécialisé
Audience	Grand public	Grand public	Professionnels
Nombre de visites / mois*	450 000	90 000	10 000
Nombre de requêtes (générant au moins une visite) / mois*	37 000	19 000	4 800
Part des 25 premières requêtes dans les visites totales	2,5 %	7,5 %	14 %

Plus un site est **généraliste**, plus le pourcentage de visite "ignorées" est élevé. Les sites BtoB ou très spécialisés attirent des requêtes plus ciblées, mais les 25 premiers mots-clés concernent toujours moins de 20 % des visites.

Des requêtes rares à l'origine de la majorité du trafic

Les internautes saisissent donc **une incroyable variété de requêtes**, dont la plupart sont par conséquent très rares.

Mais globalement, ce sont ces requêtes "rares" qui fournissent **l'essentiel du trafic** issu des moteurs de recherche. Une illustration de l'effet "Long Tail" mis en lumière par Chris Anderson (voir encadré page suivante) :



De par leur nombre élevé, les requêtes de micro-traffic (moins de 10 visites/mois) génèrent plus de 80% du trafic total

Conséquence : une démarche centrée sur un nombre réduit de mots-clés est donc, sinon inefficace, du moins **inefficiente**. Afin de permettre au site de se positionner efficacement, il faut envisager son univers sémantique de manière **beaucoup plus large**.

La notion de "Long Tail"

En octobre 2004, Chris Anderson, journaliste du magazine en ligne *Wired*, crée le terme de **Long Tail**⁶ pour décrire le modèle commercial du site Amazon.com. Par cette "longue traîne" (Henri Kaufman propose quant à lui "queue de la comète"⁷), Anderson désigne un phénomène nouveau : la majeure partie du chiffre d'affaire du libraire en ligne provient, non pas de ventes importantes sur une poignée de best-sellers, mais **de ventes moyennes ou faibles sur un nombre extrêmement important de références.**

3. Le netlinking, beaucoup d'efforts, peu d'effets

Le terme *netlinking* désigne l'obtention de **liens entrants**, synonymes d'une popularité (ou d'un PageRank, pour ne parler que de Google) élevée. Il constitue le second levier d'un bon positionnement.

Moins de liens, mais plus de qualité

A une démarche quantitative (obtenir, par des moyens plus ou moins licites, des milliers de liens entrants provenant de sites n'ayant souvent aucun rapport entre eux) s'est substituée depuis quelques années **une approche plus qualitative.**

Les moteurs ont fait évoluer les méthodes d'évaluation de la qualité des liens entrants, notamment en conférant automatiquement à des sites "de référence" (universités, ministères, médias...) **un impact très fort** sur la popularité des sites qu'ils relaient.

Cette notion de **crédibilité** (Google utilise le terme de "**TrustRank**") renforce progressivement celle de **notoriété** ("PageRank"), trop facile à détourner. Désormais, l'objectif n'est donc plus d'obtenir un très grand nombre de liens, mais de privilégier ceux en provenance de sites de référence, cohérents avec la page vers laquelle ils pointent.

De l'échange au don

La stratégie traditionnelle pour bénéficier de tels liens est basée sur **l'échange** : l'obtention d'un lien du site A vers le site B est subordonnée à la mise en place d'un lien du site B vers le site A. Une belle idée, mais qui présente de nombreux points faibles :

- Il est de plus en plus difficile de trouver des **"partenaires"**. Plus le domaine est concurrentiel, plus les sites de qualité susceptibles de pointer vers le vôtre se font rares.
- L'échange de liens est **consommateur de temps**. Une fois les liens établis, il faut en assurer un suivi permanent (pour vérifier qu'ils sont correctement implémentés sur le site partenaire, et qu'ils ne disparaissent pas sans raison...).
- Et surtout, les fameux sites "de référence", non commerciaux, sont pour la plupart complètement **inaccessibles** par ce biais : inutile d'espérer voir votre offre de produits relayée – gratuitement – sur le site du quotidien *Le Monde*...

⁶ <http://www.wired.com/wired/archive/12.10/tail.html>. Le principe de la "Long Tail" est une nouvelle déclinaison de la "distribution de Pareto", loi statistique dont les applications sont nombreuses, notamment en économie et en marketing (avec le célèbre principe du "80-20").

⁷ http://henrikaufman.typepad.com/et_si_lon_parlait_marketi/2006/07/la_comete_revien.html

S'il ne faut pas abandonner l'échange de liens, il importe d'en mesurer les limites. Et de réaliser qu'une stratégie de **création de contenus éditoriaux de qualité** sur votre site peut se révéler beaucoup plus efficace pour générer des liens entrants.

C'est le principe du "contenu viral" (les anglo-saxons utilisent le terme "*linkbaiting*") : mettre en place des **contenus de référence**, sans connotation commerciale, afin que des sites externes traitant des mêmes thématiques les relaient **spontanément**. Il peut s'agir de dossiers, d'études de cas, d'interviews exclusives...

La création de contenu viral pallie efficacement les insuffisances de l'échange de liens. Elle est tout aussi consommatrice de ressources, mais ces ressources sont consacrées, au-delà du référencement, à améliorer la qualité globale du site.

Et surtout, le contenu viral permet d'obtenir des liens depuis **n'importe quel site**, aussi prestigieux soit-il. Aucune chance, par exemple, de voir une grande université relayer votre contenu commercial... Mais si vous publiez **une étude de marché exclusive**, ou un **glossaire spécialisé** qui fait référence, votre site peut apparaître parmi ses liens externes.

Oubliez l'achat de liens

Faut-il **payer** pour un lien ?

La réponse est **non**, pour une raison simple : s'il les moteurs s'en rendent compte, ils vont vous pénaliser. Et ils sont très vigilants sur ce point...

Pourquoi ?

Parce que la popularité est au cœur de leurs algorithmes, et qu'il est donc hors de question de "tricher" avec cette notion. Et surtout parce qu'ils souhaitent conserver **l'exclusivité** sur la vente de liens : c'est leur unique source de revenus !

Conclusion : dans le domaine du lien payant, préférez les actions classiques de type **sponsoring, partenariat** ou **achat d'espace**.

