

Sébastien CARON

X 99 – Hec-Entrepreneurs 2003
sebastien.caron@polytechnique.org

**Intérêt du référencement et des
référéncieurs pour un site Internet
d'entreprise**

Avant-propos

Ce travail s'inscrit dans le programme du mastère HEC-Entrepreneurs, mastère que j'ai réalisé en tant qu'école d'application de l'Ecole Polytechnique. Il s'agit d'une thèse professionnelle réalisée sous la double tutelle d'HEC-Entrepreneurs et de l'Ecole Polytechnique.

La thèse professionnelle HEC-Entrepreneurs doit répondre à quatre critères :

- Un travail de réflexion personnelle et non un simple rassemblement d'informations déjà disponibles ;
- Une thèse professionnelle, c'est à dire un travail sur un aspect concret de la vie des entreprises et non une recherche fondamentale de type universitaire ;
- Une thèse susceptible d'intéresser un large public de dirigeants et non la monographie d'une seule entreprise ;
- Un travail de haut niveau tant sur le fond que sur la forme, susceptible de contribuer au renom d'HEC-Entrepreneurs.

La thèse représente donc un véritable travail de recherche reposant sur le développement d'idées argumentées avec pertinence.

J'ai pour ma part choisi de m'interroger sur un des nouveaux métiers apparus avec le développement d'Internet : celui des professionnels du référencement. J'ai plus particulièrement choisi de répondre à deux questions :

- Une entreprise ayant un site Internet, quel qu'il soit, ou projetant d'en avoir un, doit-elle se préoccuper du référencement de celui-ci ?
- Une entreprise ayant conscience de l'importance de son référencement dans les outils de recherche doit-elle faire appel à des professionnels du référencement ?

Ma thèse s'adresse donc à des dirigeants d'entreprises soucieux de l'efficacité de leur présence sur Internet à travers leur site web, dirigeants n'ayant pas obligatoirement de connaissances particulières en ce domaine, et encore moins dans le domaine du référencement.

J'ai choisi d'étudier des entreprises et sites Internet de tout type, sans me focaliser sur une quelconque famille autre que celle des entreprises ou sites français, ou destinés à un public français, afin d'intéresser un public aussi large que possible.

Pour chacune de ces deux problématiques, j'ai pris comme point de départ des éléments prouvés déjà existants mais pas forcément déjà publiés (statistiques, algorithmes de fonctionnement des outils de recherche, techniques de référencement), puis je les ai analysés concrètement grâce à une approche terrain (efficacité et coûts réels des techniques, avantages et inconvénients des offres réelles des professionnels), afin d'en déduire mes réponses, mon opinion sur ces problématiques.

Il s'agit d'un travail de réflexion personnelle qui concerne un aspect concret de la vie des entreprises, et qui est susceptible d'intéresser un large public de dirigeants.

