HAL
open science

Le dilemme du contenu dans le référencement : entre moteurs et internautes

Johan Sellitto

To cite this version:

Johan Sellitto. Le dilemme du contenu dans le référencement : entre moteurs et internautes . Sciences de l'information et de la communication. 2010. dumas-01691455

HAL Id: dumas-01691455

https://dumas.ccsd.cnrs.fr/dumas-01691455v1

Submitted on 24 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Johan Sellitto

MASTER 2 ICD, spécialité IDEMM

(Ingénierie du Document et Médiation Multimédia)

MEMOIRE DE STAGE

Mission effectuée du 1er avril au 30 septembre 2010
à l'agence de référencement Aeronet
35 rue Delos – 59810 Lille

LE DILEMME DU CONTENU DANS LE REFERENCEMENT : ENTRE MOTEURS ET INTERNAUTES

Sous la direction de :

M. Sébastien Billard (responsable universitaire)

M. Julien Cuingnet (tuteur professionnel)

Soutenu le 16 septembre 2010 à l'UFR IDIST
Université Charles de Gaulle, Lille 3 (Campus Pont de Bois)
BP 60149, 59653 Villeneuve d'Ascq Cedex

Année universitaire 2009/2010

Remerciements

Je tiens à remercier...

M. Sébastien Billard, référenceur professionnel et enseignant de référencement naturel à l'Université Lille 3, et M. Julien Cuingnet, référenceur professionnel et co-gérant de l'agence Aeronet, pour leur disponibilité et leurs conseils avisés à mon égard.

Un grand merci à Claire, Greg, Guillaume, Miguel, Alex, Alexandre et Romain, ainsi qu'à l'ensemble des stagiaires d'Aeronet pour leur gentillesse, leur disponibilité et leur soutien.

Enfin, une pensée toute particulière à mes proches pour leur soutien moral et matériel avant, pendant et bien après cette période de stage.

Table des matières

Remerciements 3

Table des matières 5

Introduction 7
A- Présentation de la structure d'accueil 7
B- Les sites d'Aeronet 7
C- Présentation de la mission de stage 9
D- Présentation du projet d'étude 10

Partie 1 - Le référencement en tant que solution à la visibilité des sites de e-commerce 11
1. Pourquoi être visible sur le net ? 11
1.1 - L'obsession de la première page 11
1.2 - Une question de survie pour les sites de e-commerce 12
2. La solution du référencement naturel 14
2.1 - Le référencement naturel ou SEO 15
2.2 - Le référencement naturel ou SEM 16
2.3 - Une nouvelle dimension : le SMO 17
2.4 - Le référencement naturel, une science inexacte 19
3. Les fondamentaux du référencement naturel 21
3.1 - La structure 22
3.1.1 - Relation de la structure avec les moteurs de recherche 22
3.1.2 - Relation de la structure avec le contenu du site 23
3.1.3 - Biais technologiques 25
3.2 - Popularité ou Netlinking 26
3.2.1 - Relation de la structure avec la popularité du site 27
3.2.2 - Sites d'autorité et liens d'autorité 28

Partie 2 - Le contenu, pierre angulaire du référencement naturel 31
1. Le contenu d'un site internet sous toutes ses formes 31
1.1 - Le contenu textuel 33
1.2 - Les images 33
1.3 - Le contenu sonore et la vidéo 35
2. Écriture journalistique et écriture web 36
2.1 - L'écriture journalistique 37
2.2 - Les spécificités de l'écriture web 39
2.3 - Les journalistes du web : du journal en ligne au blog 41
3. L'obsession du bon mot clé dans le référencement 42
3.1 - Problème de définition au sujet du terme « mot clé » 43
3.2 - Méthodes d'élaboration d'une liste de mots clés 44
3.3 - Les particularités du français qui mettent à mal le référencement 46
3.3.1 - L'accentuation des mots clés en cause 47
3.3.2 - Fautes d'orthographe 49
3.4 - Le phénomène de longue traîne 50

Partie 3 - Plaire moteurs, plaire aux internautes **53**
1. Plaire aux moteurs 53
1.1 - Pourquoi plaire aux moteurs de recherche ? 54
1.2 - Comment plaire aux moteurs de recherche ? 55
1.3 - Comment déplaire aux moteurs de recherche ? 57
2. Plaire aux internautes 59
2.1 - Dans la tête de l'internaute 60
2.1.1 - Le langage des internautes *60*
2.1.2 - A chaque site internet son public *61*
2.2 - La plus-value d'un contenu orienté utilisateur 62
2.2.1 - Susciter l'intérêt de l'utilisateur *63*
2.2.2 - Le contenu comme image de marque *63*
2.2.3 - Fidéliser sa clientèle grâce au contenu *65*
2.3 - Comment créer du contenu ? 66
2.3.1 - Densité de mots clés dans une page *66*
2.3.2 - Création de pages de contenu *68*
2.3.3 - Les autres éléments de contenu *69*
2.4 - Ergonomie du contenu 71

Conclusion **73**

Glossaire **75**

Bibliographie/Webographie **77**

Annexes **81**
Liste des sites sur lesquels j'ai travaillé durant le stage 83
Liste des articles rédigés et parus sur le blog d'Aeronet 87

Résumé/Abstract **88**

Introduction

A - Présentation de la structure d'accueil

Aeronet Référencement est une SARL fondée en 2002 dont la spécialité est le référencement naturel. Co-gérée par Messieurs Romain Orjekh et Julien Cuingnet, Aeronet proposait à sa genèse des prestations touchant aussi bien à la création de sites internet, que de l'infographie et, bien évidemment, des prestations de référencement. L'activité de la société s'est finalement recentrée quelques années plus tard sur le référencement naturel, avant que les agences proposant ce genre de prestations ne se multiplient.

Malgré sa taille relativement modeste (moins de 10 salariés), Aeronet est parvenue à se construire une solide réputation parmi les PME, son cœur de cible. De plus, pour soutenir sa croissance et son développement en interne, elle accueille régulièrement des stagiaires qui viennent soutenir le travail des référenceurs ou sont orientés vers des postes de développeurs internet. Enfin, afin de faire face à la recrudescence de concurrents dans son secteur, Aeronet a depuis quelques années développé sa présence sur la toile en mettant en ligne plusieurs sites connexes qui sont autant de vitrines pour son activité.

B - Les sites d'Aeronet

> http://www.aeronet-referencement.fr/

Il s'agit du site principal de l'agence, son site vitrine en quelque sorte. A travers ce site, Aeronet fait la promotion de ses prestations grâce à la présentation de ses meilleurs résultats et le témoignage de ses clients. Il constitue également un relais pour les autres sites, le blog, le compte *Facebook* et le compte *Twitter* de l'agence.

> http://www.ils-ont-reussi-en-ligne.com/

Le site a pour objectif de mettre en avant les clients d'Aeronet en leur donnant la parole à travers des témoignages vidéos. Les clients y font part de leur expérience, de ce que leur a apporté le référencement payant ainsi que de leur relation avec Aeronet.

http://blog.aeronet-referencement.fr/

Le blog d'Aeronet a un rôle prépondérant dans la communication de l'agence et confirme le statut de spécialiste du référencement de cette dernière. Des articles d'actualité liés au monde du référencement, des réseaux sociaux ou des moteurs de recherche viennent régulièrement agrémenter le blog. Plus ponctuellement, ce dernier sert également de porte-voix aux informations liées à l'actualité de l'agence, par exemple lors de la mise en ligne du nouveau site dédié aux outils de référencement gratuits[1]. Chaque nouvel article est par ailleurs relayé sur les comptes *Facebook* et *Twitter* de l'agence afin de lui offrir une visibilité accrue.

http://www.mon-referencement-gratuit.fr/

Mon référencement gratuit offre la possibilité à n'importe quel internaute d'utiliser gratuitement des outils de référencement développés par Aeronet. Il est par exemple possible de comparer les positionnement de plusieurs sites sur un ou plusieurs mots clés, générer des balises méta, analyser les pages d'un site afin d'y déceler de potentielles erreurs de code, etc. Si les outils sont forcément limités, ils permettent tout du moins une entrée en douceur dans le monde du référencement.

Les médias sociaux

Enfin, Aeronet est également présent sur les réseaux sociaux tels que *Facebook*[2] et *Twitter*[3] par le

1 *L'article est disponible sur le blog d'Aeronet:* http://blog.aeronet-referencement.fr/index.php/actu/outils-de-referencement-gratuits-1665/
2 http://www.facebook.com/pages/Lille-France/Aeronet-Referencement/161359686786
3 https://twitter.com/AgenceAeronet

biais de compte dédiés. Cette présence lui permet d'étendre sa visibilité et de toucher un public plus large. Les réseaux sociaux servent également à diffuser les nouveaux articles du blog, dès leur parution.

C - Présentation de la mission de stage

Au sein d'Aeronet, ma mission de stage a consisté à assister les référenceurs dans leur travail quotidien ainsi qu'à gérer des campagnes de référencement pour le compte des clients d'Aeronet. J'ai ainsi pu réaliser des pistes de travail, de l'optimisation de code, de la recherche de mots clés, de l'analyse de positionnement et de trafic ainsi que de la rédaction de contenu pour le compte d'Aeronet et de ses clients.

Bien que cela ne constitua pas l'ensemble des mes activités quotidiennes au sein de l'agence, mon travail s'est fortement orienté vers la rédaction de contenu. Au sein d'Aeronet, rédaction et référencement ne sont pas formellement séparés. Si certains salariés, qui ont de fortes compétences en développement pur, ont des profils typiques de référenceurs au sens technique du terme, d'autres officient davantage dans la rédaction. Toutefois, d'une manière général, chacun contribue aux sites dont il a la charge tant sur le plan technique que sur le plan rédactionnel. Au sein d'autres sociétés, qui regroupent davantage de salariés, rédaction et référencement sont généralement séparés en deux pôles bien distincts.

A titre tout à fait personnel, il m'a semblé que mêler ces deux aspects ne pouvait être que bénéfique pour la qualité de la prestation dans la mesure où chaque référenceur, responsable du suivi de plusieurs clients, a une vision très précise du travail à apporter aux sites dont il a la charge, que cela concerne la *netlinking*, la structure ou encore le contenu.

D - Présentation du projet d'étude

Ma mission de stage chez Aeronet m'aura permis de découvrir le métier de référenceur et, d'une certaine manière, celui de rédacteur web. Cette plongée au sein des multiples activités liées au référencement de sites internet divers et variés, m'a rapidement confronté à une situation qui me paru sur le moment des plus problématiques, sans doute en grande partie en raison de mon manque d'expérience dans le domaine. En effet, le contenu des sites, qui constitue, comme on le verra par la suite, un élément parmi les plus importants pour le bon référencement d'un site, doit répondre aux attentes de deux cibles qui me semblèrent en apparence antinomiques et sous certains aspects inconciliables : Les moteurs de recherche d'un côté, et les internautes de l'autre. Pourtant, à cette même occasion, je fus surpris de constater que ce qui me paraissait un problème insoluble, ne l'était pas nécessairement pour les référenceurs professionnels qui m'entouraient.

C'est à partir de ce surprenant constat que s'est tissée ma réflexion sur le traitement du contenu dans le référencement : Quels sont les outils et les moyens en la possession du référenceur pour optimiser le contenu d'un site internet ? D'autre part, et à partir de ces éléments en sa possession, de quelle manière peut-il marier idéalement les préconisations des moteurs de recherche aux attentes des internautes ?

La première partie posera un cadre théorique à nos investigations en pointant les différents aspects invoqués dans la pratique du référencement. La seconde partie sera l'occasion pour nous d'entrer un peu plus en profondeur dans la notion de contenu et de mot clé. Nous aborderons également les spécificités de l'écriture de contenu web en prenant l'écriture journalistique comme point de départ et point de comparaison. Enfin, la troisième partie tentera d'apporter des solutions à notre problématique initiale en déterminant les meilleures façons de plaire aux moteurs ainsi qu'aux internautes.

Partie 1 - Le référencement en tant que solution à la visibilité des sites de e-commerce

1. Pourquoi être visible sur le net ?

207 millions, c'est approximativement le nombre de sites internet dénombrés en Mai 2010 par la société Netcraft[4]. Un chiffre considérable pour un média âgé d'un peu moins de deux décennies. Le paradoxe est que si le public ayant accès à internet a vu son nombre naturellement s'accroitre depuis le milieu des années 1990 (58% de la population française aurait accès à internet en 2008 contre 14,4% en 2000[5]), pour un site perdu parmi une myriade d'autres sites, il est devenu bien plus difficile aujourd'hui de capter son attention.

Pour espérer être consulté par les internautes, il faut bien évidemment se faire le plus visible possible. Si cela passait essentiellement par les annuaires aux prémisses du web, tous les regards se sont rapidement tournés vers les moteurs de recherche qui offraient l'avantage d'une indexation automatique donc une chance égale pour chaque site d'atteindre un bon positionnement.

1.1- L'obsession de la première page

Si la première page du moteur de recherche est si convoitée, ce n'est évidemment pas par le plus grand des hasards. Quand on sait que « *moins d'un tiers des internautes ne consultent que la première page de résultats* [6]» et que seulement moins d'un tiers dépasse la quatrième page,

4 **Netcarft**, *May 2010 Web Server Survey*, 14 mai 2010 http://news.netcraft.com/archives/2010/05/14/may_2010_web_server_survey.html

5 **Internet World Stats**, *France Internet Usage Stats and Telecom Reports*. http://www.internetworldstats.com/eu/fr.htm

6 **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010, p.9

l'obsession de se trouver, ne serait-ce qu'avant la quatrième page semble complètement justifiée. La raison principale en est qu'un internaute n'a généralement aucun intérêt à parcourir l'ensemble des pages de résultats. Aussi sait-il sans doute par expérience ou par pure déduction que les premières pages de résultats qui lui sont offertes sont les plus pertinentes par rapport à sa requête. Et quand bien même, les résultats ne le satisfont pas, il a tout le loisir de reformuler sa pensée, à tout moment, en entrant une nouvelle requête. Agir autrement, en voulant à tout prix parcourir l'ensemble des pages de résultats d'une requête ne constituerait tout au plus qu'une pure perte de temps et d'énergie. Si certains internautes s'échinent à consulter les pages de résultats au-delà de la quatrième page, ils sont toutefois très largement minoritaires.

Dans ces conditions, on comprend sans peine qu'atteindre les premières pages des moteurs de recherche est devenu l'objectif premier de bon nombre de responsables de sites internet. Si être présent en première page ne garantit en rien une hausse du taux de transformation d'un site, se trouver devant ses concurrents constitue tout de même un avantage loin d'être négligeable. Les clients d'Aeronet souhaitent, comme tout candidat au référencement naturel, se retrouver dans cette situation avantageuse. Qu'ils fassent le commerce de chocolats, de déguisements, de produits bio, de méthodes d'apprentissage de la musique ou qu'ils proposent du matériel de protection pour les professionnels, chacun d'entre-eux tapent à la porte d'une agence de référencement avec le même désir : Améliorer leur positionnement sur les moteurs de recherche avec pour finalité l'augmentation de leur chiffre d'affaires.

1.2 - Une question de survie pour les sites de e-commerce

Si ne pas bénéficier d'une bonne visibilité pour un site vitrine peut être dommageable pour l'image de marque qu'il est censé véhiculer en direction des internautes, il en est tout autrement d'un site de commerce électronique (ou e-commerce).

En pareil cas, quand une entreprise ne peut compter que sur son site internet pour réaliser son chiffre d'affaires, il n'est pas feint d'employer le terme de « survie » pour signifier tout l'enjeu d'une bonne visibilité. C'est d'autant plus vrai lorsque le site de e-commerce en question ne peut pas compter sur son nom, ni sur sa notoriété, mais sur ses seuls produits et prestations. A titre de comparaison, les grands noms du e-commerce comme *Amazon*, *Cdiscount* ou *Fnac* sont beaucoup moins sujets à ce problème de référencement étant donné le nombre considérable d'internautes qui se rendent sur ses sites en tapant directement l'URL du site ou le nom du site dans la barre de recherches. Ces sites, déjà très présents sur les moteurs, concentrent en règle générale leurs efforts sur leur image de marque et leur notoriété. La question n'est plus tant « *L'internaute va-t-il nous trouver ?* » mais plutôt des questions du type : « *L'internaute va-t-il être intéressé par nos offres ?* », « *Qu'est-ce qui va le décider de commander chez nous plutôt que chez notre concurrent ?* », etc.

Par ailleurs, boudé il y a encore peu de temps pour son manque de sécurité (avéré comme fantasmé), le e-commerce ne fait aujourd'hui plus peur. En 2005, si les ventes en ligne ne représentaient que 8,4% des ventes en général, en 2009, elle avaient atteint les 25%[7] en France. La progression de l'accès au haut débit dans les foyers français (44% des ménages[8]), le temps passé sur internet, propulsés par la multiplication des offres en ligne, expliquent clairement cette tendance.

Internet est donc tout naturellement devenu un marché juteux qui pousse un nombre croissant d'entreprises à ouvrir leur propre boutique en ligne. Cependant, il ne faut pas s'y tromper, comme tout secteur porteur, le commerce en ligne est monopolisé par un nombre restreint de sites. En 2009, 500 sites enregistraient chacun plus de 10000 transactions par mois pendant que 19100 sites

7 **La Fevad**, *Baromètre e-commerce ICE - 1er trimestre 2009*, 12 mai 2009
http://www.fevad.com/index.php?option=com_content&task=view&id=497&Itemid=861

8 **Le Journal du Net**, *Europe : Population d'internautes*, 14 juin 2010
http://www.journaldunet.com/cc/01_internautes/inter_nbr_eu.shtml

peinaient à en réaliser 10 sur la même échelle de durée[9]. Parmi les sites les plus visités, citons les incontournables *Ebay*, *Priceminister*, *La Redoute*, *Amazon.fr* ou encore *Cdiscount*.

Les sites de e-commerce sur lesquels nous porterons notre attention lors de notre analyse ultérieure ne font pas partie de cette sphère de géants du commerce électronique. S'apparentant davantage à des PME, les sites suivis par l'Agence Aeronet ont pour la plupart une activité bien plus modeste et surtout bien plus spécialisée (Vente de gazon synthétique, pompes à bière, baptême de l'air, etc).