Points-clés

- Les techniques traditionnelles de référencement naturel sont confrontées à des **freins structurels**, liés à leur dimension essentiellement technique.
- Essoufflement des **échanges de lien**, disparition des **pages satellites**, limites de l'approche par **mots-clés**... nécessitent de mettre en place une démarche de référencement fondée sur une base sémantique beaucoup plus vaste.

Bon pour le lecteur, bon pour le moteur

Les limites de l'optimisation "mécanique" des pages Web, conséquence d'une approche centrée sur les attentes (supposées) des moteurs de recherche, rendent indispensable **d'inverser la perspective**, et de se placer résolument **du côté du lecteur**.

1. Priorité à l'internaute

Il existe au moins deux bonnes raisons de **privilégier les lecteurs** :

- **Ce sont eux qui font le succès d'un site.** Ce sont les lecteurs, pas les robots, qui effectuent des achats, réagissent aux messages, téléchargent des documents, s'abonnent à une newsletter... Un site bien positionné, mais pensé uniquement pour les robots, bénéficiera d'un fort trafic mais d'un **taux de conversion** déplorable. Il ne suffit pas d'attirer des utilisateurs, il faut également satisfaire leurs attentes en matière de contenu.
- **Le comportement des crawlers se rapproche chaque jour un peu plus de celui d'un lecteur "humain".** C'est le meilleur moyen pour les moteurs de maintenir la pertinence de leurs résultats. Et sur les requêtes les plus sensibles, Google fait d'ores et déjà appel à des équipes de « validateurs » en chair et en os. Il est temps d'oublier les robots !

Conséquence : les bonnes pratiques éditoriales destinées à **faciliter l'accès du lecteur au contenu du site**, à mieux organiser l'information et à la délivrer de manière progressive, seront profitables aux robots des moteurs – et ce, sur le long terme.

Bon pour le lecteur, bon pour le moteur : vous pouvez améliorer **à la fois** l'expérience de vos utilisateurs, et votre référencement naturel.

Pour parvenir à ce résultat, il est indispensable de maîtriser un certain nombre de **règles**, touchant à la **structuration de l'information** et à **l'écriture en ligne**. Elles sont la clé d'un contenu éditorial efficace, qui fera progresser votre site dans l'estime de vos internautes... et dans les pages de résultats.

	Bon pour vos lecteurs...	... Bon pour les moteurs
Structurer l'information		
Organiser les contenus	Navigation et repérage facilités	Parcours et indexation des pages plus efficaces
Multiplier les points d'entrées	Meilleur accès à l'information	
Limiter la profondeur	Pas d'accès fastidieux, ni d'URLs indigestes	Indexation plus rapide et plus fréquente
Respecter la "pyramide inversée"	Accès rapide aux informations essentielles	Indexation complète des contenus, granularité
Faire court		
Créer des contenus concis	Rapidité de consultation	Efficacité de l'indexation, spécificité des pages
Exploiter les niveaux de lecture		
Rédiger des titres de page explicites	Identification rapide de la page, clarté des pages de résultats des moteurs	Mesure de pertinence de la page, premiers éléments du référentiel lexical
Hiérarchiser à l'aide de titres, intertitres... et autres niveaux de lecture	Lecture et compréhension facilitées	Élargissement du référentiel lexical
Soigner les liens		
Rédiger des liens signifiants	Navigation et repérage facilités	Mesure de pertinence de la page cible

2. Structurer l'information

Lecteurs comme moteurs souhaitent **accéder rapidement à l'information qu'ils recherchent**. Cela nécessite de structurer clairement les pages du site, d'en hiérarchiser les contenus et de mettre en œuvre des points d'accès multiples.

Organiser les contenus

Une structuration claire des contenus facilite la **navigation** et le **repérage** de l'internaute au sein du site. Elle lui permet de savoir à tout moment où il se trouve, d'où il vient et où il peut aller.

Par ailleurs, la façon dont **l'architecture de l'information** est perçue par les robots des moteurs de recherche impacte l'efficacité du référencement. Elle leur permet notamment de parcourir plus ou moins efficacement les pages du site, et de les pondérer en fonction de leur position dans la structure globale.


En la matière, une **organisation arborescente** des rubriques et des contenus est généralement idéale, tant pour les moteurs que pour les utilisateurs :

- Elle accompagne progressivement la **découverte des contenus** (la page d'accueil propose des contenus génériques, les pages de niveau 2 des contenus plus précis, et la spécificité augmente au fur et à mesure de la progression dans les niveaux).
- Elle favorise la **contextualisation** (chaque élément "parent" participe à la définition des éléments de niveau inférieur).
- Elle offre une structure adaptée à n'importe quel niveau de **granularité** souhaité (si un thème doit être traité dans le détail, il est toujours possible d'ajouter un niveau inférieur constitué d'un ensemble de pages plus spécifiques).

Exemple : Accompagner une démarche d'achat / vente

Prenons l'exemple d'un site de e-commerce spécialisé "Image & Son", dont le cœur est constitué par un catalogue de produits. Il est fondamental que l'organisation des catégories et des produits épouse aussi bien :

- La **progressivité** de la formulation des besoins, pour l'internaute naviguant à partir de la page d'accueil.
- La **maturité** du besoin tel qu'il peut s'exprimer à travers les requêtes saisies sur les moteurs de recherche.

Étapes de la démarche	Maturité	Requête-type	Formats de contenus Fonctionnalités
Découverte		"acheter", "moins cher", "audio", "vidéo"...	Présentation, Aide en ligne, Conseils, E-magazines
Approche "rayon"		"lecteur DVD", "téléviseur LCD", "baladeur MP3"...	Arborescence des catégories (branches)
Approche par besoin, comparaison de produits		"lecteur DVD / DivX", "téléviseur LCD 82 cm", "baladeur Creative 4Go"...	Arborescence des catégories (feuilles) Filtres par marques, caractéristiques, prix...
Produit identifié		"Lecteur DVD LG DVX9900H", "TV LCD SAMSUNG LE32R71", "Creative ZEN V 4Go"...	Fiche produit, caractéristiques complémentaires, contextuelles

Une organisation arborescente des catégories et produits s'inscrit parfaitement dans cet objectif.