Sébastien Caron

Sommaire

I. <u>UNE ENTREPRISE DOIT-ELLE SE PRÉOCCUPER DU RÉFÉRENCIEMENT DE SON SITE ?</u>	7
A. <u>QUEL EST L'INTÉRÊT DU RÉFÉRENCIEMENT POUR UN SITE INTERNET ?</u>	7
i. <u>Quelques notions de base</u>	7
ii. <u>Quelques statistiques et tendances intéressantes</u>	8
iii. <u>Extrapolation du nombre moyen de visiteurs par site web</u>	11
iv. <u>Étude du type d'accès aux sites Internet français : l'importance des outils de recherche</u>	13
B. <u>QUEL EST L'INTÉRÊT D'UNE OPTIMISATION DU RÉFÉRENCIEMENT D'UN SITE INTERNET ?</u>	15
i. <u>Nombre de pages web visitées par un internaute au cours d'une recherche</u>	15
ii. <u>Nombre de pages de résultats visualisées par un internaute au cours d'une recherche</u>	17
iii. <u>Études de quelques durées significatives</u>	19
C. <u>COMMENT FONCTIONNE LE RÉFÉRENCIEMENT ?</u>	21
i. <u>Les principaux moteurs de recherche et annuaires</u>	21
ii. <u>Les caractéristiques principales de ces moteurs et annuaires</u>	23
1. <u>Google</u>	23
2. <u>Yahoo</u>	23
3. <u>Wanadoo / Voila</u>	24
4. <u>AOL France</u>	24
5. <u>MSN France</u>	25
6. <u>Lycos France</u>	25
7. <u>Nomade</u>	25
iii. <u>Aperçu des offres proposées par les professionnels du référencement</u>	26
1. <u>La mise en place d'une stratégie</u>	26
2. <u>Le suivi du référencement</u>	27
II. <u>ANALYSE DE L'EFFICACITÉ ET DU COÛT DES PRINCIPALES TECHNIQUES D'OPTIMISATION DU RÉFÉRENCIEMENT : LES PROFESSIONNELS SONT-ILS INDISPENSABLES ?</u>	27
A. <u>DANS QUELS MOTEURS DE RECHERCHE ET ANNUAIRES FAUT-IL ÊTRE PRÉSENT ?</u>	28
i. <u>Récapitulatif des principaux moteurs et annuaires français : tarifs et audiences</u>	28
ii. <u>Faut-il payer pour être indexé par les moteurs de recherche ?</u>	30
iii. <u>Faut-il payer pour être indexé par les annuaires ?</u>	30
iv. <u>Quel est l'intérêt des offres proposées par les professionnels du référencement ?</u>	32
B. <u>QUELLE EST L'IMPORTANCE DE L'OPTIMISATION PAR MOTS-CLÉS D'UN SITE INTERNET ?</u>	34
i. <u>Rappel du fonctionnement des moteurs de recherche : l'importance des mots-clés</u>	34
ii. <u>Comment choisir les mots-clés importants ?</u>	35
iii. <u>Quel est l'intérêt des offres proposées par les professionnels du référencement ?</u>	37
C. <u>QUELLE EST L'IMPORTANCE DE L'OPTIMISATION DE L'ORGANISATION D'UN SITE INTERNET ?</u>	38
i. <u>Cas de Google : qu'est-ce que le PageRank ?</u>	38
ii. <u>Cas de Google : comment optimiser le PageRank ?</u>	39
iii. <u>La structure de fichiers a-t-elle un rôle dans le référencement ?</u>	42
D. <u>QUELLE EST L'IMPORTANCE DE L'OPTIMISATION DU CODE ?</u>	44
i. <u>Quelles sont les « balises HTML » prises en compte par les outils de recherche ?</u>	45
ii. <u>Faut-il une optimisation du code particulière : le cas des frames</u>	47
iii. <u>Faut-il une optimisation du code particulière : le cas des pages dynamiques</u>	48
E. <u>QUEL EST L'INTÉRÊT DES PROFESSIONNELS DU RÉFÉRENCIEMENT ?</u>	50
i. <u>L'utilité des professionnels pour l'optimisation de l'organisation interne et du code d'un site</u>	50
ii. <u>L'utilité des professionnels pour leurs propres techniques d'optimisation</u>	51
III. <u>MISE EN PRATIQUE, COMPARAISON DES COÛTS ET MODÈLE MATHÉMATIQUE</u>	54
A. <u>MISE EN PRATIQUE DES RECOMMANDATIONS</u>	54
B. <u>ÉTUDE DU TARIF DES PROFESSIONNELS DU RÉFÉRENCIEMENT</u>	55
i. <u>Prestations de base</u>	55
ii. <u>Prestations moyennes</u>	56
iii. <u>Prestations supérieures</u>	57
iv. <u>Comparatif des offres des professionnels du référencement</u>	58
C. <u>CRÉATION D'UN MODÈLE MATHÉMATIQUE</u>	59

<i>i.</i>	<i>Formulation mathématique des différentes composantes</i>	59
1.	<u>Présence dans les outils de recherche</u>	59
2.	<u>Optimisation autour des mots-clés</u>	60
3.	<u>Optimisation de l'organisation interne</u>	62
4.	<u>Optimisation de la structure de fichiers</u>	62
5.	<u>Optimisation du code</u>	62
6.	<u>Liens externes</u>	62
7.	<u>Pages satellites et cloaking</u>	63
<i>ii.</i>	<i>Formulation du modèle mathématique théorique</i>	63
<i>iii.</i>	<i>Validation du modèle par des cas concrets</i>	64

Introduction

De plus en plus d'entreprises sont présentes sur Internet. Dans les plus grosses d'entre elles, il existe parfois des services spécialisés qui se consacrent à cette présence ; c'est plus rarement le cas dans les PME, PMI et autres TPE : elles font généralement appel à un ou plusieurs prestataires extérieurs.