2. La solution du référencement naturel

Commençons par un problème de définition. Le terme de « référencement » est en effet ambigu. Ambigu car, si l'on s'en tient à sa définition première, il ne recouvre finalement qu'une infime partie de ce qu'un référenceur peut décrire de son activité. Comme nous le rappelle Olivier Andrieu[10], référencer un site ne revient finalement qu'à le placer dans la base de données d'un moteur de recherche afin qu'il soit « trouvable » par l'internaute. Bien qu'il s'agisse d'une étape forcément obligatoire dans l'existence d'un site internet (car, autrement, comment pourrait-on trouver ce qui n'est pas trouvable, quand bien même il existe bel et bien en un point inconnu du web ?), elle ne peut en rien résumer l'activité d'un référenceur. D'autant plus que la plupart des sites confiés aux référenceurs ont par ailleurs déjà été « référencés » par la majorité des moteurs.

Dans la bouche d'un référenceur, et sans aucun doute par un abus de langage plus que toléré, le référencement désigne une somme d'activités bien plus étendue que la seule prise en compte d'un site par les moteurs. En effet, l'objectif essentiel du référenceur va être de positionner du mieux qu'il le peut un site internet sur des mots clés, performants si possible, en rapport étroit avec la

9 **La Fevad**, *Baromètre e-commerce ICE - 1er trimestre 2009*, 12 mai 2009 http://www.fevad.com/index.php?option=com_content&task=view&id=497&Itemid=861

10 **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010, p.2

thématique du site en question. Il possède différents moyens pour parvenir à cet objectif : optimisation du code, rédaction et optimisation des contenus, développement d'un réseau de liens, etc. Des pratiques qui seront plus amplement décrites lors de la partie de ce mémoire consacrée aux fondamentaux du référencement.

2.1 - Le référencement naturel ou SEO

Le référencement naturel, également appelé référencement organique, ou encore *Search Engine Optimization* par les anglophones, représente l'essentiel de l'activité de l'agence Aeronet. Ce type de référencement repose sur le principe que tous les sites internet, quels qu'ils soient, partent sur un même pied d'égalité lors de leur mise en ligne. Ce qui va les départager, c'est finalement tout le travail qui aura été mis en œuvre pour que le code, le contenu et les liens soient pris en compte de la meilleure des manières possibles par les moteurs.

Ainsi, le référencement naturel est par principe « gratuit » dans le sens où un *webmaster* n'a pas à payer Google, par exemple, pour que son site soit bien indexé et bien positionné sur le moteur. Tout à chacun peut donc travailler son référencement naturel, à condition évidemment d'en posséder les connaissances et les compétences techniques. Mais la dimension « naturelle » de ce référencement sous-entend également que tout site est « naturellement » référencé par les moteurs dès lors qu'il est « remarqué » par ces derniers, à condition que rien n'entrave ce référencement (biais techniques ou démarche volontaire de la part de la personne en charge du site).

C'est d'ailleurs la principale différence qui existe entre le fonctionnement des moteurs de recherche et le fonctionnement des annuaires généralistes ou thématiques. Dans ces derniers, il est nécessaire de soumettre (à un être humain, responsable de l'annuaire) alors que le moteur de recherche indexe automatiquement par le biais de robots, les *spiders**. La soumission d'un site dans un moteur de

recherche reste toutefois possible dans la plupart des cas afin d'inviter les robots à venir indexer le contenu du site, si cela n'a pas encore été fait.

Bien que le sujet de ce mémoire porte pour l'essentiel sur le traitement du contenu par les référenceurs et les moteurs de recherche, les différents aspects du référencement naturel y seront néanmoins abordés, ne serait-ce que pour replacer la thématique du contenu dans son contexte initial mais aussi parce qu'elle est indubitablement inséparable des autres aspects qui touchent au référencement.

2.2 - Le référencement sponsorisé ou SEM

Deuxième dimension que recouvre le terme de référencement, le référencement payant, également nommé référencement sponsorisé ou *Search Engine Marketing* peut se révéler à la fois une alternative et un complément au référencement naturel. Les liens naturels (et gratuits) à la base du principe du référencement naturel cèdent ici leur place aux liens sponsorisés ou commerciaux. Un lien sponsorisé est un encart en marge des résultats issus de la recherche naturelle qui est acheté à une régie publicitaire (Par exemple *Google Adwords* chez *Google* ou *Yahoo Search Marketing* chez *Yahoo*).

Le fonctionnement est très simple et il s'agit sans aucun doute là d'une des raisons pour lesquelles ce service rencontre tant de succès auprès des acteurs du e-commerce. Prenons l'exemple de *Google Adwords,* soit la plus utilisée des régies publicitaires sur internet : Un annonceur achète un emplacement sur une expression qui lui permet d'apparaître dans les résultats sponsorisés de Google. Il ne paye que quand un internaute clique sur son lien, le montant du *coût par clic* étant déterminé par un système d'enchères. Par exemple, une expression comme « *jouer poker* » est

estimée à 0,81€ du coût par clic moyen, alors que « *cartable scolaire* » est estimé à 0,25€[11].

Aeronet ne propose que des prestations de référencement naturel. Néanmoins, les clients qui effectuent également des campagnes de référencent sponsorisé en parallèle du travail effectué sur leur référencement naturel bénéficient d'un suivi sur l'ensemble de leur trafic (naturel comme sponsorisé) ainsi que des conseils des référenceurs sur les mots clés à privilégier pour chacun de ces deux types de référencement.

2.3 - Une nouvelle dimension : le SMO

La SMO (pour *Social Media Optimization)* est une dimension nouvellement prise en compte par les agences de référencement et d'une manière plus générale par les agences en charge de la communication des marques et des produits. On pourrait résumer cette technique comme du référencement naturel centré sur les réseaux sociaux, principalement constitués par les incontournables *Facebook*, *Twitter*, *Linkedin* ou encore *Viadeo*. Mais à plus grande échelle, on peut également y inclure les blogs, les forums de discussion, les portails d'information tels que *Wikio*, les jeux vidéo à dimension sociale, et bien d'autres plateformes encore (Voir ci-dessous l'illustration de la typologie des différents services apparentés aux réseaux sociaux).

11 *Chiffres fournis pas le service* Google Keywords *en Août 2010*

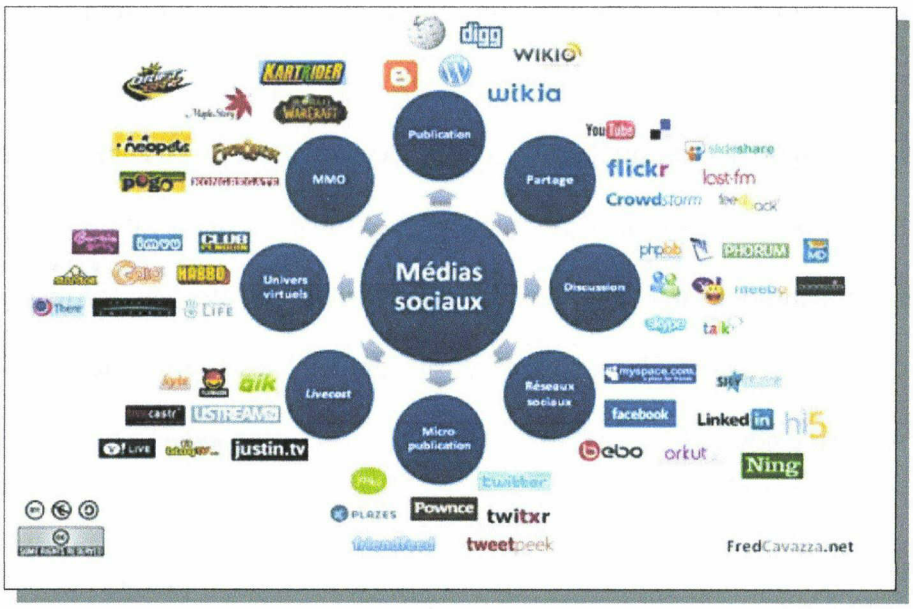

Illustration 1 - Typologie des médias sociaux

Si l'impact des réseaux sociaux par rapport aux moteurs de recherche reste encore loin de convaincre tous les référenceurs[12], le SMO est une tendance aujourd'hui difficilement contournable, en particulier pour la communication des grandes marques. Pour résonner en terme de référencement (et pas seulement en terme de présence), les règles d'un contenu déposé dans un réseau social sont les mêmes que pour le SEO : Il faut faire preuve de pertinence et d'accessibilité afin de plaire aux internautes autant qu'aux moteurs de recherche, même si les internautes sont la cible principale visée par le SMO.

L'agence Aeronet fait elle-même appel aux réseaux sociaux pour sa propre promotion et la promotion de certains de ses clients. Ainsi, une page fan sur *Facebook* et un compte *Twitter* sont régulièrement alimentés, en particulier pour mettre en avant les nouveaux articles mis en ligne sur le blog Aeronet. De cette manière, les réseaux sociaux permettent à l'agence, par ailleurs déjà positionnée sur de nombreux mots clés dans les moteurs de recherche, « *d'assoir sa popularité et de*

12 **Scanblog**, *Baromètre ECHO : les français face aux médias sociaux et la e-réputation*, 14 avril 2010 http://www.scanblog.com/blog/etude-echo-ereputation-scanblog-172/

mettre en avant ses contenus »[13]. Un contenu qui devient visible pour tous les internautes inscrits sur les plateformes sociales (quand une inscription est exigée) et qui peut être réutilisé très simplement, et commenté par les internautes. Le trafic généré peut ainsi affluer de n'importe qu'elle source : article ou commentaire dans un blog, lien dans un réseau social (généralement en *nofollow* donc non-suivi par les moteurs de recherche). La sphère des réseaux sociaux constitue un marché en pleine ébullition qui n'a sans aucun doute pas encore révéler tous ses enjeux. Il est fort à parier qu'il s'agit-là d'une dimension qui deviendra véritablement incontournable pour tout à chacun qui souhaitera étendre sa promotion sur l'ensemble de la toile.

2.4 - Le référencement naturel, une science inexacte

Enfin, avant de débuter notre partie consacrée aux fondamentaux du référencement naturel, il convient au préalable de se pencher sur une spécificité liée au référencement naturel. Si nombre d'agences de référencement naturel promettent des positionnements très précis à leurs clients, il faut toutefois rester lucide : le référencement naturel n'est en rien une science exacte et aucun référenceur sérieux ne saurait assurer son client d'un positionnement sur un mot clé ou sur une expression, quelle qu'elle soit.

A cela plusieurs raisons : Le fonctionnement des moteurs populaires (en particulier *Google*) est généralement connu des référenceurs, d'autant plus que les concepteurs des moteurs[14] communiquent sur les bonnes pratiques à adopter pour être le mieux référencé. Il demeure toutefois certaines parts d'ombres qui, à elles-seules, ne peuvent assurer le résultat d'une prestation de référencement. Ainsi, l'algorithme de ces moteurs est assez souvent mis à jour, ne permettant pas d'être tout à fait certain qu'une décision prise à un instant T sera toujours aussi pertinente quelques

13 **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010. p.233

14 *Chez Google, c'est en particulier le célèbre Matt Cutts qui prodigue ses conseils avisés aux référenceurs à l'occasion de conférences à travers le monde et d'article sur son blog* http://www.mattcutts.com/blog/

mois après. En résumé, « *il est impossible pour un prestataire de fournir plus de garanties sur le résultat de son travail, puisque celui-ci est basé sur des technologies qui appartiennent aux moteurs de recherche et, donc, qu'il ne contrôle pas.* [15]»

De plus, un référenceur, quand bien même puisse-t-il parfaitement connaître son métier et être en mesure de conseiller au mieux ses clients, doit souvent travailler au quotidien en faisant de nombreux compromis. L'objectif du référenceur ne s'accorde pas toujours au mieux avec les souhaits parfois tout à fait légitimes des clients et les décisions des acteurs extérieurs, quand ces derniers existent (agence en charge de la communication, agence en charge de la création du site, etc.). Au final, bien rares sont les prestations de référencement qui ne se retrouvent pas à un moment ou un autre face à ce genre d'obstacle extérieur.

Il ne faut pas non plus négliger le fait que la plupart des concurrents des sites qui réclament une prestation de référencement naturel sont eux-mêmes suivis par une agence. Dans ces conditions, quand de nombreux acteurs se battent sur un terrain commun, d'autant plus quand ce dernier est fortement concurrentiel, il est bien difficile d'assurer un positionnement relativement stable. Le positionnement peut varier, parfois très rapidement selon les moteurs, mais si aucun problème technique ne survient, les variations restent généralement assez minimes et il est rare que le positionnement d'un site s'écroule d'un seul coup sans qu'on en connaisse les raisons.

Enfin, le référencement naturel, de même que sponsorisé, ne peut garantir à lui tout seul une hausse du chiffre d'affaires d'un site de e-commerce. Le référenceur ouvre des portes, facilite l'accessibilité à un site, certes, mais une fois arrivé sur le site internet, la décision finale de l'internaute n'est plus du ressort du référenceur. Seuls les prix pratiqués, la crédibilité de l'offre (qui passe toutefois en grande partie par le contenu), et bien d'autres éléments sont à même de peser dans la balance.

15 **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010. p. 429

3. Les fondamentaux du référencement naturel

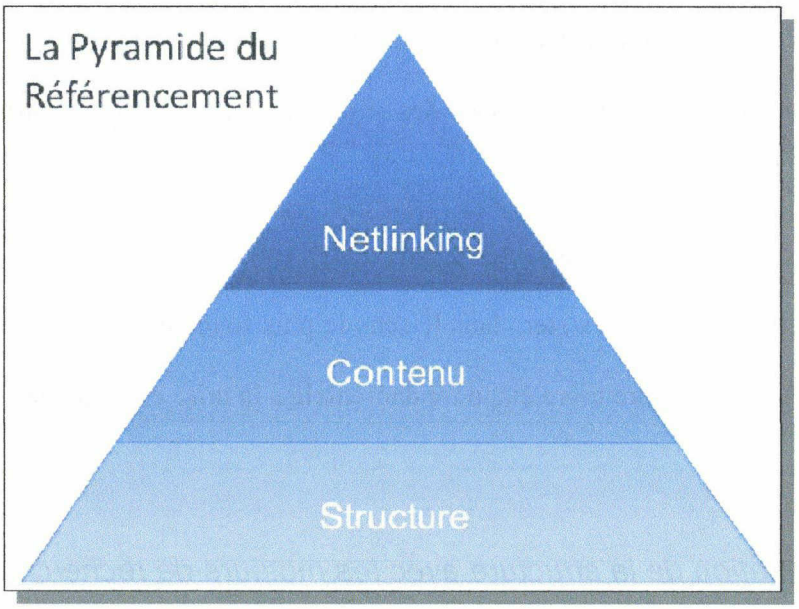

Illustration 2 - Pyramide du référencement simplifiée

La plupart des référenceurs ont tendance à présenter le référencement comme une pyramide constituée de trois piliers interdépendants : A la base, le pilier correspondant à la structure du site; au sommet, le pilier représentant la popularité du site; et enfin, entre ces deux piliers, celui consacré au contenu, étage central qui retiendra toute notre attention dans les deux prochaines parties de ce mémoire. Pour prolonger l'analogie, il suffirait d'enlever un étage à cette pyramide pour qu'elle devienne tout à coup complètement bancale, voire qu'elle s'écroule sur elle-même. Nous allons à présent entrer plus en détail dans chacun de ces piliers afin d'en mettre en avant leurs singularités et leur caractère indispensable pour la mise en œuvre d'un travail de référencement naturel.

3.1 - La structure

Premier pilier du référencement mais non des moindres, la structure représente littéralement les fondations du site. Plus prosaïquement, il s'agit du code. Qu'il s'agisse de code *html* ou *php*, la structure n'est finalement qu'une succession de caractères et de balises. De simples caractères qui peuvent néanmoins être à l'origine de bien des problèmes et peuvent aboutir en définitive au mauvais positionnement d'un site.

La structure va donc avoir pour objectif principal de rendre accessible le site, à la fois aux internautes (elle permet au site d'exister, dans le sens le plus fondamental du terme) mais aussi aux moteurs, ce qui constitue une première étape fondamentale à la mise en route du référencement.

3.1.1 - Relation de la structure avec les moteurs de recherche

Sans prétendre entrer dans le détail technique du fonctionnement des moteurs de recherche, il convient toutefois d'en décrire les grandes lignes, ne serait-ce que pour définir clairement la relation qu'ils entretiennent avec la structure des sites internet.

Cela semble une évidence, mais pour qu'un site soit indexé, faut-il encore qu'il puisse l'être. Un site est une composante de technologies dont la plus courante et la mieux perçue par les moteurs est le code *html.* Ainsi, un site dont les fondations (structure et code) sont solides n'a, a priori, pas particulièrement de raison de ne pas se voir indexer par les moteurs de recherche.

Pour comprendre la manière dont les moteurs de recherche perçoivent une page *html*, le mieux est encore de se mettre à leur place. A ce niveau, tout le travail du référenceur consiste à « lire » les sites internet comme le ferait un robot, ou tout du moins, à toujours avoir en tête la manière de

percevoir des *spiders*. C'est à dire, dans la pratique, en omettant les images, les feuilles de styles (*CSS*), le *Flash*, le *Javascript*, les *cookies*, les formulaires et en lisant le code de manière linéaire. *Lynx*[16], un navigateur issu des prémisses d'internet, au sein duquel seul le texte est affiché, sans aucunes fioritures, donne tout le loisir de naviguer sur un site de cette manière.

Plus pratiques néanmoins, un certain nombre d'extensions permettent aux navigateurs actuels (tels que *Mozilla Firefox* ou *Google Chrome)* d'obtenir un résultat relativement similaire en passant outre la rebutante navigation de *Lynx*. On peut ainsi citer *Firebug*[17] ou encore *Web Developper*[18], parmi les plus connus et les plus efficaces. C'est par ailleurs par ce biais que les référenceurs, tout du moins les référenceurs d'Aeronet, parviennent la plupart du temps à mettre en exergue des problèmes de codes difficilement identifiables autrement. Pour ma part, j'ai grandement utilisé ces extensions durant l'ensemble de mon stage. En particulier, lors de recherches de *bugs* et la rédaction de pistes de travail visant à l'optimisation de certains sites de clients de l'agence.