Multiplier les points d'entrée

Si elle présente de nombreuses vertus, l'arborescence comporte deux inconvénients principaux :

- D'une part, **elle limite les chemins d'accès à l'information**, en ne permettant pas de prendre en compte la diversité des concepts et points de vue qui peuvent mener à un contenu, via la navigation ou les requêtes saisies sur les moteurs de recherche.
- D'autre part, elle ne peut être bâtie sans **choix de classement parfois arbitraires**, qui masquent les différentes acceptions que peut revêtir un thème.

Il convient donc de considérer l'arborescence comme la **colonne vertébrale** des contenus (autrement dit, le classement principal), à laquelle il faut adjoindre des dispositifs permettant de multiplier :

- Les **moyens d'accès** (par un **moteur de recherche interne**, par thèmes, par marques, par caractéristiques, par prix...).
- Le **maillage** (établissement de liens connexes, permettant de naviguer parmi les pages du site de manière **transversale**).

Limiter la profondeur

Pour l'internaute, un contenu "profond" signifie **un accès fastidieux** (plusieurs clics) ou la saisie d'une URL peu digeste.

Par ailleurs, à partir d'un certain niveau, la profondeur de l'arborescence peut constituer **un frein à l'efficacité du référencement** et de la navigation. Google, par exemple, indexe moins fréquemment les contenus les plus "profonds", au-delà des niveaux 4 ou 5.

Pour cette raison, il est nécessaire de permettre **un accès plus rapide** aux derniers niveaux de l'arborescence (tant via la navigation qu'à travers l'URL), voire de "court-circuiter" celle-ci, lorsqu'il s'agit de **contenus stratégiques**.

Respecter la "pyramide inversée"

La structuration de l'information, fondamentale au niveau global du site, doit également s'appliquer **au niveau spécifique de la page**.

La "pyramide inversée" est un principe journalistique qui consiste à placer les informations importantes **en tête d'un article**, avant de les aborder de manière plus détaillée dans le corps de celui-ci.

Cette progression **linéaire**, qui atteint sa plus pure expression dans les journaux anglo-saxons, s'oppose à la construction classique des articles de presse écrite (accroche, développement plus ou moins hiérarchisé, "chute").

Elle est particulièrement adaptée au sein des pages Web, car :

- Les lecteurs sont **pressés**. Ils souhaitent accéder rapidement à l'essentiel de l'information, et n'approfondir qu'une fois qu'ils sont sûrs d'avoir identifié un sujet qui les intéresse.
- Si la plupart des crawlers indexent l'intégralité des contenus texte d'une page Web, ils en surpondèrent en général **les premières lignes** : c'est donc là que doivent figurer les informations importantes.

Accessibilité et référencement

Il existe une abondante littérature consacrée aux solutions de développement les mieux adaptées au référencement. Inutile en effet d'initier une stratégie de référencement naturel, si pour des raisons techniques (utilisation de frames, développement entièrement en Flash) votre site ne peut pas être correctement indexé par les robots.

Un critère simple permet de savoir si un site est optimisé techniquement pour le référencement : son **accessibilité** aux personnes handicapées.

Un site adapté aux lecteurs vocaux destinés aux malvoyants est idéal pour les crawlers. On oublie trop souvent que **les robots sont aveugles...**

Plus largement, **un site qui respecte strictement les critères d'accessibilité édictés par le W3C⁸ bénéficiera de la meilleure indexation possible.**

⁸ <http://www.w3.org/WAI/>

3. Faire court

La lecture est plus difficile sur un écran que sur le papier. Zappeurs compulsifs, les internautes apprécient donc les pages et les textes **courts**, qu'ils peuvent parcourir rapidement. Sur le Web, la **concision** est une forme de politesse envers le lecteur.

Dans le même temps, les *crawlers* privilégiant les premières lignes de texte, une page trop longue ne sera indexée qu'imparfaitement. De plus, des pages courtes permettent une maîtrise plus fine des **thématiques** abordées, et donc du référentiel lexical.

Conséquence : **les contenus courts doivent être privilégiés**, et lorsqu'il est nécessaire de développer une thématique, il est conseillé de segmenter le texte initial sur plusieurs pages (relayées dans une page de sommaire).

La recherche de la concision s'inscrit dans le droit-fil d'une **structuration** efficace du site, qui a pour conséquence des pages brèves et spécifiques.

Des phrases courtes, regroupées en paragraphes courts, au sein de pages courtes... facilitent le parcours des lecteurs, qu'ils soient humains ou robots.

4. Exploiter les niveaux de lecture

Sont regroupés sous les termes "niveaux de lecture" l'ensemble des **mises en valeur** du texte : titres et intertitres, gras... Ils facilitent le **parcours** de la page, en offrant à l'internaute des points de repères lorsqu'il la "scanne" en quelques secondes, à la recherche d'informations précises. Et ils sont prisés des moteurs, qui, encore une fois, les surpondèrent lors de leurs calculs de pertinence.

La rédaction d'une page bien **structurée**, offrant de multiples niveaux de lecture **hiérarchisés** et **riches de sens**, engendrera donc naturellement un code HTML favorable au référencement.

Titre de page

Pour le lecteur, le titre de la page (qui s'affiche dans la barre supérieure de la fenêtre du navigateur) joue un rôle dans la manière dont les moteurs de recherche affichent leurs **résultats**. Souligné et bénéficiant d'une taille de police plus importante, ce titre est le premier niveau d'information affiché ; il est généralement complété par le contenu de la balise de description.

Le titre de la page est également utilisé pour la désigner, lorsque le lecteur la place parmi ses "favoris".

Pour les robots, le contenu de cette balise, placée dans les premières lignes du code HTML de chaque page, **est un des éléments déterminants en termes de référencement**.