Le développement d'Internet a entraîné le développement de nouveaux métiers comme les créateurs de sites, ou encore les professionnels du référencement suite à l'apparition des outils de recherche sur Internet.

S'il ne paraît pas utile de justifier l'existence et le besoin de spécialistes de la création, on peut se poser la question pour les professionnels du référencement, voire pour le référencement en général.

Que ce soit en interne ou à l'aide de prestataires extérieurs, toute entreprise désireuse d'avoir un site Internet doit créer celui-ci : la phase de création est indispensable.

Par contre, trois attitudes différentes sont constatées vis-à-vis du référencement :

- Soit l'entreprise n'y attache aucune importance et elle laisse vivre son site tel quel, voire en utilisant simplement des techniques publicitaires pour augmenter son trafic.
- Soit l'entreprise est consciente d'un rôle joué par le référencement dans les outils de recherche, et fait ou fait faire le strict minimum afin d'apparaître dans les résultats des outils de recherche.
- Soit l'entreprise attache une importance réelle au référencement de son site et y consacre une optimisation particulière, en interne ou en externe.

La multiplication du nombre d'entreprises présentes sur Internet à travers un site, l'apparition continue de nouvelles techniques de publicité en ligne accompagnée par l'apparition tout aussi continue de désillusions concernant certaines d'entre elles, ainsi que la forte médiatisation des outils de recherche depuis environ quatre ans, entraînent une prise de conscience générale du poids de ces outils et de l'importance du référencement.

Cependant, contrairement à la création de sites dont beaucoup d'internautes connaissent les rudiments, le référencement et le fonctionnement des outils de recherche sont très mal connus du public et des entreprises. Cette méconnaissance entraîne de nombreuses incohérences et permet de nombreux abus de la part des professionnels du référencement.

Il est non seulement légitime, mais surtout indispensable pour un dirigeant d'entreprise souhaitant être présent sur Internet à travers un site web de s'interroger sur la nécessité d'un travail particulier sur le référencement, mais également de s'interroger sur le contenu de ce travail et donc sur la façon de l'exécuter : en interne, en externe, par quel prestataire ?

A partir de statistiques vérifiées, de témoignages et d'observations réelles, nous examinerons tout d'abord les comportements les plus répandus des internautes, afin de répondre à la problématique de la nécessité d'un travail spécifique au référencement.

Cette nécessité démontrée, nous passerons en revue les étapes apparemment essentielles de ce travail par ordre de difficulté d'exécution croissante, celles-ci ayant été identifiées par documentations et entretiens avec différents acteurs du secteur.

Pour chacune de ces étapes, nous analyserons leur pertinence, leur coût, et les connaissances mises en œuvre à partir de résultats concrets, afin de répondre aux deux interrogations : qu'est-ce qui doit être fait ou évité ? Les professionnels du référencement sont-ils les mieux placés pour ce travail, du point de vue de l'entreprise cliente ?

Enfin, nous essaierons de bâtir un modèle mathématique économique qui permettrait, le cas échéant, de calculer la variation de chiffre d'affaires due à ces étapes.

I. Une entreprise doit-elle se préoccuper du référencement de son site ?

a. Quel est l'intérêt du référencement pour un site Internet ?

i. Quelques notions de base

Pour comprendre au mieux la problématique du sujet, il peut être nécessaire de redéfinir quelques notions de bases afin d'appréhender correctement la suite de l'étude.

Navigateur

Logiciel dont la fonction première est d'interpréter les adresses des pages web, de les afficher et d'exploiter les liens hypertextes à l'intérieur de celles-ci. Ainsi, l'internaute peut se déplacer de page en page, par un simple clic sur un lien, comme un sportif se déplace de vague en vague sur sa planche de surf. D'où l'expression surfer sur le Net. Créé en 1993, à l'université de l'Illinois, Mosaic a été le premier navigateur à parcourir les pages du web et en a favorisé l'explosion. Depuis, les fonctions de ces logiciels ont bien évolué. Les deux chefs de file, Internet Explorer de Microsoft et Netscape Navigator, permettent entre autres de gérer des bibliothèques de signets, de définir des politiques de sécurité, de se connecter à des messageries instantanées, de lire des fichiers MP3, Flash, etc.