3.1.2 - Relation de la structure avec le contenu du site

A l'image d'un contenant ébréché qui laisserait échapper et gâcherait l'ensemble de son contenu, un site dont la structure est instable ne peut prétendre mettre en valeur son contenu, qu'il s'agisse de texte ou de tout autre type de contenu. Bien que rares soient les structures qui peuvent s'enorgueillir d'être parfaites, dénuées de toute failles, il convient au *webmaster* d'apporter tout de même le plus grand soin lors de l'élaboration du site dont il a la charge.

Un certain nombre des sites des clients qui se présentent à Aeronet ont une structure défaillante avant même d'avoir un problème de référencement. C'est-à-dire que nombre de ces clients réclament l'amélioration de leur positionnement dans les moteurs de recherche sans se douter qu'une

16 *Le navigateur Lynx est disponible gratuitement à l'adresse suivante : http://csant.info/lynx*
17 *Le plugin Firebug pour Firefox et Chrome est disponible à l'adresse suivante : www.getfirebug.com*
18 *Le plugin Web Developper est disponible à l'adresse suivante : https://addons.mozilla.org/fr/firefox/addon/60/*

part non négligeable de leur manque de performance réside en particulier dans les négligences dont ont fait preuve les *webmasters* lors de l'élaboration de leur site internet.

Les causes peuvent être multiples mais certaines reviennent régulièrement. Par exemple, l'architecture de certains sites est basée sur des *solutions clé en main,* qui ne nécessitent aucune connaissance particulière en *html* et permettent la mise en ligne rapide de sites de e-commerce ou de sites vitrines. On peut citer, à titre d'exemple, parmi une myriade de prestataires de ce type, la société *Oxatis*[19] qui s'adresse spécifiquement aux e-commerçants. Si la formule du *clé en main* a de quoi séduire, elle s'accompagne malheureusement d'un certain nombre de désagréments lorsqu'il s'agit de configurer en profondeur le site internet. Ainsi, il arrive qu'un commerçant ne puisse avoir qu'une marge de manœuvre résolument limitée dans l'édition de son contenu : des *URLs* qui ne peuvent être modifiées, des *balises titles** et *balises meta** générées automatiquement, etc. Dans ce genre de situation, la tâche du référenceur apparaît d'autant plus complexe qu'il doit parfois « jongler » entre le client, le prestataire ou le *webmaster* pour réussir à faire entendre sa voix et enfin optimiser le site.

Enfin, la structure est également synonyme de mise en page du contenu. Paragraphes, hauteur des lignes, cadres, tout ces éléments structurels sont régis par du code. Un contenu brut, dénué de toute mise en page, n'aurait pas grand sens dans un support qui vise justement à l'accessibilité et à l'attractivité. Et si la mise en page est nécessaire pour une lecture optimale du contenu par les internautes, elle l'est tout autant pour les moteurs. Comme nous le verrons plus en détail par la suite, les balises qui structurent le contenu sont aussi des indicateurs pour la compréhension du texte par les *spiders.* Un texte brut, sans l'ombre d'une structure, aurait autant de sens pour un moteur qu'un texte sans signes de ponctuations pour un être humain : les mots sont bien présents mais le sens de l'ensemble apparaît avec difficulté.

19 *Les prestations de la société Oxatis sont disponibles sur http://www.oxatis.fr/*

3.1.3 - Biais technologiques

Pour illustrer l'importance du code dans le référencement, prenons l'exemple certes classique mais néanmoins révélateur de la technologie *Flash*. Les sites totalement réalisés en *Flash*, plébiscités par les graphistes pour leur apport esthétique à l'identité d'une marque ou d'un produit, ont longtemps été et sont toujours quelque peu considérés comme contreproductifs en matière de référencement. A raison, puisque si les *spiders* sont capables d'interpréter le langage *html*, il en est tout autrement du *Flash* dans lequel les robots ne sont pas disposés à entrer. Ainsi, quand un client demande l'amélioration d'un site majoritairement constitué de *Flash*, peu de solutions s'offrent alors au référenceur : Il lui faut pour cela insérer du texte, assorti de balises *html*, afin que le moteur de recherche qui visite le site puisse se « raccrocher » à un contenu, aussi limité soit-il. Une autre solution consiste à proposer une version alternative du site, cette fois réalisée en *html* afin de contenter le moteur (et l'internaute qui n'aurait pas accès au *Flash*, au passage).

D'autres éléments peuvent également bloquer le travail des *spiders*, parmi lesquels : les langages *Javascript* et *Ajax*, les erreurs de code, les mauvaises redirections, certains formats tiers, les formulaires ou encore les indications inscrites dans le fichier *robot.txt**. Toutefois, dans ce dernier cas de figure, il peut tout simplement s'agir d'un blocage global ou ciblé (sur une ou plusieurs pages du site) né de la volonté du *webmaster* (souvent dans le but d'empêcher l'indexation d'un contenu en particulier).

A la lumière des ces exemples, il n'est donc pas difficile de comprendre qu'une structure défaillante peut se révéler catastrophique pour le référencement d'un site internet. Le travail d'un référenceur nécessite donc, dans un premier temps, de localiser les potentiels problèmes de code et de structure à résoudre. C'est le pré-requis nécessaire pour pouvoir déployer son travail au niveau du contenu et de la popularité du site.

3.2 – Popularité ou Netlinking

Dans le milieu du référencement, la notion de popularité est intiment liée à la notion de « lien ». Qu'ils soient internes au site (liens menant de l'accueil à une page catégorie, par exemple) ou qu'ils soient externes (d'un site A à un site B), les liens ont une importance capitale dans l'établissement de la popularité d'un site internet. Olivier Andrieu insiste particulièrement sur ce point : « *Les liens sont très importants pour les moteurs de recherche car ils permettent à leurs robots d'explorer votre site pour y cueillir d'autres documents.* »[20] Cela sous-entend également qu'un *spider* suivra aussi bien les liens menant vers une page interne du site qu'il est en train de visiter que les liens menant vers des sites externes.

En ce qui concerne les liens externes, ces derniers se doivent d'être pertinents, c'est-à-dire que pour un site qui propose de la location de gîtes de vacances en montagne, il sera toujours bien plus pertinent d'effectuer un échange de liens avec un site du même domaine ou d'un domaine connexe (par exemple, une agence de tourisme) qu'avec un site dont le domaine est éloigné (site de vente de matériel de Poker, par exemple). Un moteur comme Google a tendance à prendre mieux en compte les liens entre des sites dont les thématiques sont proches ou semblables, c'est-à-dire qui partagent en commun une partie de leurs mots clés.

De la même manière, les liens internes pointant des pages peu intéressantes pour les moteurs, par exemple la page contact ou la page relatant les mentions légales des sites, soit des contenus peu dissociables selon les sites, doivent être affublées d'un attribut *nofollow*, indiquant par là-même aux *spiders* de ne pas suivre les liens en question. A ce propos, un lien en *nofollow* n'aura d'intérêt que pour les internautes, il ne peux rien apporter au référencement du point de vue technique.

20 **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010. p.157

Cependant, tout lien, même en *nofollow,* peut amener du trafic sur la page d'un site, ce qui en soi ne peut être que positif. L'attribut *nofollow* est particulièrement utilisé sur les réseaux sociaux afin que ces derniers ne deviennent pas l'espace idéal pour un « *bourrage de liens* » sans vergogne. Quand bien même une marque ou une société serait tentée de multiplier les liens vers son site sur un réseau, comme *Twitter*, elle ne finirait par ne susciter qu'agacement chez les internautes et risquerait de tacher son image.

3.2.1 - Relation de la structure avec la popularité du site

Pour qu'un lien soit bien suivi par un moteur, faut-il encore qu'il soit rédigé d'une manière optimale. Cette règle élémentaire est parfois bien loin d'être respectée dans la pratique. En cause, des problèmes liés au code qui ne résultent parfois que d'anodines fautes de frappes. Des fautes de frappes cependant lourdes de conséquences puisqu'elles ont tout le potentiel pour annihiler l'intérêt d'un lien. Lors d'un échange de liens, j'ai souvent pu constater que le code que j'avais fourni à mon contact avait été ajouté à la va-vite, sans vérifications, ce qui rendait l'intérêt du lien caduc. Parmi les problèmes de code qu'il m'ait été donné d'observer, on peut citer à titre d'exemple la récurrente mauvaise utilisation des balises *<a>*. Par exemple, *Vaisselle jetable* où les balises ** ne sont pas à leur place. Dans ce cas, le moteur considérera « *Vaisselle jetable* » comme ancre du lien, ce qui peut potentiellement brouiller l'intérêt du lien aux yeux du moteur.

Certains liens peuvent également perdre de leur substance à cause des redirections multiples (appelées *redirections en cascade**) qu'ils subissent car « *il y a de fortes chances pour que les moteurs ne suivent pas la deuxième redirection.* »[21], comme l'indique Olivier Andrieu. L'apprenti référenceur doit donc rapidement acquérir le réflexe de vérifier ses liens, soit en vérifiant la source

21 *Ibid.*, p. 301

du lien, soit en utilisant des applications[22] prévues à cette effet.

Enfin, terminons cette partie consacrée à la relation entre la structure et la popularité des sites en évoquant l'attribut *nofollow*. Placé dans le code d'un lien, ce dernier permet tout simplement d'indiquer aux moteurs de ne pas suivre le lien, donc de ne pas le prendre en considération lors de l'indexation. Le lien perd donc de tout son intérêt pour le calcul du *PageRank**, ce qui ne lui empêche pas de rester vecteur de trafic auprès des internautes. Ce n'est donc pas sans hasard que cet attribut est devenu véritablement indissociable du monde des réseaux sociaux. En effet, que ce soit sur *Facebook* ou sur *Twitter* mais aussi dans les commentaires de la plupart des blogs, tout lien posté est automatiquement affublé de son attribut *nofollow*. Un choix des concepteurs néanmoins totalement justifiable car, sans cela, les abus seraient légion.

3.2.2 - Sites d'autorité et liens d'autorité

A l'heure actuelle, mettre en ligne un site internet en voulant se passer de l'environnement proche incarné par les autres sites est une véritable hérésie et une décision des plus absurdes si l'on souhaite assurer sa viabilité. Véritable écosystème dont le fondement-même repose sur le principe d'hyperliens, internet est en soi un maillage complexe de sites interconnectés.

Comme on peut s'en douter, chaque site ne possède pas le même poids dans ce microcosme. Qu'ils aient acquis leur renommé par leur contenu de qualité, leur ancienneté sur la toile, leur audience particulièrement importante, leur expertise dans un domaine particulier ou leur réseau de partenaires, les sites qui font autorité en leur domaine font partie intégrante de la dynamique des moteurs de recherche. Parmi ces sites, on va trouver :

22 *Voici, parmi d'autres, une application gratuite qui teste les redirections : http://www.annuaire-info.com/outil-referencement/test-redirection*

- Des sites et des annuaires dont l'autorité est dite « *absolue* ». Leur prédominance est incontestable sur internet et on les retrouve en bonne place sur énormément de requêtes et d'autres sites (Exemple : *Facebook*, *Wikipedia*, *lemonde.fr*, etc.).

- Ainsi que des sites et des annuaires dont l'autorité est dite « *relative* », qui cumulent également des excellents positionnements dans des domaines toutefois moins étendus que les sites d'autorité absolue (*jeuxvideo.com* dans le domaine du jeu-vidéo, *webrankinfo.com* dans le domaine du référencement, etc.).

Il existe donc des sites omnipotents qui brassent à eux-seuls quotidiennement un flux considérable d'internautes issus de sources diverses et variées. Tout naturellement, les liens qui émanent de ces sites sont sources de toutes les convoitises. Deux principales raisons à cela :

- Le moteur qui détecte des *liens d'autorité* (issus donc d'un site d'autorité) favorise le site qui en bénéficie en partant du principe qu'un site qui obtient des liens d'autorité doit être un site de confiance, un site utile ou tout simplement un site pertinent pour l'internaute. Le site est « bien vu » du moteur en quelque sorte et est récompensé comme tel.

- Un modeste site va pouvoir tirer les bénéfices d'un trafic bien plus considérable qu'à l'accoutumé étant donné la potentialité du nombre d'internautes offerte par le site d'autorité. Par exemple, un site évoqué dans un article de *lemonde.fr* va tout naturellement gagner de l'audience, sans doute durant une période limitée mais il s'agit-là d'une publicité gratuite rarement boudée par les responsables de sites.

Chez Google, ce principe est représenté et formalisé par le système de *PageRank* qui indique, sur une échelle de 0 à 10, la « valeur » de la page d'un site par rapport aux autres pages indexées par le moteur. Le plus intéressant reste que cette note de Google va directement avoir une influence sur le positionnement des sites.

Ce système de valeur reste toutefois à relativiser car il ne faut pas oublier qu'il est issu de le réflexion des ingénieurs de Google qui, même en l'ayant conçu avec un maximum d'objectivité, restent liés aux intérêts de leur entreprise, ce qui a pour incidence un moteur potentiellement critiquable à certains points de vue. Néanmoins, étant donné le statut de quasi-monopole du moteur dans un grand nombre de pays (en particulier en France), il apparaîtrait absolument absurde de ne pas se conformer au système établi par Google si l'on souhaite un site bien positionné.

Enfin, le rôle du contenu dans l'amélioration d'un *PageRank* apparaît clairement. Comme on a pu le voir, les liens émanant des sites d'autorité ont un poids considérable. De cette manière, un contenu original et de qualité pourra se révéler une formidable occasion de capter l'attention des sites d'autorité qui pourront se faire le relais de ce contenu.

Partie 2 - Le contenu, pierre angulaire du référencement naturel

1. Le contenu d'un site internet sous toutes ses formes

« *Content is King* ». C'est sous cette célèbre formule qu'il est de coutume de résumer le rôle à accorder au contenu dans le succès de tout site internet. Le « *contenu est ROI* », dans son adaptation française, revêt également une signification bien plus marketing. Ici, ROI est aussi *Return On Investment*, soit « *revenu sur investissement* ». Une formule qui trouve néanmoins nombre de contestataires dans les rangs des référenceurs professionnels. Sans doute cette notion de contenu-roi apparaît-elle un peu trop restrictive, mettant au plan secondaire les piliers de la structure et de la popularité pourtant essentiels au référencement. En 2001, Andrew Odlyzko, alors mathématicien et économiste chez Bell, allait jusqu'à réfuter totalement le rôle moteur du contenu, dans un contexte certes légèrement différent de celui du référencement, en déclarant : « *Le contenu n'a jamais été le roi, il n'est pas le roi aujourd'hui* (en 2001), *et il est peu probable qu'il le soit un jour. L'internet s'est très bien débrouiller sans contenu, et peut continuer à prospérer sans lui. Le contenu aura une place sur l'internet, peut-être une place substantielle. Cependant cette place restera subordonnée à celle du monde des affaires et des communications interpersonnelles.* »[23]

S'il est vrai que la vision industrielle de A. Odlyzko envisage le « contenu » dans un sens très général, c'est à dire comme l'amalgame des œuvres culturelles, des documents et de tout autre produit d'information pouvant servir une industrie, il apparaît que le contenu, tel que nous l'envisageons, c'est à dire comme tout élément informationnel susceptible d'être présent sur un site internet (texte, image, vidéo, etc.), nous semble néanmoins incontournable pour son rôle joué à la fois pour les moteurs mais aussi pour les internautes.

23 **SALAUN**, Jean-Michel, *Le contenu n'est (décidément) pas le roi*, 25 avril 2008. http://blogues.ebsi.umontreal.ca/jms/index.php/post/2008/04/25/475-le-contenu-n-est-decidement-pas-le-roi

Ainsi, le terme de « contenu » est un des mots qui revient le plus souvent dans le vocabulaire des référenceurs. A raison puisque sans ce contenu, le site n'est finalement plus qu'une coquille vide, un contenant sans contenu, qui perd par là-même tout son intérêt auprès des visiteurs et donc des moteurs de recherche[24]. Et pourtant, un site internet qui pourra paraître vide de contenu aux yeux des moteurs ne le sera pas pour autant aux yeux des internautes (site en *Flash*), et inversement (contenu caché). Ce paradoxe tient à la notion même de contenu, une notion qui, dans le contexte d'internet, représente tout aussi bien « *ce qui est dans un contenant* » que « *ce qui est exprimé* »[25] au sein de ce même contenu (par exemple, les idées contenues au sein d'un texte lui-même contenu dans une page internet, etc.). A vrai dire, ces deux notions, non contradictoires, ont tendance à se confondre dans le contexte qui nous intéresse.

Précisons enfin qu'à l'heure de l'internet multimédia, au sein duquel cohabite une multitude de représentations de l'information, cette notion s'est étendue bien au-delà du simple contenu purement textuel. Depuis quelques années, les sources d'information se sont multipliées sur l'espace offert par la page et les internautes semblent s'être parfaitement habitués à passer d'un texte à une vidéo en faisant un saut sur une image. Le fondement hypertextuel du document *html* renforce par ailleurs ces nouvelles pratiques.

Pourtant, si du point de vue de l'internaute, un site, à travers sa dimension multimédia, forme un contenu lui-même subdivisé en différents types de contenus, tout contenu n'est pas logé à la même enseigne du point de vue des moteurs de recherche. Nous allons à présent nous attarder sur la typologie des contenus (au sens « médias » du terme) que l'on peut trouver sur internet afin d'en caractériser l'utilité pour les moteurs de recherche.

---

24 *Rappelons que le discours officiel des concepteurs des grands moteurs est d'offrir les sites les plus en adéquation avec les attentes des internautes, et donc, a fortiori, le contenu le plus pertinent.*

25 **Le Robert**, p. 207

1.1 - Le contenu textuel

Le contenu textuel est la forme d'expression la plus représentée sur internet. Quoi de plus normal finalement pour un média qui était à l'origine exclusivement composé de texte ? Aujourd'hui, même si différents autres médias viennent agrémenter la navigation des internautes, presqu'aucun site n'oserait se passer de texte si ce n'est certains sites à vocation conceptuelle dont l'ambition est de justement communiquer « autrement ».