Stratégique tant pour les moteurs que pour les lecteurs, le titre de la page doit donc être rédigé avec soin, afin de **décrire** clairement le contenu de la page et d'en relayer les principaux **mots-clés**, tout en restant **concis**. La **spécificité** est de mise : un titre de page comme "Produits - toutes nos références" n'apportera aucune information concrète au lecteur, et aucun élément exploitable au *crawler*.

Chaque page du site devra par ailleurs bénéficier de son propre titre, résumant son contenu et incluant des mots-clés spécifiques.

Dans ce domaine aussi, une **structuration** cohérente du site permet d'obtenir des titres efficaces, par "héritage" :

- Page d'accueil : [Objet du site (champ lexical générique)] - [Nom de la société]
- Page carrefour : [Objet de la rubrique (champ lexical plus spécifique)] - [Nom de la société]
- Page de niveau 2 : [Objet de la page (champ lexical plus spécifique)] - [Nom de la rubrique] - [Nom de la société]

Titres et sous-titres

On considère qu'un bon titre doit être à la fois **informatif** (il doit résumer en quelques mots le texte qu'il précède) et **incitatif** (il doit donner envie de poursuivre la lecture, par l'usage de l'humour, du second degré ou d'une référence culturelle).

Mais dans une optique de référencement naturel, **le curseur doit être résolument placé du côté de l'informatif**. Pourquoi ?

Les lecteurs sont **pressés**. Ils scannent les pages, et cherchent à en appréhender le contenu en une fraction de seconde. Un titre qui ne se comprend pas immédiatement est donc considéré comme un frein.

Les crawlers accordent beaucoup d'importance aux titres (lors de l'indexation des pages, ils accordent un "poids" plus importants aux termes compris entre les balises HTML <H1>, <H2>, etc.)... et ils sont totalement imperméables au second degré.

Les titres ne doivent pas seulement décrire les contenus qu'ils précèdent, ils doivent également être **spécifiques**, c'est-à-dire contenir les mots-clés relatifs à la page. Et ils doivent être **concis**, autant pour mettre en valeur les mots-clés qu'ils contiennent que pour satisfaire le lecteur pressé.

Les sites de presse, incitation vs référencement

Les journaux sont, par tradition, portés aux titres incitatifs (qui sont souvent une véritable marque de fabrique).

Mais l'impact croissant du référencement naturel sur la fréquentation des sites de presse soumet les rédactions à un dilemme : doivent-elles désormais **préférer les titres informatifs**, plus adaptés aux robots d'indexation, au risque d'uniformiser leur discours et de décevoir leurs lecteurs ?

Le débat est loin d'être clos, et une **solution mixte** est déjà testée par certains sites, comme celui de BBC News : des titres incitatifs en page d'accueil ("Le tourisme de la fertilité"), et plus factuels en pages intérieures ("Témoignage : une fécondation in vitro à Istanbul")? !

Exemple

Sur le site d'un opticien, le titre "Notre gamme" est suffisamment informatif pour l'internaute, qui déduit de l'environnement qu'on va lui présenter des lunettes. Mais il n'a aucune valeur en référencement naturel (le terme "gamme" pouvant difficilement être considéré comme un mot-clé stratégique) ; il faudra donc lui préférer "Notre gamme de lunettes", ou "Toutes nos lunettes", plus concis.

⁹ Article d'Helen Briggs, édition du 19 décembre 2006

La rédaction de titres efficaces en termes de référencement naturel privilégiera donc la **spécificité** et la **concision**. Ce n'est qu'une fois ces critères respectés qu'il est possible de mettre en œuvre des éléments incitatifs. "Nos lunettes en un coup d'œil", par exemple.

Intertitres, gras

Le texte mis en gras et les niveaux de titre intermédiaires, sont autant d'étapes dans le parcours visuel de l'internaute.

Les crawlers estiment que ces termes mis en **exergue** (c'est-à-dire situés entre des balises HTML spécifiques) ont plus de valeur que les autres, et ils leur donnent plus d'importance lors de l'indexation.

Il est donc souhaitable de résumer chaque paragraphe par un **intertitre** (respectant les principes des titres, évoqués plus haut – information, spécificité, concision) et d'en mettre en valeur les éléments les plus significatifs (et en particulier les mots-clés) en usant du **gras**.

L'optimisation des PDF : une nécessité ?

Désormais, les principaux moteurs de recherche (Google, Yahoo!, Live Search...) indexent les documents PDF.

Sur un site dont le contenu est largement constitué de PDF proposés en téléchargement, il est donc nécessaire :

- Soit de retravailler le référentiel lexical de chacun d'entre eux.
- Soit de les relayer via **une page HTML optimisée**.

5. Soigner les liens

Les liens constituent le fondement même du Web. Ils permettent aux internautes de circuler au sein des pages d'un site, ou d'un site vers un autre.

Éléments clés dans le référencement d'une page Web, les liens hypertextes sont exploités par les robots pour évaluer la **pertinence** de la page qui les contient... et de celle **vers laquelle** ils pointent. Leur texte doit donc être cohérent avec le champ lexical de la page qu'ils désignent.

Un bon lien est **prédictif** : son texte annonce clairement le contenu vers lequel il pointe. Cela évite au lecteur la frustration qu'engendre le fait de cliquer sur un lien peu précis, puis d'attendre le chargement de la nouvelle page avant de constater que son contenu ne correspond pas à ses attentes. Le texte du lien doit, pour être prédictif, respecter les principes **d'information** et de **concision** qui s'appliquent à la rédaction des titres.

Afin de rédiger des liens **courts**, il est souvent intéressant de les **contextualiser**, de les **thématiser**. Ils peuvent ainsi être regroupés dans une rubrique "Tous les dossiers", ou "Documents à télécharger".

Une règle absolue : éviter les "[Cliquez ici](#)"... sauf si vous considérez que le terme "cliquez" est stratégique pour votre référencement naturel¹⁰.

Points-clés

- Parce que votre site est destiné à vos internautes, il doit leur offrir une information claire, nettement **hiérarchisée** et facilement **accessible**.
- Ces attentes sont également celles des robots : le respect de critères **structuraux** et **éditoriaux** précis permet donc de satisfaire **à la fois lecteurs et moteurs**.