Moteur de recherche

Site web utilisant un programme qui indexe de façon automatisée le contenu de différentes ressources Internet, et plus particulièrement de sites web, et qui permet à l'internaute qui utilise un navigateur web de rechercher de l'information selon différents paramètres, en se servant notamment de mots-clés. Google, Voila, AltaVista, Infoseek... sont des exemples de moteurs de recherche.

Annuaire

Site web se présentant comme un inventaire, spécialisé dans un domaine ou non, dans lequel les sites référencés sont classés par catégories et accessibles au moyen de liens hypertextes. Afin de localiser plus rapidement un site ou une catégorie, les répertoires donnent accès à un module limité de recherche. Yahoo est l'exemple type de l'annuaire. Les annuaires sont alimentés manuellement contrairement aux moteurs qui indexent les pages web de façon automatique.

Portail

Site qui se propose d'être une "porte d'entrée" sur le web mais qui fournit un très large éventail de services et de contenus susceptibles de retenir l'internaute. L'offre de base consiste en un moteur de recherche et/ou un annuaire de sites, mais aussi des informations venant de sources multiples le plus souvent agrégées, généralistes ou thématiques. Le portail est souvent le site par lequel un internaute entame son "surf". Les grands portails sont les oeuvres de moteurs ou d'annuaires (Yahoo, Voila...) ou bien de fournisseurs d'accès Internet (Wanadoo, Club-Internet...). Le "portail thématique" joue le même rôle sur un secteur limité : un portail médical, un portail high-tech... Le portail d'entreprise est, lui, un Intranet qui donne au personnel d'une entreprise et éventuellement à ses partenaires, accès, d'une part, à l'ensemble des données et des informations qui appartiennent à l'entreprise en question, que ces données soient structurées (ex. : bases de données internes) ou non (ex. : documents de traitement de

texte), et, d'autre part, à une série de sites web ou de portails verticaux qui se rapportent à la sphère d'activité de l'entreprise.

Référencement

Corps de techniques visant à optimiser la visibilité d'un site web dans les différents outils de recherche disponibles sur Internet : annuaires (qui indexent les sites manuellement) ou moteurs (qui les indexent automatiquement). Le référencement recouvre toute une série d'actions très différentes qui vont de la simple définition de mots-clés à la création de pages spéciales facilitant le repérage des sites par les moteurs. La fugacité et la relativité des résultats obtenus (un bon référencement n'est que provisoire) font du référencement une discipline un peu ésotérique qui laisse souvent sceptique le non-initié. En 2001, est apparue la notion de référencement payant : le "payeur" voit alors son site apparaître en tête des résultats (sur Overture.com) ou, plus subtil, dans une fenêtre en haut de page (sur Google). Le référencement devient alors de la publicité et le référenceur une centrale d'achat d'espace.

ii. Quelques statistiques et tendances intéressantes

De nombreux chiffres et statistiques sur Internet et sur les internautes sont publiés tous les jours, mais ils ne sont que très rarement réunis, comparés, et surtout analysés les uns par rapport aux autres.

Si ces statistiques varient toujours de quelques points en fonction de l'organisme dont elles sont issues, les tendances qui en ressortent restent toutefois identiques.

Le but de ce paragraphe n'est en effet pas d'obtenir des données chiffrées précises, mais bien d'analyser les statistiques existantes pour en faire ressortir les tendances les plus intéressantes. De la même façon, les différences des périmètres d'étude des statistiques et des modes de calcul et définitions des différentes désignations utilisées par ces organismes peuvent engendrer des différences dans les résultats chiffrés, mais pas dans les ordres de grandeur de ces résultats.

J'ai donc choisi d'ignorer ces différences plutôt que de les expliquer et de les retraiter une à une, afin d'alléger l'analyse des tendances qui se dégagent de ces ordres de grandeur.

Malgré tous les progrès des organismes de statistiques spécialisés dans l'étude d'Internet, il est encore difficile aujourd'hui de connaître précisément la taille de celui-ci. En effet, l'Internet peut tout d'abord être divisé en deux sous parties : l'Internet, ou net, « visible », et le net « invisible ».

Le premier est le net que l'on côtoie le plus souvent, celui sur lequel on navigue de liens en liens. C'est le plus connu et le plus facile à évaluer en taille.