Par ailleurs, c'est le texte et rien d'autre qui est indexé par les *spiders* des moteurs, et c'est également par le biais du texte que les internautes entrent leurs requêtes dans les formulaires des moteurs de recherche. Dans cette configuration, difficile pour un site qui espère être indexé de faire l'impasse sur un contenu textuel. Et ce n'est pas, comme nous pourrons le voir par la suite, en renseignant quelques *balises meta* que l'on pourra espérer résoudre ce problème.

Nous ne décrirons pour le moment pas davantage le traitement du contenu textuel dans le référencement. Une place plus importante y sera en effet accordée dans les prochaines parties de ce mémoire.

1.2 - Les images

A l'heure actuelle, la très grande majorité des sites internet contiennent des images. Qu'elles servent de simple illustration ou de véritable contenu central au site, ces dernières constituent autant de points d'entrée potentiels pour les internautes. D'autant plus que la recherche d'images via les moteurs de recherche représente une part considérable des activités des internautes. En juillet 2010, Benjamin Ling, directeur des produits recherche de Google, annonçait, à l'occasion de la refonte de *Google Images*, que le service avait indexé dix milliards d'images et capitalisait dorénavant un

milliard de pages vues par jour[26]. Au vue de ces chiffres, il serait donc déraisonnable ne pas prendre en compte cette dimension dans l'optimisation d'un site internet.

Il faut néanmoins en convenir : rien ne saurait nous assurer qu'une recherche d'images sur un produit particulier s'accompagne forcément d'une attention d'achat, encore moins d'une transformation sous forme d'achat. L'intérêt pour un site de e-commerce de miser une partie de son trafic sur les images reste donc a priori relativement limité. Cependant, cette incertitude vis à vis du comportement des internautes ne doit pas contraindre le responsable d'un site à mettre de côté des éléments ayant le potentiel pour amener du trafic.

Comme le référencement du contenu textuel, le référencement des images s'accompagne d'un travail d'optimisation : nom de l'image, ancre du lien et texte alternatif sont les principaux éléments requis lors de l'implémentation d'une image au sein d'un site. Les éléments textuels de la page, en particulier le texte entourant l'image ont également une grande importance. Ces diverses informations permettent au moteur de comprendre un tant soit peu ce qu'il est incapable de « voir » autrement, comme le verrait un être humain. N'oublions pas que les algorithmes des moteurs déchiffrent uniquement du code et donc, du contenu textuel.

Si les moteurs de recherche « orientés image » ne permettent pas encore de comprendre ou d'interpréter le contenu d'une image en tant que tel, de nombreuses avancées technologiques ont été réalisées ces dernières années dans le domaine de la reconnaissance des couleurs, de la disposition des éléments et des formes. Un logiciel comme *TinEye*[27] permet par exemple de trouver sur la toile les multiples utilisations qui peuvent avoir été faites d'une image-source. L'application analyse pixel par pixel l'image soumise par l'internaute pour l'identifier. Ainsi, une copie de l'image, même altéré

26 **SIEGLER**, MG, *Google Image Search: Over 10 Billion Images, 1 Billion Pageviews A Day*, 20 juin 2010. http://techcrunch.com/2010/07/20/google-image-search

27 http://www.tineye.com/

au niveau des couleurs ou de la taille, ou amputée d'une partie, est facilement retrouvée par *TinEye*. Cette application, particulièrement pratique pour la recherche des transgressions de droits d'auteur, s'avère toutefois relativement limitée dans un domaine comme le référencement, il faut en convenir.

Dans ce même ordre d'idée, depuis peu de temps, *Google Images* propose à l'internaute de combiner sa recherche d'image avec une dominante de couleur et également de filtrer les résultats par type d'images (visages, dessins, etc.). Cette mise à jour est particulièrement représentative d'une tendance générale chez les concepteurs de moteurs de recherche : celle d'ajouter aux mots-clés et aux balises accompagnant les images un tout nouvel aspect qui prend davantage en compte l'image dans sa dimension graphique, et qui la prendra peut-être, à bien plus long terme, dans sa dimension esthétique.

1.3 - Le contenu sonore et la vidéo

Grâce à la recherche universelle, les vidéos occupent aujourd'hui les pages de résultats des grands moteurs, au même titre que les images ou les liens textuels conventionnels. Cela montre l'importance prise par ce média dans le quotidien des internautes, qui n'ont plus de difficulté particulière pour lancer une visualisation de vidéo en quelques secondes. Sur le seul mois d'octobre 2009, 28 milliards de vidéos ont été visionnées sur internet par les internautes américains. Sur ces 28 milliards de vidéos, 37,7% ont été visionnées sur des sites appartenant à Google (*Youtube*, *Google Videos*)[28].

Grâce aux technologies du *streaming*, les vidéos et contenus sonores (*podcast*, musique, etc.) n'ont à l'heure actuelle, aucun mal à se faire une place de choix sur les sites internet. Des sites de partage de vidéos comme *Youtube* et *Dailymotion* et l'émergence du haut débit ont par ailleurs démocratisé

28 **Le Journal du Net**, *Monde : Vidéo et télévision numérique*, 12 juillet 2010. http://www.journaldunet.com/cc/27_video/video_marche_mde.shtml

l'implémentation à la volée d'un média auparavant beaucoup trop imposant en termes de poids pour être directement intégré dans une page internet.

Puisqu'à l'heure actuelle aucun moteur n'est encore capable de lire « dans une vidéo »[29], le référencement d'une vidéo doit donc se réaliser par le biais des métadonnées et des diverses informations contenues sur la page. Tout comme les images, le nom du fichier, ainsi que le titre, l'*URL* et la description de la vidéo tiennent place de contenu textuel, qui peut, pour sa part, être parfaitement indexé par les moteurs.

Si le média vidéo n'est en soi pas un média déterminant pour le référencement d'un site internet, il peut toutefois constituer un formidable support de promotion pour les sociétés. A titre d'exemple, depuis quelque temps, l'agence Aeronet a basé une grande partie de sa communication sur le format vidéo en donnant la parole à ses clients par le biais d'interviews. Des vidéos qui, par la suite, peuvent être diffusées sur les réseaux sociaux, ce qui permet d'accroitre d'autant plus leur visibilité et, par conséquent, la visibilité de l'agence sur internet.

Le référencement des fichiers audio suit à peu de choses près les mêmes problématiques que le référencement des vidéos. Toutefois, le média sonore, qui concerne en l'occurrence la musique, semble posséder bien moins d'importance en terme de référencement que le format vidéo, tout du moins en ce qui concerne les clients d'Aeronet.

2. Écriture journalistique et écriture web

Lorsque l'on évoque la création de contenu web, la notion d'écriture web n'est jamais loin. Depuis

29 *Des systèmes de reconnaissance vocale directement implémentés dans les services de streaming vidéo comme* Blinkx *offrent la possibilité d'effectuer des recherches sur le texte ainsi transcrit. Toutefois, ils sont encore loin d'avoir été démocratisés et n'ont donc pas encore véritablement de retombées sur le monde du référencement.*

bientôt deux décennies, l'écriture web a la lourde tâche d'adapter un modèle de communication traditionnel aux enjeux d'un média qui exige une conjugaison de l'accessibilité, de la lisibilité et du référencement.

A travers cette partie, l'occasion nous sera donnée de nous pencher quelques instants sur l'écriture journalistique qui partage avec l'écriture web un grand nombre de conventions communes mais présente également un certain nombre de divergences qu'il sera intéressant de caractériser.

2.1 - L'écriture journalistique

L'écriture journalistique répond à des conventions depuis longtemps établies et admises par toute une profession de journalistes. Il convient d'en rappeler dès à présent les principes fondateurs dont certains, comme nous le verrons par la suite, se retrouvent également dans les fondements de l'écriture adaptée aux spécificités d'internet.

Les questions clés sont la clé de voute de tout article journalistique qui se respecte. A travers son article, un journaliste doit répondre aux questions suivantes : *Qui* ?, *quoi* ?, *où* ?, *quand* ?, *comment* ?, *pourquoi* ?

Ces questions trouvent généralement leur réponse dès les premières phrases indiquant par là-même au lecteur le sujet de l'article qu'il a sous les yeux. Tout l'intérêt de cette introduction, qui doit par ailleurs être écrite de la manière la plus percutante possible, est de capter l'attention du lecteur pour le convaincre de poursuivre sa lecture.

Par ailleurs, le principe de la pyramide inversée est le moyen généralement employé pour expliquer de manière schématique la manière dont doit être structuré un article de presse écrite. Ainsi, l'essentiel de l'information doit se trouver dans le titre ainsi que dans l'introduction. Suivi du développement où davantage de détails sont révélés. Et enfin, une « chute » vient clôturer l'article

sur un bon mot ou une anecdote, suivant la nature du sujet. Le schéma ci-dessous résume le principe de la pyramide inversée.

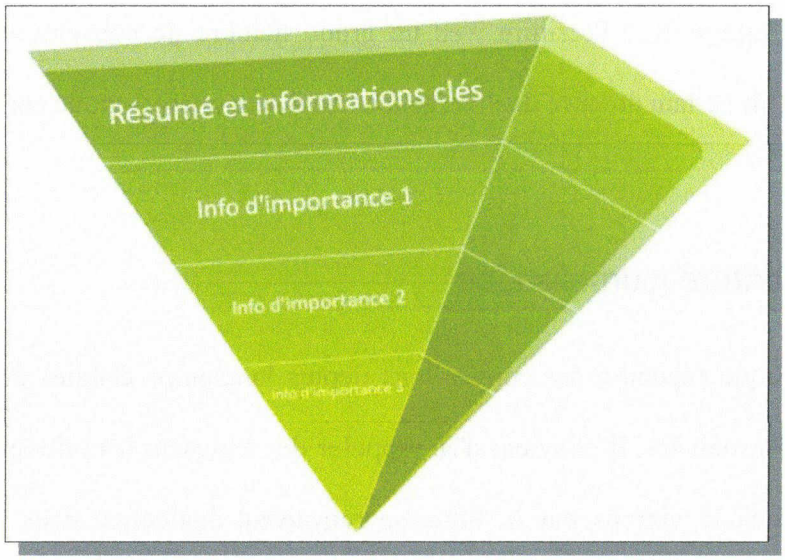

Illustration 3 - Pyramide inversée

On peut opposer cette structure à la forme classique de l'écrit romanesque où l'information est distillée au fur et à mesure du récit pour atteindre son paroxysme lors du dénouement. Cette structure en pyramide inversée se voit par ailleurs renforcée par des éléments visuels et une disposition particulière du texte et des images au sein de la page. La titraille (titre, sous-titre, chapeau) crée des points de repère visuels sur la page, permettant ainsi au lecteur de repérer dès le premier coup d'œil l'emplacement des différentes parties de l'article, sans nécessairement avoir à les lire.

Enfin, terminons ce tour d'horizon très succinct des conventions journalistiques en matière d'écriture en évoquant le respect des règles typographiques. Les règles typographiques françaises, concernent la ponctuation, les espaces ou les abréviations, et diffèrent parfois quelque peu des règles anglo-

saxonnes. Sur le web, si les erreurs de ponctuation n'ont pas véritablement d'incidences sur le référencement, elles peuvent en avoir sur la qualité de lecture des internautes. Par exemple, un texte exempt de ponctuation peut devenir un véritable frein pour le lecteur.

Bien évidemment, la mission du journaliste, qui consiste à informer le lecteur, diffère quelque peu de celle du rédacteur, terme bien plus générique, qui peut se voir confier la rédaction d'une multitude de types de contenus (d'un contenu purement promotionnel à de l'informationnel sans fondement nécessairement journalistique). Si en soi, la plupart des journalistes de presse écrite sont eux-mêmes rédacteurs, l'ensemble des rédacteurs est loin d'être constitué uniquement de journalistes. Ainsi, au sein des agences web évoluent des rédacteurs qui n'ont généralement rien à voir avec des journalistes.

2.2 - Les spécificités de l'écriture web

L'écriture web déroge finalement assez peu aux règles académiques de l'écrit journalistique. Malgré tout, l'environnement dans lequel elle s'inscrit est bien différent de celui de la presse papier. A fortiori, elle se voit confrontée à des problématiques typiques du web : un public d'internautes qui recouvre une infinité de profils différents et des habitudes en ligne bien particulières, une lecture sur écran spécifique, une logique d'hyperliens et, surtout, une prise en compte constante de l'indexation des moteurs de recherche. L'écriture web doit donc relever un double défi. Celui de plaire aux lecteurs, comme la presse papier en somme, mais dans le même temps, elle doit plaire également moteurs qui à défaut de cerveau, fonctionnent grâce à leurs algorithmes.

A propos de l'adaptation de l'écriture traditionnel au média internet, une des spécificités les plus notable concerne la lecture sur écran. Environ 25% plus lente que sur un support papier, la lecture sur écran nécessite un processus cognitif ainsi qu'une sollicitation des yeux plus intenses.

Contrairement au support papier où l'œil connait les limites physiques de la page et ne subit pas certains phénomènes liés au support (le phénomène de rémanence, les problèmes de contraste, entre autres, y sont inexistants), l'écran offre une multitude de configurations différentes. Des informations peuvent se trouver sur toute la page-écran, voire au dehors. Car là où un livre ou un journal papier constituent un espace fini, la page *html* peut s'étendre bien plus loin que l'espace offert par l'écran, voire sur plusieurs niveaux de profondeur (hyperliens). Ainsi, pour Isabelle Canivet : « *Les comportements de lecture en ligne induisent une architecture de l'information particulière, tant au niveau de la hiérarchisation (macrostructure) que de l'organisation du texte (microstructure). On devra adapter le présentation de l'information au niveau de l'architecture de l'information, en allant de l'essentiel au détail : c'est ce qu'on appelle la granularité de l'information.* »[30]

Plus que jamais, le rédacteur doit donc veiller à fournir à l'internaute l'information qu'il est venu chercher, en évitant au maximum d'inutiles intermédiaires qui pourraient le freiner dans sa quête d'information. C'est encore plus vrai dans le cas d'un site de e-commerce dont l'objectif est de convaincre l'internaute de passer commande. Pour ces raisons, le rédacteur web doit devenir bien plus qu'un simple rédacteur de contenus : « *Il* (lui) *faut acquérir des notions d'architecture de la page web, de communication, de marketing. Il s'agit également d'avoir une macrovision sur le site, afin d'intégrer la page dans une architecture plus large (celle du site, celle du Web), de la relier aux autres pages en créant un réseau, un maillage de liens.* »[31] Là réside sans aucun doute la différence la plus notable entre l'écriture journalistique et l'écriture web.

Comme on a pu s'en rendre compte, le contenu possède un statut central sur les sites internet et sur internet en général. Il suffit pour cela de considérer l'attention qui lui est quotidiennement porté dans les réflexions des référenceurs pour s'en convaincre. Pourtant, aussi paradoxal que cela puisse

30 **CANIVET**, Isabelle, *Bien rédiger pour le web... et améliorer son référencement naturel*, Éditions Eyrolles, Paris, 2010, p.23

31 *Ibid.*

paraître, ce n'est pas là l'aspect qui prévaut lors de la conception d'un site internet. Si la plupart des concepteurs de sites sont conscients du rôle prépondérant du contenu dans la réussite de leur projet, c'est un aspect bien souvent négligé en réalité.

2.3 - Les journalistes du web : du journal en ligne au blog

L'apparition de la presse en ligne a amené avec elle une cohorte de nouvelles problématiques au sein de la communauté des journalistes de la presse écrite. Habitués à une complémentarité plus qu'une concurrence avec les autres médias dits « traditionnels » (Télévision et radio), une partie des journalistes de la presse papier ont rapidement perçu internet comme une sérieuse menace pour leur travail. Un frayeur qui n'aura durée qu'un temps puisque la profession aura finalement abordé la virage de la presse en ligne sans trop de dégâts. Le web a semble-t-il davantage apporté aux journalistes qu'il ne leur a soustrait. Pour Isabelle Canivet, « *l'écriture hypertexte est une écriture de la rupture et de l'enrichissement qui rompt avec la linéarité de la lecture sur papier et qui s'enrichit du multimédia.* » Le journalisme de presse écrite a donc dû adopter cette nouvelle dimension et devenir une profession éminemment polyvalente. Selon le journaliste et blogueur Alain Joannès, « *L'esprit de polyvalence s'entend comme une agilité intellectuelle au service de l'enrichissement* »[32]. Un enrichissement de l'information qui passe par l'image, le son, la vidéo et, bien évidemment, par l'hypertextualité offerte par le support.

Les journalistes sont toujours autant lus, si ce n'est bien davantage qu'avant, si l'on conjugue la presse écrite sur support papier et la presse en ligne. Le véritable problème actuel de la presse en ligne réside finalement bien plus dans la recherche de son mode de financement. Après avoir proposé du contenu gratuit aux internautes, en échange de sollicitations publicitaires diverses, la plupart des grands quotidiens locaux, nationaux ou internationaux tentent de faire machine arrière,

32 **JOANNES**, Alain, *Le journalisme à l'ère électronique,* Éditions Vuibert, Paris, 2007.p. 37.

avec difficulté. Difficile en effet de faire payer une information aux internautes quand ceux-ci ont été habitués à la gratuité. Sans compter que la concurrence dans le secteur est particulièrement rude.

Enfin, aujourd'hui, un grand nombre de journalistes ont choisi de tenir des blogs personnels, une tendance finalement naturelle pour des professionnels qui vouent une partie de leur existence à informer leurs concitoyens. Ces blogs de journalistes, au sein desquels la parole est bien plus relâchée, sont souvent l'occasion pour les journaliste de livrer une vision de l'actualité qui flirte davantage avec la subjectivité qu'à l'accoutumée[33].

En choisissant de nous arrêter quelques instants sur la mutation du journalisme à l'heure d'internet, il s'agissait en particulier de soulever une problématique connexe à celle du traitement du contenu. En effet, le *cyberjournalisme*, pour reprendre un néologisme à succès, doit lutter avec les mêmes armes que tout autre site en quête de popularité. En raison de la prise en compte de la dimension « référencement », le journaliste du web doit donc envisager la rédaction des contenus, non plus seulement en terme purement informationnel et selon le respect de sa déontologie, mais aussi dans une optique d'optimisation, d'ouverture aux moteurs de recherche et aux médias sociaux.