¹⁰ Pour information, le site classé en première position par Google sur la requête "cliquez" est [un convertisseur de devises](#) (sur les pages duquel le terme "cliquez" n'apparaît qu'une seule fois). La raison ? C'est un outil pratique, relayé par plus de 18 000 sites externes... Souvent sous la forme d'un lien "Cliquez ici" !

Vers le référencement global

Le respect des règles détaillées dans le chapitre précédent assure au contenu **existant** de votre site un référencement optimal. Mais le véritable enjeu est de voir au-delà de cet existant, de développer les contenus du site afin de coloniser un **univers sémantique plus large**. Il existe des continents entiers de requêtes qui attendent d'être conquis.

1. Une approche plus large de l'univers sémantique

Chaque jour, la majorité des internautes qui parvient sur votre site a saisi dans les moteurs de recherche des requêtes auxquelles vous n'auriez jamais songé. Inutile, donc, d'essayer d'optimiser vos pages en fonction d'expressions définies. La seule solution est **d'étendre au maximum votre univers sémantique**, pour capter un nombre croissant de ces visiteurs imprévisibles.

- Première étape : la **définition** de cet univers, en étudiant de manière très étendue les attentes de vos internautes ; puis sa traduction sous forme de référentiel lexical (système de mots et d'expressions clés organisés **en grandes thématiques**).
- Deuxième étape : le développement de contenus traitant ces thématiques, et jouant sur une grande variété de **formats**.

Puiser à toutes les sources

Lors de la définition de l'univers sémantique d'un site Web (que ce soit lors de sa création ou de sa refonte), un simple brainstorming ne suffit pas.

Pour viser l'**exhaustivité**, les sources internes sont très largement insuffisantes. Elles ne déboucheront que sur une parcelle de l'univers sémantique, **propre à l'entreprise**, avec un vocabulaire et des codes parfois très éloignés de ceux des internautes – et de leurs requêtes.

Sur le Web, répondre aux attentes des utilisateurs est une loi cardinale. Il est donc primordial d'exploiter un maximum de sources, **internes** mais aussi – et surtout – **externes**.

1. Moteur de recherche interne

Si le site dispose de son propre moteur de recherche, il constitue une véritable mine d'informations sur les attentes des internautes. Il permet notamment de détecter des "**lacunes sémantiques**" : des requêtes saisies régulièrement, mais auxquelles aucune page du site ne correspond.

2. Outils d'analyse de fréquentation

Les outils d'analyse d'audience (Xiti, Google Analytics...) autorisent également une étude très fine des requêtes réelles des internautes. Ils permettent de mettre à jour des **décalages** entre l'offre en contenu de votre site et les attentes des utilisateurs, par exemple :

- Une requête très courante, générant un fort trafic sur une page de votre site mal valorisée : pas d'incitation, de liens vers d'autres pages ou de documents à télécharger... (il est alors primordial d'améliorer le contenu de cette page).
- Des pages entières de votre site n'enregistrant aucun trafic depuis les moteurs, malgré leur dimension stratégique (il convient de se demander si elles ne sont pas stratégiques uniquement pour l'entreprise, ou si c'est simplement leur champ lexical qui est mal adapté).

3. Appels téléphoniques et mails entrants

Si vous disposez d'un SAV ou d'un service d'assistance téléphonique, il est souvent intéressant d'analyser les requêtes les plus courantes et leur **formulation** – afin de vérifier que le site y répond, et que son champ lexical est adapté au langage employé par les utilisateurs.

De même, si votre site propose un **forum**, sa consultation peut être riche d'enseignements.

4. Concurrents

Les sites concurrents constituent une source d'inspiration qu'il ne faut pas négliger. Ils offrent un moyen supplémentaire de révéler des lacunes sémantiques (des termes employés sur leur site et pas sur le vôtre) ; et un **benchmark éditorial** peut également être intéressant lors de l'étape suivante, pour détecter des formats de contenus innovants ou des thématiques inédites.

5. Contradicteurs

Si l'activité de votre site touche à des domaines **sensibles** (environnement, santé publique, sécurité alimentaire, propriété intellectuelle...), il existe probablement un réseau très organisé, entièrement consacré à délivrer un message assez différent du vôtre. Il est indispensable d'étudier **l'univers sémantique de ces contradicteurs**, d'une part parce qu'il reflète les attentes des internautes, d'autre part parce que vous devez absolument y positionner votre site (voir encadré).

Des contradicteurs bien référencés

Avec le développement du Web 2.0, le principal rival en termes de référencement n'est plus forcément le site du concurrent, identifié et dont les évolutions sont prévisibles. Il peut désormais s'agir d'une multitude d'acteurs isolés et incontrôlables : le blog d'un ex-employé mécontent, un article publié par le site d'une association et épinglant des pratiques de discrimination au sein de l'entreprise, une vidéo tournée avec un téléphone portable démontrant le dysfonctionnement d'un produit¹¹...

Les actions de référencement initiées par les entreprises doivent désormais tenir compte de ces nouveaux acteurs, ce qui passe notamment par :

- Le renforcement des contenus institutionnels (pour faire du site un espace de référence pour toute information touchant à l'entreprise) et une transparence accrue.
- Une veille permanente sur ces nouveaux médias.
- La création de contenus traitant des sujets conflictuels, afin d'être référencé aussi bien – et si possible, mieux – que ses adversaires sur ces thèmes (notamment en cas de communication de crise).
- L'exploitation des médias CtoC, par des interventions ciblées, la mise en place d'un blog d'entreprise ou l'implication de blogueurs influents (des actions délicates à mettre en œuvre, mais qui peuvent se révéler payantes).

Penser en trois dimensions

Une fois déterminées les attentes de vos utilisateurs, reste à extraire de cet univers sémantique le **référentiel lexical** du site : une série d'idées, de mots et d'expressions clés (identifiés lors de l'étape de définition), organisés en **grandes thématiques**.

Trois axes permettent d'étendre ce référentiel lexical, et de définir dans leurs grandes lignes les formats de contenus du site.

¹¹ L'exemple le plus célèbre est celui de Sony, dont les bénéfices nets se sont effondrés lors du troisième trimestre 2006 suite à un problème survenu avec des batteries pour ordinateur portable fabriquées par le constructeur japonais. Relayée par le site YouTube, une vidéo montrant un portable en flammes a fait le tour du Web. <http://www.pcinpact.com/actu/news/32353-Sony-Nintendo-benefices.htm>

1. Traiter en profondeur

Sur votre cœur de métier, vous devez viser **l'exhaustivité**. Il est donc nécessaire d'approfondir au maximum le référentiel lexical correspondant à votre offre et à vos principaux messages, et de leur consacrer **des contenus détaillés**.

Il peut s'agir de livres blancs, de fiches techniques, d'études de cas... A condition de structurer efficacement l'information, vous ne fournirez jamais **trop** de précisions sur votre activité.

Exemple : adapter son référentiel lexical au cycle de vente

Dans le cas d'un site marchand, il est primordial de développer l'univers sémantique de l'offre produit en respectant les différentes étapes de l'acte d'achat. Il faut toujours tenir compte du fait que l'internaute effectue souvent **une série de recherches successives**, de plus en plus spécifiques.

Étape	Exemples de requêtes (l'utilisateur cherche une solution d'affacturage)
Découverte	Affacturage Solution d'affacturage Étude de cas affacturage ...
Intérêt	Consultants affacturage Spécialistes affacturage Cabinets affacturage ...
Considération	Simulation affacturage BNP affacturage Natexis affacturage ...
Acquisition	Contrat d'affacturage Contact affacturage Devis affacturage

2. Étendre le territoire de communication

La parfaite maîtrise de votre domaine d'activité est bien sûr indispensable. Mais pour élargir au maximum le périmètre de votre référencement, il est nécessaire de communiquer **au-delà de vos produits et de votre offre**. Et de vous demander sur quelles **thématiques** relatives à votre activité (mais sans rapport direct avec vos produits et votre discours marketing) vous pouvez communiquer **légitimement**.

Vous pouvez toujours développer ce type de thématiques connexes, qu'elles soient issues de votre **expérience métier**, de votre **veille concurrentielle** ou des **connaissances** de vos collaborateurs.

Exemples

- Un spécialiste de l'électroménager haut de gamme anime une chronique consacrée au design ou à l'art de vivre.
- Une banque en ligne commente l'actualité économique par le biais de ses conseillers financiers.
- Un site de vente de *street wear* propose des portraits de rappeurs.
- Un constructeur automobile consacre un dossier aux biocarburants...

3. Multiplier les angles

Une fois le référentiel lexical approfondi et étendu, il lui manque encore une dimension : celle qu'apporte la variété des points de vue. En donnant la parole à des **clients** ou à des **tiers** (experts, partenaires...), vous enrichissez le champ lexical, et **vous en reformulez les termes** sous des formes plus proches des requêtes réelles des internautes.

Une étude de cas illustrée par les témoignages de vos clients, un forum destiné aux utilisateurs de vos logiciels, un blog permettant aux internautes de laisser des commentaires sur vos derniers produits... permettent de faire entendre sur votre site **une voix différente**, plus conforme à celle de votre audience.

Variation des formats

Le référentiel lexical de votre site est parfaitement défini et exploité ? Les thématiques abordées sont déterminées ? Reste à les traduire sous forme de **contenus** concrets.

Il est nécessaire de respecter un certain nombre de **critères éditoriaux** (définis de manière détaillée dans le chapitre précédent) : information structurée, concision, titres informatifs... Mais vous devez dans le même temps tirer parti de la **multiplicité de format qu'offre le Web** : article, dossier, fiche produit, interview, quiz...

Cela permet à la fois de procurer aux internautes **une expérience de lecture riche et variée**, et de faire vivre le référentiel lexical, en répétant les expressions stratégiques sans redondances.

Il faut donc trouver le point d'équilibre entre rigueur structurelle et liberté formelle.

2. Une démarche dynamique

L'univers sémantique d'un site est amené à **évoluer**, pour s'adapter à de nouvelles attentes des internautes, aux nouveaux messages de l'entreprise, ou aux mutations du marché... Et à leur impact sur le référencement.

Le système doit rester **ouvert**, en contact permanent avec son environnement ; et être régulièrement **actualisé**.

Se maintenir au cœur du système

Un site Web doit être considéré dans un contexte plus vaste ; un **écosystème** avec lequel il interagit en permanence, par le biais, notamment, de liens sortants et entrants. Ces derniers jouent un rôle majeur dans le référencement naturel, il est nécessaire de favoriser leur obtention, sur le long terme.

Pour se positionner en tant qu'acteur **pérenne** de cet écosystème, la proactivité est de mise. Le site doit en particulier :

- Proposer des **contenus de référence**, utiles à tous (c'est le principe du contenu viral, évoqué plus haut).
- **Intervenir activement** dans la vie du système (par le biais d'interventions sur des blogs, de posts sur des forums, d'alimentation d'un fil RSS...).

Soutenir l'attention

Autre impératif : l'actualisation régulière des contenus – et la création de contenus nouveaux.

Des contenus régulièrement mis à jour permettent :

- De fidéliser les lecteurs et d'améliorer l'image du site (un contenu actualisé est l'un des principaux critères qualitatifs cités lors des tests utilisateurs).
- De maintenir l'attention des moteurs (qui indexent à un rythme plus soutenu les sites régulièrement actualisés).
- D'adapter en permanence le référentiel lexical du site – et donc son référencement.

Cette mise à jour peut être effectuée **à des rythmes différents** ; le taux de renouvellement étant proportionnel à la spécificité du référentiel lexical :

Formats de contenu	Actualisation	Référentiel lexical
Articles de fond, présentation de l'activité	Faible	Générique
Dossiers, e-magazine...	Régulière	Spécifique
Actualités, communication de crise...	En temps réel	Stratégique

3. Se fixer des objectifs et vérifier qu'ils sont atteints

Une fois des contenus couvrant l'ensemble de l'univers sémantique mis en place, et leur rythme d'actualisation déterminé... Reste à valider **leur impact réel** sur votre référencement.