3. L'obsession du bon mot clé dans le référencement

Mot clé, *mot-clé*, *mot clef* ou *mot-clef*. Quelle que soit la graphie usitée, le terme résonne dans la bouche de tous les référenceurs comme une véritable lubie. Si ce terme est tellement employé au quotidien, cela n'est en rien dû au hasard. La notion de « mot clé » est au centre de toutes les attentions car elle est le carburant des moteurs de recherche. En effet, que ce soit *Google*, *Yahoo*,

33 *Rappelons qu'une des règles de déontologie des journalistes leur donne obligation de restituer les faits de la manière la plus objective qui soit.*

Bing, ou *Orange*, tout moteur de recherche indexe avant tout des mots, parmi lesquels certains sont plus signifiants que d'autres : les « mots clés ».

3.1 - Problème de définition au sujet du terme « mot clé »

En quoi un mot clé ne serait-il pas un mot comme un autre ? Le mot clé, dans la manière où l'envisage les référenceurs et les rédacteurs du web, est directement hérité du vocable des documentalistes et des informaticiens. Une locution abusive si l'on en croit Marc Duval, bibliothécaire professionnel : « *L'incompréhension* (au sujet du terme mot clé) *est née de la confusion, surtout chez les néophytes de l'indexation, entre le «mot représentant le document» et le «mot recherché par l'internaute* »[34]. Il est vrai que le terme est aujourd'hui employé à tort et à travers, aussi bien pour qualifier les descripteurs d'un document, les requêtes des internautes que pour parler des unités sémantiques indexées par les moteurs de recherche. Si la confusion est entretenue par les référenceurs, c'est sans doute aussi parce que les problèmes qu'elle engendre ne sont pas assez significatifs dans la pratique. Un abus de langage qui au final semble irriter essentiellement les documentalistes.

Il existe en réalité deux définitions de « mot clé », l'une issue de l'informatique où le terme recouvre « *le contenu d'un document ou d'un fichier, servant de principal critère de recherche dans un fichier ou dans un système de gestion de base de données.* »[35]

La seconde définition est issue du secteur de la bibliothéconomie : « *Mot ou groupe de mots, éventuellement dans une forme lexicographique normalisée, choisi dans le titre ou le texte d'un document, caractérisant le contenu et permettant la recherche de ce document.* »[36]

Ces deux définitions se rejoignent finalement sur un point : les mots clés représentent un contenu et

34 **Duval**, Marc, *Le mot-clé*, sur DSI-info.ca, 8 janvier 2002. http://www.dsi-info.ca/mot-cle.html
35 *Ibid.*
36 *Ibid.*

servent à la recherche de ce dit-contenu. Dans le contexte d'internet, ils représentent le contenu d'une page ou d'un site, et sont utilisés dans la recherche de contenu informationnel par les internautes.

3.2 - Méthodes d'élaboration d'une liste de mots clés

Chez Aeronet, les mots clés font leur apparition dès la première prise de contact avec les clients. Une liste d'expressions est établie entre le client et le référenceur et sert de point de départ aux optimisations diverses, échanges de liens[37], rédaction de contenus et audits.

Le référenceur vérifie alors les occurrences de chaque mot clé sur le service de générateur de mots clés *Google Keywords*. Il a ainsi la possibilité de vérifier la pertinence des mots clés sur la base du nombre de requêtes entrées par les internautes sur le moteur Google. Si le service ne fournit naturellement que des résultats ayant trait à Google, la position de quasi-monopole du moteur (en France du moins) est telle qu'il est finalement inutile d'aller voir sur les moteurs concurrents, Yahoo et Bing, pour ne citer que les plus connus. Cette consultation de *Google Keywords*, qui fournit également des suggestions en rapport avec les mots clés sélectionnés, permet au référenceur de supprimer de sa liste des mots clés peu intéressants car trop peu consultés par les internautes ou trop concurrentiels pour qu'il ne faille pas espérer y apparaître avant la dixième page.

37 *Les mots clés sont utilisés dans les ancres des liens lors des échanges de liens avec des sites partenaires.*

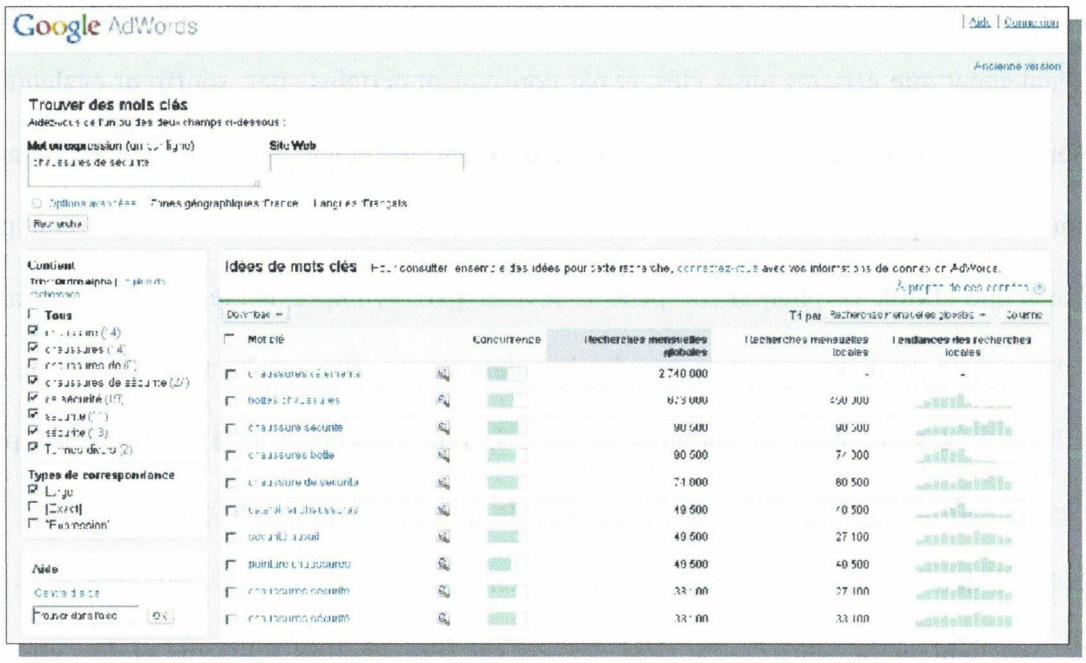

Illustration 4 - Exemple de recherche d'occurrences de mots clés

Les sites internet concurrents constituent également un champs d'investigation riche d'enseignements pour le référenceur. En observant les mots clés sur lesquels apparaissent ses concurrents au sein des moteurs de recherches ainsi que leurs *backlinks** respectifs, le référenceur se fait parfois une idée bien plus précise des choix à opérer dans sa propre sélection de mots clés. Toutefois, dans la majeure partie des cas, les clients eux-mêmes sont particulièrement renseignés sur le positionnement et les mots clés choisis par leurs concurrents.

Cette étape terminée, il ne reste plus au référenceur qu'à décliner les différentes expressions en leur ajoutant la marque du pluriel, en retirant les accents, etc. dans les limites du respect de la langue française[38]. Ainsi, à partir du mot clé « *déchet biodégradable* », on pourra décliner de cette manière : « *dechet biodegradable* », « *déchets biodégradables* », « *déchets bio* », etc. En plus de varier le contenu pour le rendre moins artificiel pour la lecture de l'internaute, cette méthode à l'avantage de travailler le phénomène de *longue traîne*.

38 *Nous verrons par la suite que le référenceur, qui base son travail sur les requêtes des internautes, est très souvent amené à flirter allégrement avec ces limites.*

Notons également que certains mots clés, et par conséquent certains sites, souffrent également d'un phénomène de saisonnalité. Ainsi, en fonction du moment de l'année, certains produits ou activités sont naturellement plus recherchés que d'autres. Pourtant, contrairement à ce que l'on pourrait penser de prime abord, ce phénomène peut se révéler parfois imprévisible. A titre d'exemple, un client d'Aeronet, spécialiste de la livraison de chocolats à domicile, a eu l'amère surprise de ne plus pouvoir honorer ses commandes durant la période la plus chaude de l'été 2010 car ses produits subissaient les affres de la chaleur. Les recherches de mots clés liés à la thématique de la livraison de chocolats à domicile (par exemple: « *livraison chocolat* », « *commander chocolat* ») ont elles-mêmes considérablement chuté durant cette période car les potentiels acheteurs étaient eux-mêmes découragés de commander du chocolat. Cette anecdote est révélatrice de la multitude de situations à laquelle peut se retrouver confronter un référenceur.

3.3 - Les particularités du français qui mettent à mal le référencement

La langue française est assurément une langue complexe. Un signe qui ne trompe pas : bien peu de francophones savent parfaitement orthographier ou parler leur langue, tout du moins tel que les « immortels » de l'Académie française souhaiteraient qu'elle soit parlée. Internet n'échappe évidemment pas à cette règle et aurait même tendance à grossir le trait car si l'on peut techniquement se passer de l'écriture au quotidien, il n'en est plus question lorsque l'on s'aventure sur la toile.

Il n'est toutefois pas question ici de faire une analyse exhaustive des comportements des internautes en matière d'orthographe ni de jeter l'opprobre sur ces derniers. Le devoir des référenceurs n'est pas d'éduquer les internautes sur les rudiments de la langue mais bien de s'adapter à la manière d'écrire de ces derniers.

3.3.1 - L'accentuation des mots clés en cause

L'omniprésence des nombreux accents présents dans la langue française n'arrange en rien les affaires des référenceurs. Si, d'un point de vue technique, un moteur de recherche ne fait pas de différence entre les majuscules et les minuscules, il en est tout autrement des caractères accentués et de leurs équivalents non-accentués. Il suffit d'effectuer des requêtes sur des expressions comportant des lettres accentuées et des requêtes sur les mêmes expressions sans les accents pour se rendre compte que les résultats affichés sur la page de résultats ne sont pas les mêmes.

Nul oubli quant à cette spécificité française chez les moteurs pourtant, il s'agit-là au contraire d'un choix qui se justifie aisément d'un point de vue purement sémantique : un certain nombre de mots deviennent en effet ambigus dès lors qu'ils perdent leur accentuation. A ce propos, l'avis de l'Académie française, qui concerne en particulier l'accentuation des majuscules mais qui vaut également pour les minuscules, est le suivant : « *Il convient cependant d'observer qu'en français, l'accent a pleine valeur orthographique. Son absence ralentit la lecture, fait hésiter sur la prononciation, et peut même induire en erreur. Il en va de même pour le tréma et la cédille.* »[39]

Par exemple, « *un individu degenere* » pourrait correspondre à la fois à « *un individu dégénère* » mais aussi à « *un individu dégénéré* », ce qui n'a pas la même signification. L'absence d'accents ne nous permet pas de pouvoir choisir parmi les deux significations possibles. Bien évidemment, une phrase complète nous permettrait de désambiguïser la proposition mais les requêtes des internautes sont rarement rédigées comme des phrases complètes.

A défaut d'accentuation, il est alors aisé de prendre un terme pour un autre, d'autant plus que les

39 **Académie française**, *Questions de langue*.
http://www.academie-francaise.fr/langue/questions.html#accentuation

mots clés entrés par les internautes sont généralement sortis de leur contexte lors des requêtes sur les moteurs. Pour réduire le champs des possibilités, et offrir aux internautes les résultats qui correspondent le mieux à leurs attentes, le moteurs respectent donc les différences entre les mots accentués et non-accentués.

Là où ce détail du fonctionnement des moteurs devient véritablement problématique pour les référenceurs, c'est qu'il existe deux grands types d'internautes : ceux qui respectent du mieux qu'ils le peuvent l'accentuation des mots et ceux qui écrivent sans les accents.

Pour une expression donnée, par exemple « *fete des meres* » et « *fête des mères* », les pages de résultats de Google ne sont sensiblement pas les mêmes. Ainsi, on remarque des variations sur le positionnement de certains sites. Il suffit également de comparer les expressions « *referencement lille* » et « *référencement lille* » pour constater que le nombre de résultats va du simple au double : 450 000 pour la première expression et 927 000 pour la seconde. Cette spécificité nationale (problème que ne possèdent pas les pays anglophones) complique donc quelque peu le travail des référenceurs qui doivent jongler entre des termes qui, s'ils se rapportent au même mot, ne sont pas considérés comme un même mot clé par le moteur. Chez Aeronet, l'emploi des termes non-accentués (mais dont la bonne graphie réclame un ou plusieurs accents) s'effectue généralement dans la balise *<title></title>* du document ou dans le pied de page du site, c'est à dire à un emplacement discret qui ne gène pas la lecture de l'internaute.

Précisons enfin que la prise en compte des accents est d'autant plus importante dans le cas du référencement payant où l'annonceur achète un emplacement sur un mot clé, à l'accent près. Un annonceur qui sera placé sur « *poupée* » ne le sera donc pas automatiquement sur « *poupee* ». L'annonceur pourra néanmoins apparaître sur le terme qu'il souhaite. Il sera donc possible d'acheter

« *poupee* » mais de rédiger une annonce ayant pour titre le mot clé « *poupée* »

3.3.2 - Fautes d'orthographe

Les fautes d'orthographe[40] sont un autre problème des plus épineux pour le référencement naturel. Un problème qui, par ailleurs, semble au premier abord bien plus insoluble que celui précédemment évoqué. Car, en partant du principe que la quasi-totalité des internautes n'écrivent pas parfaitement, il y a donc potentiellement une infinité de configurations possibles dans l'écriture des requêtes des internautes. Évidemment, certaines expressions sont plus sujettes aux mauvaises orthographes que d'autres, ne serait-ce qu'en raison de leur relative complexité. Il suffit d'ailleurs d'observer les sources de trafic offertes par *Google Analytics* pour constater l'exotisme de certaines requêtes.

De son côté, Google prend en considération depuis longtemps cette problématique. Le moteur détecte automatiquement la plupart des fautes d'orthographe courantes et propose les résultats avec la requête correctement orthographiée. Cela n'empêche toutefois pas les référenceurs de prendre en compte cette réflexion dans leur travail quotidien. Que cela concerne les fautes d'orthographe ou les problèmes d'accentuation, le référenceur tente d'agir avec méthode. Par exemple, il a la possibilité d'écrire les titres ou une partie de la balise *<title>* en majuscule afin de se soustraire discrètement des accents. En ce qui concerne les fautes d'orthographe, Isabelle Canivet propose de son côté[41] la création sur le site internet de glossaires ou dictionnaires qui incluent les fautes courantes. De cette manière, les fautes d'orthographes sont indexées sans faire perdre de la crédibilité au site.

D'une manière générale, référenceurs comme rédacteurs se doivent de respecter au maximum la langue dans laquelle ils rédigent car même s'ils sont en mesure de gagner quelques places en

40 *Nous séparons volontairement problèmes d'accentuation et fautes d'orthographe même si, en soi, une faute sur un accent constitue bel et bien une faute d'orthographe aux yeux de l'Académie française.*

41 **CANIVET**, Isabelle, Bien rédiger pour le web... et améliorer son référencement naturel, Éditions Eyrolles, Paris, 2010, p. 321

truffant les sites de fautes d'orthographe, ils mettent dans le même temps gravement en péril l'image de la société représentée par le site.

3.4 – Le phénomène de longue traîne

Terminons cette partie consacrée aux mots clés en évoquant le principe de *longue traîne*. Ce principe part du simple constat qu'un nombre très restreint de mots clés capitalise une somme relativement importante d'accès aux sites via les moteurs de recherche par rapport à leur quantité. Inversement, un très grand nombre d'autres mots clés, bien moins entrés par les internautes représentent un ratio bien plus élevé de visites sur les sites, il s'agit de la longue traîne. Le principe de longue traîne, initialement observé par Chris Anderson, s'observe donc également dans les statistiques d'accès des sites internet.

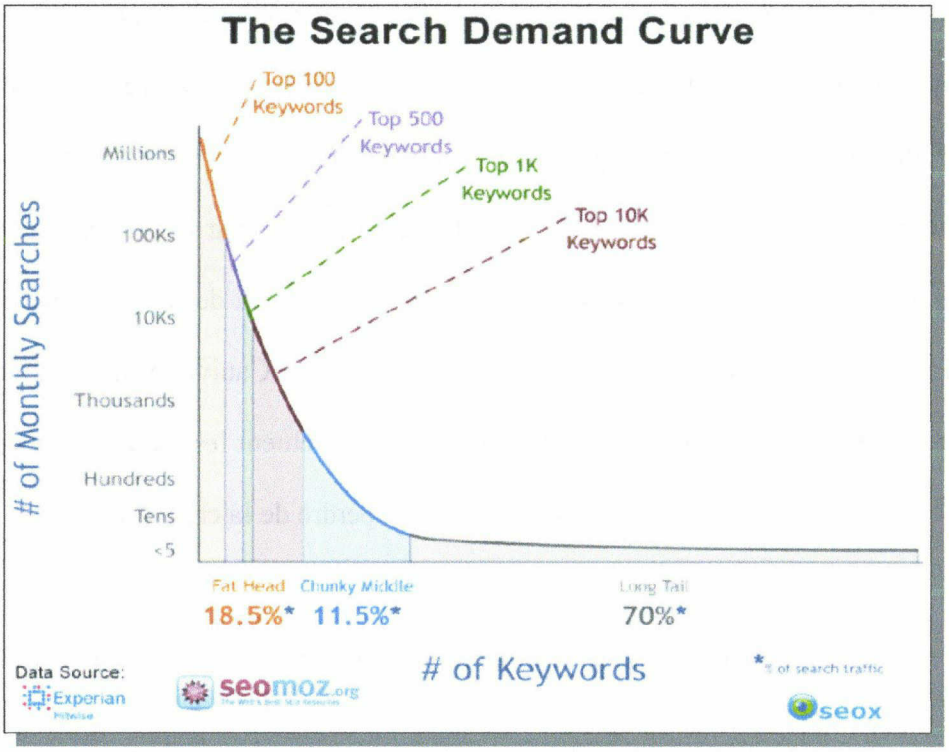

Illustration 5 - Phénomène de longue traîne

Il y a donc certes un nombre restreint de mots clés qui monopolisent à eux-seul l'essentiel du marché tels que « *poker* », « *logiciel gratuit* », « *achat voiture* », etc. Mais, la somme des mots clés plus spécifiques ou contenant des fautes d'orthographe, comme « *jouer au poker gratuitement* », « *télécharger des logiciels pc* » ou encore « *acheter une voiture occasion* » capitalise au total bien plus de recherches. D'autant plus qu'il est bien plus aisé de se positionner sur des termes spécifiques que sur un terme plus générique qui cannibalise également les efforts des sites concurrents.

Pour un responsable de site, la question n'est finalement pas tant de savoir s'il souhaite privilégier les termes génériques et fédérateurs (mais sur lesquels la concurrence fait rage) ou plutôt les termes plus spécifiques, mais plutôt de savoir en quelle proportion employer l'un et l'autre. Un référenceur va naturellement privilégier les termes génériques puisqu'il serait absurde de s'en passer au vue de l'audience qu'ils amènent, mais dans le même temps, il essaiera de varier le titre des pages, d'introduire quelques synonymes dans le contenu afin de constituer un vivier de mots clés assez diversifié.

Partie 3 - Plaire moteurs, plaire aux internautes

Lors de la première partie de ce mémoire, nous avons abordé la notion de référencement au sens général du terme. Puis, dans la seconde partie, nous nous sommes longuement attardés sur les notions de « *contenu* » et de « *mot clé* », de leurs différentes typologies à ce qu'ils engendraient dans la pratique.

Comme nous avons pu le voir précédemment, le contenu, tel qu'il est présenté par les professionnels du web en général, est véritablement sublimé dans sa fonction, loué pour son pouvoir de séduction envers les internautes et sa qualité distinctive envers les moteurs. Pourtant, plaire aux moteurs tout en plaisant aux internautes est un objectif qui flirte souvent avec les contradictions. Dans la pratique, rédiger un contenu qui puisse convenir aux deux entités est loin d'être si simple. Au sein de cette ultime partie, nous allons tenter de comprendre ce qu'est un contenu idéal en nous plaçant du point de vue des moteurs de recherche pour ensuite nous placer du point des internautes. Enfin, après en avoir présenté les spécificités de ces deux dimensions, nous vérifierons s'il peut y avoir existence d'un véritable consensus d'écriture qui puisse contenter à la fois les moteurs et les internautes.

1. Plaire aux moteurs

Les concepteurs des moteurs de recherche se targuent de vouloir proposer aux internautes un service de qualité qui leur fournira des résultats pertinents en fonction de leurs requêtes. Rappelons que l'algorithme de chaque moteur possède ses spécificités et que les référenceurs font en sorte d'en comprendre du mieux qu'ils le peuvent le fonctionnement par le biais de la *rétro-ingénierie**. De cette manière, et également par le biais de ce que les concepteurs veulent bien révéler du fonctionnement de leur moteur, un certain nombre des comportements à adopter face aux moteur

font à peu près consensus entre les professionnels (et certains amateurs éclairés du référencement). De nombreux autres aspects, généralement liés à l'évolution des algorithmes, sont pour leur part sujets à des discussions parfois vivaces entre les référenceurs, conférant à cette profession une certaine vitalité et une constante remise en question.

Tout du moins, certaines bonnes pratiques sont immuables et d'une manière générale, les référenceurs un minimum consciencieux usent tous d'une manière relativement semblable quand il s'agit de respecter les préconisations des moteurs.

1.1 - Pourquoi plaire aux moteurs ?

De toutes les missions qui incombent au référenceur, il en est une qui résume toutes les autres et qui constitue un véritable leitmotiv pour toute son activité : celle de plaire aux moteurs. Il convient d'admettre que le verbe « plaire » est forcément abusif puisqu'il sous-entendrait que les moteurs de recherche fassent appel à leur goût, dont ils n'ont assurément pas la notion.

Il faut toutefois toujours garder à l'esprit que le moteur de recherche est incontournable en ce sens où, telle une interface entre l'internaute et le contenu, il puise dans son index les pages indexées par ses robots pour les présenter à l'internaute suivant un ordre établi par son algorithme[42]. Le fonctionnement de ce moteur, qui justifie à lui-seul l'existence-même du métier de référenceur, est donc devenu bien naturellement le centre de toutes les réflexions chez les professionnels.

Enfin, rappelons qu'être indexé sur un moteur de recherche est une chose essentielle, mais y être bien positionné en est une autre. Et c'est là tout l'enjeu du rapport de séduction qui s'opère entre les sites internet et les moteurs. Un rapport de séduction finalement très déséquilibré si l'on se fie au véritablement pouvoir de vie ou de mort que possèdent les moteurs de recherche sur les petits sites

42 *Le fonctionnement d'un moteur de recherche est évidemment bien plus complexe mais cet aspect du référencement ne constituant pas le sujet du mémoire, nous n'entrerons volontairement pas davantage dans les détails techniques.*

de e-commerce, qui sans existence dans les moteurs, sont de ce fait inexistants pour les internautes.

1.2 - Comment plaire aux moteurs de recherche ?

Plaire à un moteur de recherche dépend grandement du moteur en question. Ne possédant pas les mêmes algorithmes, les moteurs ne requièrent pas forcément les mêmes règles. Toutefois, certaines règles, si elles ne possèdent pas la même impact chez tous les moteurs, ont un caractère universel.

Un balisage optimisé

Le balisage du texte est un élément primordial dans l'élaboration d'un document à destination d'internet. Chaque balise se veut un indicateur qui caractérise et explicite la fonction d'une partie de texte au sein du contenu global présent sur la page. Par exemple, grâce aux balises *<h></h>* , le *spider* peut déduire la structure générale d'un document : Parties, sous-parties, etc. De la même manière, la balise ** indique l'existence d'un élément d'importance, généralement un mot clé, qui est caractéristique du sujet abordé au sein du texte. Dernier exemple, les balises *
* ou *
* indiquent la présence d'une liste ordonnée. Ces divers éléments confèrent donc un minimum de sens au document, lui permettant d'être mieux interprété par les *spiders* et d'être plus agréable pour la lecture des internautes (quand il est associé à une feuille de style).

Dans le cas inverse, un texte brut, sans aucun balisage, se verrait tout de même indexer par les moteurs mais ses points saillants, titres en tête, auraient bien moins d'importance aux yeux des algorithmes au risque même d'être tout simplement « noyés » dans le reste du contenu. A titre de comparaison, ce serait comme fournir à un lecteur un texte journalistique dont l'ensemble du contenu textuel, titres inclus, est parfaitement uniformisé pour ne former qu'un bloque de texte sans aucun point d'accroche visuel. Le balisage du texte possède donc une importance capitale puisqu'il participe à la structuration visuelle et surtout, dans le cas qui nous intéresse, à la structuration

logique du contenu.

Des phrases simples

Par ailleurs, les moteurs aiment les phrases simples, pour la simple raison qu'ils ont beaucoup de mal à comprendre des phrases complexes incluant plusieurs propositions ou des doubles négations, par exemple. Il n'est malheureusement pas toujours évident de fournir des titres à la fois simples, informatifs et incluant des mots clés intéressants. D'autant plus, comme on le verra par la suite, qu'il ne faut au passage pas faire l'impasse sur l'internaute.

Dans le même ordre d'idée, il convient de limiter au maximum l'utilisation des phrases qui font appel aux dimensions sémantique et pragmatique de la linguistique. Même si on constate des avancées significatives sur la plupart moteurs de recherche en matière de sémantique, les algorithmes sont encore loin de « comprendre » une question comme est capable de le faire un être humain. Des recherches ambitieuses sont tout de même menées sur le sujet du *web sémantique** mais l'on touche là à une problématique réservée à des technologies bien spécifiques, beaucoup plus ambitieuses mais encore en pleine conception, qui ne touchent pas la sphère du grand public. *Wolfram*[43], lancé il y a un peu plus d'un an, est l'exemple typique du moteur de recherche à vocation sémantique, dont la conception diffère grandement de celle des principaux moteurs du marché.

On s'imagine sans peine que la dimension pragmatique, encore plus complexe que la dimension sémantique car ayant recours au contexte de la phrase, mobilisée par exemple dans l'emploi de l'humour ou des références culturelles, est encore loin de faire son apparition dans les algorithmes de nos moteurs.

Pour résumer, l'idéal serait très prosaïquement de s'adresser aux moteurs comme on le ferait avec un

43 *Wolfram est disponible dans sa version Alpha à l'adresse suivante : http://www.wolframalpha.com*

enfant, c'est à dire en utilisant des mots simples, des phrases courtes, sans équivoques, ni double sens, ne recourir à aucune métaphores, images, ni comparaisons qui nécessiteraient une connaissance culturelle ou une véritable capacité d'abstraction. Évidemment, en plus d'être inadéquate et excessivement pauvre, cette manière de s'exprimer n'est quasiment employée par personne sur internet.

1.3 - Comment déplaire aux moteurs de recherche ?

Après avoir résumé les techniques générales visant à faciliter l'indexation du contenu des sites par les moteurs, il nous faut tout naturellement en venir aux meilleurs moyens d'agir en sens contraire.

Le meilleur moyen de se mettre à dos la plupart des moteurs est principalement d'user de techniques dites *Black Hat**. Bien qu'elles n'aient en elles-mêmes rien d'illégales - de nombreux blogs[44] ayant pignon sur rue traitent d'ailleurs du sujet avec force détails - les techniques de *Black Hat* sont tout de même non-conformes aux règles établies par Google, par exemple. Les pénalités que risquent les récalcitrants pris en flagrant délit peuvent aller du simple malus de positionnement, à la *sandbox**, jusqu'au véritable blacklistage du site en « infraction » avec les règles de Google. Autant dire qu'un référenceur professionnel qui serait amené à pratiquer ces techniques risquerait fortement de mettre en péril le positionnement du site dont il s'occupe.

La plupart des techniques de *Black Hat* sont liées à l'exploitation des failles qui existent dans tout algorithmes. Elles peuvent avoir divers objectifs comme l'amélioration du positionnement d'un site mais aussi la mise à mal du positionnement ou de l'image d'un site concurrent. Néanmoins, la plupart des référenceurs autoproclamés *Black Hat* semblent se contenter de mener des « tests » de leurs techniques par le biais de sites-test créés justement pour l'occasion. Toutefois, en mettant de

44 *Voir entre autre, le très influent SEO Black Out :* http://www.seoblackout.com/

côté les considérations purement éthiques et les distinctions *White Hat*/Black Hat*[45], il tombe tout de même sous le sens de proscrire certaines pratiques. Les exemples de techniques suivants ont été sélectionnés car ils entretiennent une relation évidente avec les notions de contenu abordées précédemment.

Keyword stuffing

Le *Keyword stuffing, spamdexing* ou bourrage de mots clés est généralement utilisé en associations aux pages satellites. Il consiste à insérer dans une page une somme considérable de mots clés au mépris de la lecture et du sens des phrases. La plupart du temps, ces listes discontinues de mots clés sont cachées aux internautes par le biais de jeux sur la typographie (police blanche sur fond blanc, par exemple) ou alors elles sont dissimulées hors de la page « visible ». Leur seul objectif est de positionner un site sans véritable effort de rédaction. Durant mon stage, il m'est arrivé souvent de découvrir un bourrage de mots clés maladroitement dissimulé derrière une images ou dans des pages pourtant relativement consultées (dans le contenu d'une page « à propos » en l'occurrence). Cette technique, en plus d'être aujourd'hui complètement inefficace, car facilement détectée par les moteurs, jette un certain discrédit sur la société dont le site en est truffé, d'autant plus quand elle est pratiquée dans le dos des clients par des professionnels peu scrupuleux.

Typosquatting

Cette technique consiste à littéralement occuper des noms de domaine dont la graphie est extrêmement proche d'un nom de domaine connu et fréquenté. Dit autrement, il s'agit de profiter des fautes de frappes et fautes d'orthographe des internautes pour en tirer des bénéfices ou porter atteinte à l'image d'une société. Un exemple de *typosquatting* consisterait par exemple à acheter le nom de domaine « *googl.fr* » afin de profiter du trafic à l'origine destiné à « *google.fr* ». Cette technique fait donc directement écho aux problèmes d'orthographes des internautes déjà soulevé

45 *De nombreuses techniques étant par ailleurs à la frontière entre les deux tendances.*

dans une partie précédente.

Duplication de contenu

Enfin, la duplication de contenu consiste en une répétition du même contenu sur des pages différentes, qu'elles se trouvent au sein d'un seul et même site (on parle d'auto-duplication) ou bien sur des sites différents. A titre d'exemple, le traitement de Google vis à vis de ces pages peut varier mais généralement si plusieurs pages dont le contenu est identique sur un même site sont détectées par le moteur, une seule page sera accessible dans la page de résultats.

Indépendamment des sites dont le contenu est volontairement dupliqué de pages en pages, il existe un grand nombre de raisons qui peuvent amener les sites dans ce genre de situations : Contenu dupliqué automatiquement (très courant dans les sites qui incluent la technologie *Php*), site accessible avec et sans les *www*. Des réajustements au niveau du code peuvent alors venir en aide au référencement du site. Parfois, il suffit simplement d'indiquer aux moteurs de ne pas indexer une page pour en privilégier une autre.

La meilleure solution, néanmoins quand cela est possible, reste encore de fournir à chaque page du site un contenu original ou, tout du moins, une part importante de contenu original. Comme on le verra par la suite, dans la pratique, il n'est pas toujours simple de varier le contenu d'un site, d'autant plus si le domaine du site en question ne s'y prête pas.

2. Plaire aux internautes

Nous avons vu que plaire aux moteurs de recherche était avant tout une question de respect des préconisations des concepteurs, de bon sens vis à vis du fonctionnement des moteurs et de mise à l'écart des diverses pratiques pouvant freiner un bon référencement.

Mais quels sont les ingrédients pour plaire aux internautes, qui ne sont pas constitués d'algorithmes, mais sont doués au contraire de capacités de compréhensions linguistiques avancées, en particulier en matière de sémantique ? Des capacités qui, par ailleurs, divergent selon chaque internaute. C'est au travers de cette partie que nous allons tenter de caractériser les différents moyens de comprendre les internautes afin de leur apporter un contenu qui soit adapté à leur attentes et à leurs spécificités.

2.1 - Dans la tête de l'internaute

Pour réussir à s'adresser à l'internaute, faut-il encore comprendre comment ce dernier fonctionne : quelles sont ses attentes ? Qu'est-ce qui le motive à se rendre sur un site plutôt que sur un autre ? Quelle est sa manière de lire une page ? Comprend-t-il toutes les informations mises à sa disposition ? Etc.

Autant de questions qui ne sont pas à prendre à la légère car aussi bien référencé est un site internet, son succès n'est en aucun cas assuré si un certain nombre de facteurs viennent freiner les internautes dans leur progression sur le site.

2.1.1 - Le langage des internautes

Aussi étrange cela puisse paraître de prime abord, le référenceur ou rédacteur de contenu est constamment confronté au langage des internautes. Évidemment il n'y a pas langage universel. On pourrait même aller jusqu'à dire qu'il y a autant de langages qu'il y a d'internautes, voire davantage puisqu'un même internaute s'exprimera d'une manière différente selon l'endroit dans lequel il s'exprime, le forum de discussion auquel il participe, tout comme un être humain n'usera pas du même vocabulaire et des mêmes tournures de phrases en fonction de son interlocuteur. Dans ces conditions, comment parvenir à tirer une règle générale d'une multitude de comportements ?

On peut par exemple tirer certains enseignements en analysant la manière dont les internautes parviennent sur un site internet, en particulier observer avec attention les requêtes qu'ils tapent dans le champs de saisie des moteurs de recherche et leur satisfaction. En l'occurrence, le service *Google Analytics* offre un grand nombre de statistiques qui permettent d'avoir une idée relativement précise et exhaustive du comportement des internautes sur un site internet. De cette manière, il est possible, par exemple, de se rendre compte qu'une page est certes consultée mais subit un *taux de rebond** élevé. Cette situation peut avoir des origines diverses mais on peut parfois y déceler un contenu mal adapté au public car peu informatif ou utilisant un vocable inadapté à la cible.

Bien que des études sur le comportement des internautes aient été menées par le passé[46], regroupant de nombreuses statistiques diverses et variées, il n'existe pour le moment aucune étude qui se soit attelée à décrire en profondeur les différentes manières de s'exprimer des internautes. Aussi, les décisions des référenceurs reposent exclusivement sur leurs observations quotidiennes (forums, sites concurrents, etc.), sur leur expérience, sur les résultats offerts par des outils comme *Google Keywords* ou *Google Analytics*, mais aussi, il est vrai, sur une certaine dose de bon sens. Si ces outils ne sont pas suffisants pour restituer l'entièreté des compétences et des comportements linguistiques des internautes, ils en offrent tout de même une vision assez fidèle pour aiguiller les référenceurs dans leurs décisions.

2.1.2 - A chaque site internet son public

Finalement, s'il était besoin de formuler une règle qui prévaudrait pour tous les sites, ce serait que chaque site internet à son public et que c'est à ce public qu'il faut s'adresser en priorité. En effet, comme nous le rappelle Isabelle Canivet, « *Ne pas définir sa cible est une erreur courante dans les*

46 **SEINSIGHTS**, *L'évolution du comportement des internautes entre 2007 et 2010*, 5 février 2010. http://www.seinsights.com/emarketing/l-evolution-du-comportement-des-internautes-entre-2007-et-2010

projets web actuels. On fait un site pour tout le monde, ce qui revient à dire : pour personne. »[47] C'est d'autant plus véridique pour les sites de e-commerce et les sites vitrines (qui constituent tout deux l'essentiel de la clientèle de l'agence Aeronet) qui ont pour la plupart une activité qui s'adresse à une audience bien particulière.