Si **l'analyse de ROI** est parfaitement intégrée lors des campagnes de liens sponsorisés, elle fait encore trop souvent défaut en référencement naturel. C'est pourtant une étape stratégique, qui permet à la fois :

- De **valider la pertinence du dispositif éditorial** mis en place et du budget alloué.
- De **réajuster**, si besoin, le référentiel lexical.
- De **développer des synergies** avec les autres formes de référencement ou de communication, online mais aussi offline.

Il existe de nombreux outils permettant non seulement de déterminer le trafic issu du référencement naturel, mais également de lier **chaque visite** à l'atteinte d'un objectif, qui peut être :

- **Commercial** : achat, réservation, prise de contact avec un conseiller...
- **Qualitatif** : durée de visite, nombre de pages vues, téléchargement d'une brochure, inscription à une newsletter...

Il est ainsi possible de mesurer **la part du référencement naturel dans le développement global du site**.

Les outils d'analyse d'audience permettent également de détecter si tous les objectifs assignés au site sont atteints et si des ajustements sont nécessaires. Il peut s'agir de la création **de nouveaux contenus**, pour renforcer l'efficacité du référencement à moyen terme sur un objectif précis ; ou de la mise en place d'une campagne complémentaire de **liens sponsorisés**, pour un effet à court terme.

Mais la logique du ROI, si elle est indispensable, ne doit pas faire oublier que l'impact d'une approche éditoriale du référencement ne se limite pas à l'amélioration du chiffre d'affaires.

Un contenu de qualité **fidélise** les internautes (et augmente les taux de conversion), améliore **l'image** du site et sa **visibilité**, permet d'étendre son **territoire de marque**... Des effets positifs qui dépassent, de très loin, le seul domaine du référencement.

Glossaire

Accessibilité

L'accessibilité d'un site Web désigne le respect d'un certain nombre de critères techniques (définis par le W3C) et éditoriaux destinés à faciliter l'accès du site aux personnes porteuses de handicap (déficients visuels, notamment). Un site respectant à la lettre ces critères bénéficie par ailleurs d'une **indexation** parfaite par les **robots**, indispensable à un référencement efficace.

Actualisation / animation éditoriale

La mise à jour régulière d'un site, et notamment la création de nouveaux contenus éditoriaux, est considérée par les internautes comme l'un des principaux critères qualitatifs. Cette actualisation a également un impact positif sur le référencement du site, les **robots** indexant à un rythme plus élevé les pages fréquemment mises à jour.

Algorithme

Dans le domaine du référencement, les algorithmes des moteurs de recherche déterminent les résultats affichés lors de la saisie d'une **requête**. Tenus secrets, ces algorithmes se basent sur une multitude de critères (plus d'une centaine pour Google) : contenu de la page, **liens entrants**, URL...

Analyse d'audience / Web Analytics

Les outils d'analyse d'audience mesurent le trafic sur un site Web et fournissent de nombreux indicateurs : origine du trafic, pages vues, **requêtes** saisies... Ils permettent notamment de déterminer la part du **référencement naturel** dans le ROI d'un site.

Backlink

Voir "Lien entrant"

Benchmark éditorial

Analyse comparative des contenus éditoriaux de plusieurs sites d'un même univers concurrentiel, ce benchmark permet de détecter les bonnes pratiques et d'enrichir le **référentiel lexical**.

Cloaking

Technique désormais bannie par les moteurs de recherche, le cloaking ("masquage") permet de fournir aux **robots** une page Web "optimisée", différente de celle proposée aux lecteurs humains.

Contenu viral

Un moyen efficace d'obtenir spontanément des **liens entrants** (en parallèle du **netlinking**) consiste à mettre en œuvre des contenus de référence, inédits ou polémiques, susceptibles d'être relayés par des médias, des institutions ou la blogosphère.

Crawler

Voir "Robot"

CtoC

Construite sur le même mode que BtoB et BtoC, cette abréviation de "Customer to Customer" désigne l'ensemble des médias participatifs ou communautaires, permettant un échange d'informations directement entre les internautes, sans entremise d'un annonceur : **blogs**, forums, communautés en ligne...

Densité (indice de)

Dans le texte d'une page Web, l'indice de densité d'un terme désigne le rapport entre l'occurrence de ce terme et le nombre total de mots. C'est l'un des critères utilisés par les moteurs de recherche pour déterminer la pertinence des pages - mais son respect trop strict entraîne la rédaction de textes peu attractifs pour les lecteurs.

Granularité

Dans le domaine scientifique, la granularité désigne la taille du plus petit élément d'un système, au-delà de laquelle on ne peut plus découper l'information. Dans un site à structure arborescente, la granularité désigne la page de dernier niveau, où se trouve une information très spécifique.

Indexation

Inscription, par les **robots** d'un moteur, des pages d'un site dans sa base de données (index). L'indexation est le préalable indispensable au **positionnement** des pages : une page non indexée ne peut pas apparaître dans les résultats des moteurs.

Lien entrant

Sur une page Web donnée, est considéré comme "entrant" tout lien provenant d'une autre page, qu'elle appartienne au même site (lien entrant interne) ou à un autre (lien entrant externe). Le nombre de liens entrants d'une page, et plus encore la qualité des pages dont ils proviennent, sont utilisées par les moteurs pour en déterminer la **popularité** et la pertinence.

Lien sponsorisé

Lien provenant d'un achat de **mot clé** auprès d'un moteur de recherche, et qui apparaît dans une zone spécifique de la page de résultats.

Link baiting / Linkbaiting

Voir "Contenu viral"

Long Tail

"Longue traîne" ou "queue de la comète", ce terme forgé en 2004 par le journaliste Chris Anderson désigne la structure de vente de certains cybermarchands (Amazon, notamment). Contrairement aux enseignes traditionnelles, dont la majorité du chiffre d'affaire est concentrée sur un nombre réduit de références, Amazon réalise plus de la moitié de ses ventes grâce à un catalogue très étendu. On retrouve la Long Tail lors de l'analyse du trafic issu des moteurs de recherche : la majorité de celui-ci provient de requêtes générant peu de visites (**micro-traffic**), mais dont le nombre est très élevé.