Ainsi, un site qui vend des chaussures de ville ou des chaussures de sport ne va naturellement pas employer le même vocabulaire qu'un site qui vend des chaussures de sécurité car la cible visée n'est tout simplement pas la même. Pour autant, cela ne veut pas dire qu'il faille employer un vocable inutilement complexe mais il ne faut pas non plus hésiter à utiliser les mots qui conviennent le plus à son audience. C'est dans ce cas précis que la liste de mots clés établie en association avec le client s'avère la plus précieuse. Étant donné que la plupart des prestations de référencement s'effectuent à posteriori de la mise en ligne du site, le référenceur ne peut pas véritablement compter sur la définition de profils d'utilisateur, ou *personas*, c'est-à-dire une méthode mise en place par Alain Cooper en 1999 qui consiste à créer des utilisateurs factices, ce qui permet aux concepteurs de sites « *d'évoquer la cible utilisateur de manière concrète et proche de la réalité* »[48]. En règle générale, le référenceur doit donc se fier aux seules informations fournies par le responsable du site ainsi qu'à sa propre expérience pour cibler la clientèle des sites sur lesquels il travaille.

2.2 - La plus-value d'un contenu orienté utilisateur

Offrir à l'internaute un contenu conçu à son intention est une véritable gageure pour un site internet. Il est évidemment bien plus simple de ne s'adresser qu'aux moteurs à grand renfort d'optimisations qui vont rarement de pair avec la subtilité. Beaucoup de sites internet semblent oublier que leur audience est constituée d'êtres humains. En effet, les moteurs n'éprouvent à proprement parler aucun

47 **CANIVET**, Isabelle, *Bien rédiger pour le web... et améliorer son référencement naturel*, Éditions Eyrolles, Paris, 2010, p.42

48 **BOUCHER**, Amélie, *Ergonomie Web – Pour des sites web efficaces (2ème édition)*, Editions Eyrolles, Paris, 2010, p.78

intérêt pour le contenu qu'ils indexent, leur finalité n'étant que de fournir aux internautes ce qu'ils recherchent. L'internaute, par contre, est constamment en quête d'un contenu de qualité, tout du moins d'un contenu qui s'accorde avec ses attentes. Faire l'impasse sur l'internaute au profit du moteur est en cela un non-sens total, car aussi bien positionné soit un site internet, il n'est d'aucune utilité s'il ne s'adresse pas à une catégorie d'internautes, même infime.

2.2.1 - Susciter l'intérêt de l'utilisateur

Pour un site de e-commerce, le principal objectif d'un contenu est de susciter de l'intérêt chez l'utilisateur. Indépendamment de son positionnement dans les pages de résultats des moteurs, un site possède plusieurs solutions qui peuvent lui permettre de se différentier de ses concurrents : proposer des prix compétitifs, des prestations supplémentaires (Garanties, SAV, etc.), une charte graphique attirante ou qui met l'acheteur en confiance, etc. C'est malheureusement souvent loin d'être la préoccupation première des responsables de sites de e-commerce. Ainsi, comme le souligne Amélie Boucher, « *Il y a souvent un fossé entre, d'une part les efforts que déploient les boutiques pour vendre des produits dans le monde physique et, d'autre part, le manque de préoccupation de la qualité des descriptifs produits sur le Web.* » [49]

Le contenu fait pourtant partie des éléments véritablement différenciateurs qui stimulent la décision d'achat. Il a d'autant plus d'intérêt qu'il bénéficie également aux moteurs, qui apprécient particulièrement les sites qui présentent un contenu de qualité.

2.2.2 - Le contenu comme image de marque

Pour les grandes sociétés qui capitalisent énormément sur leur image et allouent en ce sens une part considérable de leur budget à la publicité, la qualité du contenu est déterminante. « *L'image positive*

49 *Ibid.*, p. 286

ou négative repose sur une information partielle et partiale (toutefois estimée complète et objective par l'acheteur) et est susceptible d'influencer la décision d'achat. (...) Malgré sa démarche rationalisée, l'acheteur est en effet sensible à des éléments d'image tels que la pérennité supposée du fournisseur, son professionnalisme, le classement obtenu dans telle revue professionnelle, son expertise dans tel secteur, sa nationalité, sa qualité et la fiabilité de son offre. [50]» Autant dire qu'un certain nombre de ces éléments (professionnalisme et fiabilité de l'offre en tête) sont fortement liés à la présence d'un contenu de qualité.

Cela vaut aussi bien pour les multinationales, les grandes marques que pour les petits sites de e-commerce, qui sont au centre de notre réflexion. On peut même inférer que dans le cas d'une PME, la qualité du contenu du site est d'autant plus importante qu'elle peut véritablement contrebalancer son manque de notoriété dans son secteur d'activité. Le prix, la prestation, les délais de livraison, etc. constituent de leur côté un argument supplémentaire mais toutefois décisif quant à la décision finale de l'acte d'achat.

De plus, si pour un média non-figé comme internet, les fautes d'orthographes et de grammaires semblent moins déterminantes que dans le cas d'une publicité sur format figé (tel que la publicité dans les magazines), elles ne sont pas pour autant à prendre à la légère. Elles constituent de ce fait généralement un des premiers champs d'investigation du référenceur quand il a la main sur le site. Durant mon stage, j'ai eu la charge de corriger le contenu de quelques sites qui, en plus d'être truffés de fautes d'orthographe dans une proportion véritablement alarmante, étaient pour ainsi dire rédigés dans un français totalement approximatif. Cette situation est-elle l'indication d'une véritable incompétence en matière de rédaction de la part de l'auteur ou plutôt d'une prise à la légère d'un aspect pourtant essentiel de la communication du site ? Il semble que ce soit plutôt vers la seconde option qu'il faille chercher les raisons de la présence de ce genre de contenus bâclés. Ainsi, comme

50 **MALAVAL**, Philippe, **BENAROYA**, Christophe, *Marketing Business to Business : Du marketing industriel au marketing d'affaires*, Éditions Pearson Éducation, Paris, décembre 2005, p. 294

le constate amèrement Isabelle Canivet : « *La création de contenu est considérée comme accessoire, confiée à des personnes en interne, désignées volontaires. (...) Comme tout le monde sait écrire sans, a contrario, avoir forcément des compétences en graphisme ou en développement, la rédaction des contenus est prise en charge par n'importe qui ou presque, souvent sans aucune sensibilisation à l'écriture web et à ses enjeux.* »[51]

Les clients d'Aeronet n'entrent certes pas dans la catégorie des sites à forte notoriété, néanmoins, ils ont tout autant besoin que n'importe quel autre site d'un contenu de qualité, reflétant leur activité, capable de promouvoir leurs produits et, surtout, de laisser aux internautes une image professionnelle de la société.

2.2.3 - Fidéliser sa clientèle grâce au contenu

Bien évidemment, un site de e-commerce est constamment en quête de nouveaux clients. Nous avons vu que à quel point le contenu pouvait être déterminant dans l'acquisition de nouveaux acheteurs. Mais n'oublions pas qu'il l'est tout autant pour la fidélisation de la clientèle qui aurait par ailleurs déjà été convaincue par une transaction passée.

Le renouvellement du contenu peut apporter une véritable plus-value à un site internet. Pour des raisons liées au référencement du site, ces actualisations ne doivent toutefois pas avoir lieu trop souvent. De plus, indépendamment du référencement, l'actualisation régulière d'un site employant la technologie *html* peut rapidement devenir contraignante en matière de temps et de moyens mobilisés.

La meilleure solution, souvent privilégiée par les sociétés qui souhaitent entretenir une grande proximité avec leur clientèle sans toutefois mettre à jour quotidiennement leur site vitrine, est la

51 **CANIVET**, Isabelle, *Bien rédiger pour le web... et améliorer son référencement naturel*, Éditions Eyrolles, Paris, 2010, p. 2.

mise en ligne d'un blog. Sans risque pour le référencement du site principal, il permet en outre une communication spontanée autour des nouveaux produits, des prestations ou d'autres sujets en relation avec l'activité de la société. Par exemple, les responsables de *Gaspard et Zoé*[52] ont choisi de faire la promotion de leur site par le biais d'un blog[53] lancé quelques mois avant la mise en ligne effective du site de e-commerce. Le contenu n'est ici pas directement lié au site mais sa finalité en est la même : communiquer sur la marque et les produits du site. L'agence Aeronet elle-même possède un blog ainsi qu'une page *Facebook* qui lui permettent d'offrir régulièrement du contenu d'actualité aux internautes, ce qu'il serait difficile à réaliser avec le seul site vitrine de l'agence.

2.3 - Comment créer du contenu ?

L'idéal pour le référenceur est de pouvoir mettre en place les stratégies de référencement en amont de la mise en ligne du site. Il s'agit-là d'un luxe malheureusement bien rare. Il est en effet autrement plus fréquent que les demandes de prestation de référencement s'effectuent un certain nombre de mois ou d'années après la mise en ligne d'un site, souvent pour corriger le mauvais positionnement de ce dernier. Dans ce cas de figure, c'est donc au référenceur de s'adapter aux limitations auxquelles il fait face. Ces limitations ne touchent pas seulement l'aspect technique[54] des sites, elles sont parfois liées au contenu-même du site. Ainsi, il arrive qu'un client ne souhaite pas que l'on modifie sa ligne éditoriale, ce qui est en soi un vœu totalement compréhensible, et que le contenu déjà existant constitue ainsi le point de départ du travail des référenceurs/rédacteurs.

2.3.1 - Densité de mots clés dans une page

D'une manière générale, que la rédaction des contenus s'effectue en partant de zéro ou par rapport à

52 *Voir http://www.gaspardetzoe.com*
53 *Voir http://www.gaspardetzoe.com/fr/le-blog*
54 *Parfois, un site est difficilement modifiable en raison de son architecture ou de sa technologie. C'est en particulier le cas des sites mis en place à partir de solutions « clé en main », comme celles fournies par la société Oxatis.*

un contenu déjà existant, le référenceur s'appuie sur la liste de mots clés élaborée en coopération avec le client. L'objectif n'est pas de remplir la page de mots clés (n'oublions pas que la page est tout à fait accessible aux internautes) mais d'avoir une bonne densité de mots clés sur la page par rapport aux expressions sur lesquelles on souhaite apparaître. La densité d'un mot clé se calcule comme le nombre d'apparition de ce mot clé par rapport à la somme totale des mots présents sur la page. Ainsi, si le mot clé « *voiture* » est utilisé 10 fois dans un texte composé de 100 mots, sa densité sera de 10%.

Il n'existe pas de densité qu'on peut véritablement qualifier d'idéale. Il convient de répéter un mot clé un nombre de fois suffisant pour qu'il soit pris en compte par les moteurs tout en essayant de maintenir une écriture naturelle qui n'apparaît pas artificielle (en particulier pour le respect de l'internaute mais également pour ne pas être suspecté de *spamdexing* par les moteurs). Dans la pratique, cette règle de bon sens est loin d'être aisée à mettre en place. En particulier, pour des textes qui réclament des mots clés très spécifiques. Il apparaît en effet bien plus naturel de placer à plusieurs reprises le mot clé « *chaussures de sécurité* » dans un texte plutôt que plusieurs occurrences du mot clé « *surchaussure en polyéthylène* », pour la simple raison que l'utilisation d'un mot qui sort de l'ordinaire retient davantage l'attention du lecteur, lui demandant un effort cognitif supérieur en comparaison d'un mot couramment employé. Il en va de même des « *mots " chauds ", provocants ou à connotation sexuelle* »[55]. La difficulté effective de répéter certains mots clés est en cela fortement liée à la nature-même du mot clé en question.

Pour remédier au problème, il peut être envisageable de recourir de temps à autres à des synonymes. En plus de bénéficier à la *longue traîne*, les synonymes vont naturellement fluidifier le propos en évitant certaines répétitions trop artificielles. Il faut néanmoins veiller à ce que le contenu ainsi créé

55 **CANIVET**, Isabelle, *Bien rédiger pour le web... et améliorer son référencement naturel*, Éditions Eyrolles, Paris, 2010, p.71.

reste compréhensible, relativement naturel et surtout en rapport avec la thématique abordée. Il ne s'agit donc en aucun cas d'utiliser des synonymes en dépit du bon sens.

2.3.2 - Création de pages de contenu

Il existe certains cas de figure où l'ajout de contenus est limité voire impossible en raison d'une structure de site qui ne s'y prête pas. Le référenceur est alors contraint d'insérer du contenu dans de nouvelles pages, appelées « pages de contenu ». Il arrive que la fonction de ces pages soient rapprochées de celle des *pages satellites*. D'un certain point de vue, c'est exact car pages satellites comme pages de contenu ont pour objectif de créer de nouvelles portes d'entrée pour les internautes mais aussi de la « matière à indexer » pour les moteurs.

Une distinction fondamentale néanmoins : les pages de contenu, tout du moins telles qu'elles sont conçues chez Aeronet, ne sont pas cachées aux internautes. Contrairement aux pages satellites, ces pages n'ont pas pour objectif de rediriger automatiquement l'internaute vers le contenu principal du site, au contraire, elles sont pleinement inscrites dans l'architecture du site et généralement accessibles sur la page d'accueil. C'est pour cette raison que leur contenu est élaboré selon la même rigueur que peut l'être le contenu des pages principales du site. Elles constituent donc un apport bénéfique pour le site mais demandent en contrepartie une charge de travail relativement importante étant donné que le référenceur doit trouver une thématique qui s'accorde avec le site (et qui n'y est pas déjà abordée) et rédiger le contenu tout en tenant compte des mots clés à insérer. De plus, étant donné que ces pages se présentent comme un contenu supplémentaire, elles doivent généralement être validées par le client avant leur mise en ligne.

Il est également primordial de respecter la structure des pages principales afin que l'internaute qui accède au site par l'une de ces pages de contenu n'ait pas la sensation de se trouver sur une page isolée du reste du site. Ainsi, par exemple, en gardant le même menu de navigation que sur le reste

du site, l'internaute n'aura aucun mal à comprendre qu'il se trouve sur un site de e-commerce et non pas sur un « simple » site proposant uniquement du contenu textuel. En effet, l'internaute qui se rend sur un site de e-commerce s'attend naturellement à trouver un menu de navigation, la mention d'un panier d'achat, etc. et ce, quelle que soit la porte d'entrée qu'il empreinte pour se rendre sur le site. Le priver de ces repères essentiels ne peut être que contre-productif. La page de contenu est donc indissociable du reste du site en terme de structure et d'ergonomie.

Enfin, ces pages sont également l'occasion pour le référenceur d'insérer des liens hypertextes au sein-même du contenu. En plus d'être bénéfiques pour l'indexation du site (une page isolée a beaucoup moins de poids qu'une page pleinement inscrite dans l'architecture du site), les liens hypertextes consolident le maillage du site et offrent dans le même temps de nouvelles options de navigation à l'internaute.

2.3.3 - Les autres éléments de contenu

Jusqu'ici, la notion de contenu à laquelle nous nous sommes intéressés concernait surtout le texte de la page, visible des moteurs comme des internautes. Il ne faut toutefois pas oublier qu'une page *html* n'est pas un simple conteneur de texte faisant sens uniquement pour l'internaute. En effet, un certain nombre d'éléments de la page *html* sont situés à la frontière entre le contenu et le code, en ce sens où les informations qui s'y trouvent, bien que rédigées dans un langage relativement naturel, sont pourtant davantage destinées aux moteurs qu'aux être humains.

Passons rapidement sur les ***balises méta*** qui, du point de vue des professionnels, ont complètement perdu de leur intérêt. Elles étaient effectivement d'une grande utilité quand les moteurs étaient encore relativement incapables d'interpréter le contenu d'une page. On y indiquait, entre autres, des informations sur la page comme les mots clés (*meta keywords*), la description (*meta description*),

l'auteur de la page (*meta author*), etc. Certains clients d'Aeronet restent toutefois persuadés de leur utilité dans le positionnement de leur site et ne daignent pas faire l'impasse dessus. Dans ce cas de figure, le travail du référenceur se concentre en particulier sur la constitution d'une balise *meta description* tout de même capable susciter l'intérêt de l'internaute (la description est affichée dans les *snippets** des moteurs de recherche). Elle n'est toutefois techniquement en rien obligatoire puisque Google se charge lui-même de fournir une description sur la base du contenu présent sur la page.

Le **titre des pages** (balise *<title></title>*) est pour sa part un élément essentiel dans le référencement d'un site. En plus de constituer une accroche pour les internautes qui parcourent les pages de résultats des moteurs de recherche, son contenu a du poids lors de l'indexation des moteurs. Ainsi, chaque page d'un site requiert un titre différent et significatif du contenu de la page. On y inclus généralement des mots clés correspondant à la thématique ou à la catégorie relayée sur la page. Il n'est pas rare qu'un titre générique se voit dupliqué sur l'ensemble d'un site, par faute de temps ou par simple méconnaissance de l'impact de la balise *<title></title>* sur le référencement. Ce qui est en soi aussi dommageable que la présence d'un contenu bâclé sur le site.

Les **ancres des liens** constituent pour leur part la partie émergée des liens, leur aspect saillant. C'est non seulement par le biais de ce mot clé que le moteur va parcourir le lien mais c'est aussi grâce à des accroches efficaces que l'internaute va être poussé à cliquer sur le lien. Notons qu'une ancre de lien qui ne correspond pas au contenu du site visé (par exemple, choisir le mot clé « *voitures d'occasion* » pour un lien menant vers un site de poker) pourra paraître fortement suspecte aux yeux des moteurs.

Enfin, un autre élément souvent négligé, les attributs *alt* des images n'ont finalement d'utilité que

pour les internautes chez qui l'image ne s'afficherait pas ou chez les personnes utilisant une synthèse vocale. Toutefois, cet élément n'a vraiment d'utilité que quand l'image se révèle significative pour la compréhension ou l'illustration d'un document. Si elle n'a qu'une fonction décorative, il apparaît parfois bien plus pertinent de faire l'impasse sur un contenu alternatif qui ne viendrait qu'embrouiller la lecture du document.

2.4 - Ergonomie du contenu

Enfin, afin de clore cette partie consacrée aux différentes manières de plaire aux internautes, il nous semblait nécessaire d'aborder l'aspect lié à l'architecture de l'information et par extension à l'ergonomie des contenus. S'il est vrai que l'intérêt d'un contenu se situe essentiellement dans les informations qu'il contient et la plus-value qu'il est susceptible d'apporter à un site internet, il n'est malheureusement plus d'une grande utilité s'il ne s'inscrit pas dans une conception fonctionnelle du site murement réfléchie. Il ne s'agit pas non plus de placer le contenu à la va-vite sur la page en espérant qu'il soit remarqué par l'internaute au détour de sa navigation. Au contraire, l'idéal est que l'internaute fasse le moins d'effort cognitif possible pour discriminer un contenu présent sur la page.

Sans avoir la prétention d'entrer dans des détails qui nous demanderaient un développement trop important et qui nous éloignerait par la même occasion de notre thématique initiale, on pourrait résumer l'importance de l'ergonomie des contenus en indiquant que la prise en compte de la dimension ergonomique ne constitue pas qu'un simple fard esthétisant appliqué sur le contenu, mais qu'il s'agit au contraire d'une véritable mise en valeur du contenu au sein de laquelle sont pris en compte la théorie des formes, les lois de proximités, etc., parmi les nombreux concepts invoqués dans le domaine de l'ergonomie web.

Tout comme le contenu lui-même, les considérations d'ordre ergonomique sont parfois mises de

côté ou au dernier plan par les *webmasters* lors de la conception d'un site internet. Pour Amélie Boucher, spécialiste de l'ergonomie des sites internet, « *le mieux est de considérer les problématiques d'ergonomie comme étant aussi importantes que le reste, au début, au milieu, à la fin d'un projet web, bref, en permanence* »[56]. De cette manière, un site bien conçu aura pour résultat un contenu mis en valeur, accessible, agréable à lire, s'inscrivant parfaitement dans une continuité fonctionnelle et d'autant mieux armé pour donner pleine satisfaction à l'internaute.

Et c'est plus généralement la structuration du site qui doit recevoir une intention particulière de la part des concepteurs de sites. Un contenu difficilement accessible à cause d'un site mal-conçu, tant sur le plan de l'architecture que de l'accessibilité, ne peut pleinement remplir son rôle prépondérant.

56 **BOUCHER**, Amélie, *Ergonomie Web – Pour des sites web efficaces (2ème édition)*, Editions Eyrolles, Paris, 2010, p.33.

Conclusion

Le référenceur, quand il se double du rôle de rédacteur de contenu, se retrouve tiraillé entre, d'une part, sa volonté de respecter les conformités induites par les exigences et les limitations technologiques des moteurs de recherche, et d'autre part, le souhait d'apporter aux internautes un contenu qualitatif dont le fond et la forme s'éloignent de l'artificiel et du préformaté. Dans un premier temps, cette situation cornélienne nous pousserait très naturellement à penser qu'il faille opérer un choix inconditionnel pour l'une ou l'autre option. La réalité se trouve pourtant dans un consensus. Un consensus qu'il est obligatoire d'appréhender pour le référenceur s'il désire à la fois atteindre ses objectifs et ne pas léser la cible d'internautes de chacun des sites dont il s'occupe.

En l'absence d'un web véritablement sémantique, qui serait par là-même en mesure de comprendre le contenu d'un document sans en réclamer une simplification parfois bien malheureuse, le référenceur-rédacteur doit en somme éviter à tout prix de penser que la finalité de son travail est le moteur de recherche. Au contraire, le moteur de recherche doit être envisagé comme une étape sur le chemin qui mène à l'internaute. Une étape certes nécessaire et ô combien décisive pour la création de trafic, mais une « simple » étape tout de même. C'est principalement à cause de cette vision embrouillée de la finalité du métier de référenceur que l'existence de deux cibles a priori inconciliables m'avait fait grandement douté du traitement réservé au contenu des sites internet. En résumer, il semble que penser à l'internaute en premier lieu en incluant évidemment les spécificités des moteurs soit la solution la plus raisonnable qui s'offre au référenceur. Ne doutons pas que le web sémantique puisse à l'avenir apporter une nouvelle issue à cette situation tout de même frustrante.

Terminons enfin en soulevant un paradoxe. Les concepteurs des moteurs de recherche, les têtes pensantes derrière Google en particulier, n'ont de cesse d'inciter les responsables de sites à se

démarquer grâce à des contenus originaux, capables de constituer une véritable valeur ajoutée pour les internautes. Cependant, d'un autre côté, le respect des préconisations issues de ces mêmes concepteurs poussent les rédacteurs à homogénéiser leur contenu afin de le rendre le plus attractif possible pour le passage des *spiders*. Ce constat est également partagé par Pascal Beria qui va même jusqu'à se demander si le web ne contribue pas à « *l'appauvrissement du style journaliste* »[57]. Il est vrai qu'en optimisant leur contenu et principalement leurs titres, les journalistes du web font également perdre un peu de ce qui faisait leur personnalité et par extension le charme de la presse écrite dite « traditionnelle ».

57 **BERIA,** Pascal, *Le web contribue-t-il à un appauvrissement du style journalistique ?*, 25 novembre 2009. http://pascalberia.wordpress.com/2009/11/25/le-web-contribue-t-il-a-un-appauvrissement-du-style-journalistique/

Glossaire

Balise title : Sert à indiquer le titre de la page, qui s'affichera dans la partie haute du navigateur ainsi que dans les résultats des moteurs de recherche.

Balises meta : Ensemble de balises servant à renseigner les moteurs sur le contenu de la page. Elles se placent dans la balise *<head></head>*. Elles n'ont aujourd'hui quasiment plus d'utilité.

Black Hat : Ensemble des techniques utilisant les failles d'un moteur en vue d'une amélioration d'un positionnement ou du torpillage d'un site concurrent.

Backlinks (ou *liens retour* ou *liens entrant*) : Liens issus de sites extérieurs. S'opposent aux liens sortant, qui pointe un autre site.

Page satellite : Page créée uniquement dans le but de générer artificiellement du trafic. Une fois rendu sur une page satellite, l'internaute est généralement redirigé vers le site en lui-même.

PageRank : Système de classement des pages d'un site internet en grande partie basé sur la pertinence des liens entrant présents sur les dites-pages.

Redirections en cascade : Succession de redirections. Un internaute est redirigé sur une page elle-même redirigée sur une autre page. Les redirections en cascades peuvent se révéler problématique pour les moteurs.

Rétro-ingénierie: Méthode d'analyse d'un objet, par exemple d'un moteur de recherche, afin d'en

comprendre le fonctionnement et l'utiliser à bon escient.

Robot.txt : Simple fichier texte placé à la racine d'un site donnant des indications aux spiders (ne pas indexer certaines pages par exemple).

Sandbox : État momentané durant lequel un site récent n'est plus positionné à cause de méthodes de référencement trop rapides ou trop agressives (beaucoup de *backlinks* en très peu de temps). Ce phénomène est considéré comme de l'histoire ancienne par certains professionnels.

Snippet : Court extrait d'une page présent dans les pages de résultats des moteurs de recherche.

Spiders ou crawlers : Noms donnés aux robots chargés de naviguer de lien en lien afin d'indexer le contenu des sites internet.

Taux de rebond: Indicateur clé de performance illustrant l'attitude générale des internautes à leur arrivée sur un site. Un taux de rebond élevé indiquera que la majorité des internautes quitte le site immédiatement après avoir visité une page, et un taux faible indiquera qu'ils continuent leur navigation sur d'autres pages. L'interprétation que l'on donne à ce pourcentage dépend grandement du type de site.

Web sémantique : Ensemble de technologies encore à l'étude qui visent à apporter une dimension sémantique, c'est-à-dire un véritable sens, au contenu présent sur le web.

White Hat : Ensemble des techniques conformes aux recommandations des moteurs.

Bibliographie

- **ANDRIEU**, Olivier, *Réussir son référencement web* (2ème édition), Éditions Eyrolles, Paris, 2010, 450 pages. ISBN : 9782212126464
- **CANIVET**, Isabelle, *Bien rédiger pour le web... et améliorer son référencement naturel*, Éditions Eyrolles, Paris, 2010, 418 pages. ISBN : 9782212124330
- **JOANNES**, Alain, *Le journalisme à l'ère électronique,* Éditions Vuibert, Paris, 2007. ISBN : 2711787109
- **BOUCHER**, Amélie, *Ergonomie Web – Pour des sites web efficaces (2ème édition)*, Éditions Eyrolles, Paris, 2010, 460 pages. ISBN : 9782212124798
- **MALAVAL**, Philippe, **BENAROYA**, Christophe, *Marketing Business to Business : Du marketing industriel au marketing d'affaires*, Éditions Pearson Éducation, Paris, décembre 2005, 704 pages. ISBN : 9782744071232

Webographie

- **Académie française**, *Questions de langue*. [en ligne] (consulté le 20 juin 2010) http://www.academie-francaise.fr/langue/questions.html#accentuation
- **BERIA**, Pascal, *Le web contribue-t-il à un appauvrissement du style journalistique ?*, sur Connexion[s], 25 novembre 2009. [en ligne] (consulté le 5 août 2010) http://pascalberia.wordpress.com/2009/11/25/le-web-contribue-t-il-a-un-appauvrissement-du-style-journalistique/

- **DUVAL**, Marc, *Le mot-clé*, sur DSI-info.ca, 8 janvier 2002. [en ligne] (consulté le 15 juin 2010)
  http://www.dsi-info.ca/mot-cle.html
- **Netcarft**, *May 2010 Web Server Survey*, 14 mai 2010. [en ligne] (consulté le 22 juin 2010)
  http://news.netcraft.com/archives/2010/05/14/may_2010_web_server_survey.html
- **Internet World Stats**, *France Internet Usage Stats and Telecom Reports*. [en ligne] (consulté le 22 juin 2010)
  http://www.internetworldstats.com/eu/fr.htm
- **La Fevad**, *Baromètre e-commerce ICE - 1er trimestre 2009*, 12 mai 2009. [en ligne] (consulté le 26 juin 2010)
  http://www.fevad.com/index.php?option=com_content&task=view&id=497&Itemid=861
- **Le Journal du Net**, *Europe : Population d'internautes*, 14 juin 2010. [en ligne] (consulté le 26 juin 2010)
  http://www.journaldunet.com/cc/01_internautes/inter_nbr_eu.shtml
- **Le Journal du Net**, *Monde : Vidéo et télévision numérique*, 12 juillet 2010. [en ligne] (consulté le 19 juillet 2010)
  http://www.journaldunet.com/cc/27_video/video_marche_mde.shtml
- **SALAUN**, Jean-Michel, *Le contenu n'est (décidément) pas le roi*, 25 avril 2008. [en ligne] (consulté le 15 juillet 2010)
  http://blogues.ebsi.umontreal.ca/jms/index.php/post/2008/04/25/475-le-contenu-n-est-decidement-pas-le-roi
- **Scanblog**, *Baromètre ECHO : les français face aux médias sociaux et la e-réputation*, 14 avril 2010. [en ligne] (consulté le 24 juin 2010)
  http://www.scanblog.com/blog/etude-echo-ereputation-scanblog-172/
- **SEINSIGHTS**, *L'évolution du comportement des internautes entre 2007 et 2010*, 5 février

2010. [en ligne] (consulté le 27 juin 2010)

http://www.seinsights.com/emarketing/l-evolution-du-comportement-des-internautes-entre-2007-et-2010

- **SIEGLER**, MG, *Google Image Search: Over 10 Billion Images, 1 Billion Pageviews A Day*, sur TechCrunch.com, 20 juin 2010. [en ligne] (consulté le 14 juillet 2010)

http://techcrunch.com/2010/07/20/google-image-search/

Annexes

Les sites sur lesquels j'ai travaillé durant mon stage

- **Abisco** : Protections individuelles pour les professionnels

  http://www.abisco.fr/

- **Abitasud** : Agence immobilière du Tarn

  http://www.abitasud.com/

- **Blog d'Aeronet**

  http://blog.aeronet-referencement.fr/

- **Alex Location** : Location de voitures en Guadeloupe

  http://www.locations-voitures-guadeloupe.com/

- **Bébé Maman Nature** : Produits naturels pour bébés et femmes enceintes

  http://www.bebe-maman-nature.com/

- **Bonus Casino en Ligne**

  http://www.bonus-casino-en-ligne.info/

- **La Boutique du Mariage** : Organisation de mariage dans le Nord-Pas-de-Calais

  http://www.lbdm.fr/

- **Céline Robert** : Créateur de chapeaux pour hommes et femmes

  http://www.celinerobert.com/

- **Chocadom** : Livraison de chocolats à domicile

  http://www.chocadom.com/

- **C Mon Bio** : Vente de produits bio

  http://www.cmonbio.com/

- **Discount Mariage** : Accessoires et décoration pour le mariage

  http://www.discount-mariage.fr/

- **Elyatel** : Opérateur mobile international

  http://www.elyatel.com/

- **Encre Hermane** : Vente de cartouches d'encre et toners

  http://www.encre-hermane.fr/

- **Eurotrade Pallet** : Recyclage de palettes et emballage industriel

  http://www.eurotradepallet.com/

- **La Boutique Exelgreen** : Vente de gazon artificiel

  http://www.gazonstar.com/

- **Façades et traditions** : Rénovation de façades

  http://www.facades-et-traditions.com/

- **Feel-Zen** : Fontaines et murs d'eau

  http://www.feel-zen.com/

- **Fêtes et réceptions** : Décorations pour les fêtes

  http://www.fetes-et-receptions.com/

- **Gaspard et Zoé** : Créateur de vêtements pour bébés et enfants

  http://www.gaspardetzoe.com/

- **Gite Roche des Ducs** : Location de gîtes dans les Vosges

  http://www.giterochedesducs.com/

- **Guy de Hainaut** : Voix off professionnelle

  http://www.guydehainaut.be/

- **I.D. Graphique** : Communication visuelle

  http://www.idgraphique.fr/

- **Jardin Bio Harmonia** : Protection biologique pour jardin

  http://www.jardin-bio.com/

- **Jeu Légal France** : Poker et paris sportifs légaux

  http://www.jeu-legal-france.fr/

- **JFC Bien-être** : Spas et Saunas

  http://www.jfc-bienetre.com/

- **Kaplan Aspect** : Cours de langues à l'étranger

  http://www.aspect-france.fr/

- **Kis Jardins** : Création de jardins et aménagement intérieur

  http://www.06kisjardins.fr/

- **Magellan Organisation** : Organisation de séminaires en France et à l'étranger

  http://www.magellan-organisation.com/

- **Mes souliers** : Boutique de chaussures

  http://www.mes-souliers.com/

- **Méthode Colin** : Méthode d'apprentissage du piano

  http://www.methode-colin.com/

- **Musclez-vous** : Compléments alimentaires

  http://www.musclezvous.com/

- **OOGarden** : Vente de meubles et accessoires pour le jardin

  http://www.oogarden.com/

- **Plushtoy** : Vente de peluches et doudous

  http://www.plushtoy.fr/

- **Printpromo** : Impressions en grand quantité

  http://www.printpromo.fr/

- **Ruchers de Lorraine** : Produits apicoles

  http://www.ruchersdelorraine.com/

- **Superkoala** : Organisation de voyages en Australie

  http://www.superkoala.com

- **VLP Concept** : Stages de pilotage automobile

  http://www.stage-pilotage-automobile.com/

- **Vinifete** : Vaisselle jetable et vins

  http://www.vinifete.com/

Articles rédigés et parus sur le blog Aeronet

- Grosse mise à jour pour Bing Webmaster Tools, 3 Août 2010
  http://blog.aeronet-referencement.fr/index.php/veille-referencement/grosse-mise-a-jour-pour-bing-webmaster-tools-1776/
- Le Google Images nouveau est arrivé, 23 juillet 2010
  http://blog.aeronet-referencement.fr/index.php/moteurs-de-recherche/google-images-mis-a-jour-1740/
- A quand une extension classée « X » ?, 1er juillet 2010
  http://blog.aeronet-referencement.fr/index.php/veille-referencement/extension-xxx-1701/
- Aeronet lance ses outils de référencement gratuits, 23 juin 2010
  http://blog.aeronet-referencement.fr/index.php/actu/outils-de-referencement-gratuits-1665/
- T.co, le raccourcisseur d'URL de Twitter, 10 juin 2010
  http://blog.aeronet-referencement.fr/index.php/reseaux-sociaux/t-co-raccourcisseur-url-twitter-1645/
- Budget référencement du mois de Mars 2010, 26 avril 2010
  http://blog.aeronet-referencement.fr/index.php/budget/budget-referencement-mars-2010-1553/
- Google Suggest s'améliore et se localise, 23 avril 2010
  http://blog.aeronet-referencement.fr/index.php/moteurs-de-recherche/google-suggest-localisation-correction-1539/

Résumé

Le contenu possède une importance capitale dans la réussite d'un site internet. D'un côté, les moteurs de recherche indexent du contenu textuel et, de l'autre, les internautes utilisent les moteurs pour trouver essentiellement du contenu textuel. Pourtant, rédiger un contenu qui puisse répondre aussi bien aux spécificités techniques des moteurs qu'aux exigences des internautes n'est pas chose aisée. La faute à deux cibles aux attentes parfois contradictoires. Du rôle du contenu au statut des mots clés dans le référencement naturel jusqu'à l'analyse des moyens à disposition du référenceur pour la création du contenu, ce mémoire jette un regard sur le quotidien du référenceur mué en rédacteur. De cette analyse découle une question essentielle : Qui du moteur ou de l'internaute constitue réellement la finalité du travail de référenceur ?

Descripteurs : REFERENCEMENT – REDACTION DE CONTENU – MOT CLE – INTERNAUTE – SITE INTERNET – COMMERCE ELECTRONIC

Abstract

Content is an important element in the success of websites. On the one hand, search engines index textual content, and on the other hand, internet users use search engines in order to find essentially textual content. However, it's a difficult work to writing content which is suitable for internet users and search engines. It is due to two contradictory targets. The aim of this dissertation is to analyse the role of keywords and content in the search engine optimization, then the digital content creation process. «What is the real purpose of the search optimizer ? Are the search engines or the internet users ? » is the question which results from this analyse.

Keywords : SEARCH ENGINE OPTIMIZATION – CONTENT WRITING – KEYWORD – INTERNET USER – WEBSITE – ELECTRONIC COMMERCE