Micro-traffic

Plus de 80% du trafic d'un site provient d'un nombre très élevé de requêtes "rares", générant chacune peu de visites. Ces requêtes de micro-traffic constituent un gisement que les techniques traditionnelles de référencement naturel (et en particulier l'approche par **mot-clé**) exploitent très imparfaitement.

Mot-clé

Élément de base des stratégies traditionnelles de référencement naturel, le mot (ou l'expression) clé n'offre qu'une visibilité et une pertinence très limitées. Une approche plus large, basée sur un référentiel lexical, permet de pallier ces limites.

Netlinking

Ensemble des actions visant à obtenir des **liens entrants** : échange, achat, inscription dans des annuaires... Il est de plus en plus difficile d'obtenir des liens entrants de qualité (c'est-à-dire provenant de sites de référence) ; la création de **contenu viral** s'impose donc souvent comme un complément indispensable au netlinking.

Niveaux de lecture

Titres et intertitres, texte en gras, listes à puces... sont autant d'éléments qui facilitent la lecture d'une page Web, et l'accès aux informations essentielles. Apprécié des internautes, ces niveaux de lecture sont également surpondérés par les **robots** lors de l'**indexation** des pages.

PageRank

Voir "Popularité"

Page satellite

Optimisées "mécaniquement" pour le référencement, les pages satellites sont des pages Web réservées aux seuls **robots** (les visiteurs humains sont automatiquement redirigés vers une autre page du site). Destinées à améliorer de manière totalement artificielle le **positionnement** du site qui les héberge, elles sont désormais bannies par les moteurs.

Popularité

La popularité d'une page Web est fonction du nombre de liens pointant vers cette page, et de la propre popularité des pages dont ces liens proviennent. C'est un critère déterminant, utilisé par les moteurs de recherche lors de la détermination du **positionnement** d'une page dans leurs résultats. L'indice de popularité le plus célèbre est celui de Google, baptisé PageRank. (voir aussi "Trustrank")

Positionnement

Le positionnement d'une page Web désigne son rang dans les résultats des moteurs de recherche, lors de la saisie d'une **requête** spécifique. Une page positionnée au-delà des trente premiers résultats générera un trafic quasi-nul.

Pyramide inversée

Principe d'organisation de l'information consistant à placer les éléments essentiels "au sommet", puis à offrir des niveaux d'information de plus en plus spécifiques. Il s'applique aussi bien au niveau d'une page Web (concentration de l'information dans le titre, le chapeau et les premiers paragraphes) que

d'un site dans son ensemble (l'information stratégique est accessible dès la page d'accueil, et devient de plus en plus spécifique à mesure que l'on parcourt l'arborescence).

Quality Score

Mis en place par Google en 2006, le Quality Score est une méthode d'évaluation de la pertinence des **liens sponsorisés**, basée sur leur intitulé et sur le contenu de la page vers laquelle ils pointent. Un système comparable est mis en œuvre dans la plate-forme publicitaire *Panama* de Yahoo!

Référentiel lexical

Le référentiel lexical d'un site Web désigne un système d'expressions et de termes organisés en grandes thématiques stratégiques. Beaucoup plus efficace et pertinent qu'une simple liste de mots-clés, il devrait constituer la base de toute action de référencement naturel.

Requête

Ensemble de termes saisis par l'internaute dans un moteur de recherche. Longtemps cantonnées à un ou deux termes, les requêtes sont désormais de plus en plus complexes, ce qui limite l'efficacité d'un référencement naturel basé sur une simple liste de **mots-clés**.

Robot

Ce logiciel d'indexation créé par les moteurs de recherche (également appelé *crawler* ou *spider*) parcourt le Web en suivant les liens hypertexte et enregistre le contenu texte des pages qu'il rencontre. C'est ce contenu qui permettra de déterminer la pertinence des pages et leur **positionnement** dans les résultats des moteurs.

SEA

Le terme de *Search Engine Advertising* désigne les actions de référencement payant, et notamment l'achat de liens sponsorisés. Le SEA constitue, avec le **SEO**, les deux versants du **SEM**.

SEM

Le *Search Engine Marketing* désigne un ensemble de méthodes destinées à augmenter la visibilité d'un site dans les pages de résultat des moteurs de recherche. Il comporte deux aspects, les liens sponsorisés (**SEA**, *Search Engine Advertising*) et les actions de référencement naturel (**SEO**, *Search Engine Optimization*).

SEO

On regroupe sous le terme de *Search Engine Optimization*, ou référencement naturel, les actions destinées à améliorer le positionnement d'un site dans les résultats organiques des moteurs de recherche, et l'obtention de liens entrants. Le SEO constitue, avec le **SEA**, les deux versants du **SEM**.

Spider

Voir "Robot"

Titre de page

Dans le code HTML d'une page Web, le titre de page est le texte situé dans la balise spécifique, <title>. Il apparaît en haut de la fenêtre du navigateur et dans les pages de résultat des moteurs de

recherche, et c'est l'un des éléments les plus importants lors de l'**indexation** des pages d'un site. Il doit donc répondre à des critères éditoriaux et structurels précis.

TrustRank

Ce terme déposé par Google désigne le résultat d'un **algorithme** destiné à mesurer la crédibilité des pages Web. Contrairement au **PageRank**, essentiellement quantitatif, le TrustRank est qualitatif et privilégie les liens en provenance de sites considérés comme faisant référence.

Web 2.0

Concept aux contours flous, le Web 2.0 désigne (a minima) une nouvelle génération d'outils (fils RSS, blogs...) et de pratiques (création d'espaces communautaires, *taggage* de contenus...) visant à centrer les fonctionnalités d'Internet sur ses utilisateurs. Parce qu'il offre une conception radicalement nouvelle de la création et de la diffusion des contenus, le Web 2.0 influence fortement les méthodes de référencement naturel.

Pour en savoir plus :



Tél. : 01 49 66 60 60

E-mail : contact@dixxit.fr

42, rue d'Arthelon - 92190 Meudon

www.dixxit.fr