Le second, le net « invisible », est composé de tout ce qui n'est pas accessible par des liens, c'est-à-dire qu'on ne peut visualiser qu'en connaissant auparavant l'adresse exacte de la page en question (ou url, du type « `http://www.domaine.com/... /...index.html` » par exemple). Cette partie du web est composée de pages personnelles d'utilisateurs qui ont créé celles-ci pour eux et pour un nombre restreints de connaissances, ou encore des différentes archives ou vieux documents qui ne sont plus accessibles à partir de pages remises à jour, ainsi que l'ensemble des pages privées protégées par un mot de passe (intranet, extranet), etc. La plupart des organismes d'étude de l'Internet s'accordent pour dire que cette partie du web

est de loin la plus importante en taille, puisqu'elle serait même environ de 400 à 550 fois plus grande que le web « visible », mais il n'en existe à l'heure actuelle aucune mesure réelle.

Intéressons nous simplement au web visible. Là encore, on peut distinguer plusieurs catégories de sites web.

Tout d'abord, on trouve les sites « actifs », ceux sur lesquels on « navigue » la plupart du temps.

Puis il y a également les sites « en construction » qui n'ont par définition pas encore de contenu exploitable pour l'internaute. Il y aurait au niveau mondial environ autant de sites « en construction » que de sites « actifs ».

Enfin, il y a tous les sites « non pertinents » qui sont les sites qui ne sont que de simples copies de sites actifs, ou encore des sites ne pouvant être visualisés pour cause d'absence de pages ou d'erreurs.

Là encore nous nous contenterons d'étudier une petite partie de l'Internet, puisque nous n'étudierons que les sites « actifs » du web « visible », et plus particulièrement les sites « actifs » du web « visible » français !

Des études récentes de plusieurs organismes montrent qu'en janvier 2004, on peut évaluer à environ 2,4 millions le nombre de ces sites en France :

Le nombre de sites actifs en Europe par pays, hors sites dupliqués (source Ripe)	
Pays	Janvier 2004
Pays-Bas	3 521 932
Grande-Bretagne	3 169 318
Allemagne	2 603 007
France	2 403 459
Finlande	1 271 634
Danemark	1 248 296
Suède	943 139
Espagne	910 677
Italie	626 536
Suisse	548 044

A titre indicatif, regardons également le nombre de sites actifs du web visible mondial et son évolution :

Monde : nombre de sites actifs, hors sites dupliqués (source Netcraft)	
Période	Nombre de sites
Juillet 2003	42 298 371
Septembre 2002	35 756 436
Septembre 2001	32 398 046
Septembre 2000	21 166 912
Septembre 1999	7 370 929
Septembre 1998	3 156 324
Septembre 1997	1 364 714
Septembre 1996	397 281

Nous remarquons très facilement au passage l'explosion du nombre de sites en 7 ans, même si cette croissance semble avoir ralenti en 2001 – 2002 et l'éclatement de la « bulle » Internet.

Ce qui est important à retenir, c'est que nous nous intéressons pour la suite de cette étude chiffrée à une très petite partie du web : les quelques 2,4 millions de sites web français actifs.

Regardons maintenant l'évolution du nombre d'internautes au monde ... :

Internaute dans le monde par continent, en millions (sources eMarketer et International Telecom Union)					
Zone	2000	2001	2002	2003	2004 (e)
Amérique du Nord	136,7	156,3	167,7	179,8	196,3
Amérique latine	19,3	26,2	33,1	43,4	60,6
Afrique	4,6	6,7	7,7	9,2	11,1
Asie/Pacifique	115,9	165,0	181,5	205,0	235,8
Europe	108,3	144,4	175,7	196,2	221,1
Total Monde	384	498,7	565,7	633,6	724,9

... puis plus particulièrement en France :

Internaute français, en millions (source Nielsen // NetRatings)	
Date	Nombre (en millions)
Avril 2003	13,8
Février 2002	13,5
Octobre 2001	11,2
juillet 2001	11,0

Là encore, nous ne nous intéressons qu'à une petite partie des internautes : les 13,8 millions d'internautes français, afin de les mettre en relation avec les 2,4 millions de sites web français vus auparavant.

Un premier calcul rapide nous donne un ratio moyen de 5,75 internautes français par site web français. Mais ce résultat n'est à prendre en compte que si l'on considère qu'un internaute français ne visite qu'un seul site web et que celui-ci est un site français, ce qui est bien évidemment faux dans la plupart des cas.

Il est essentiel de remarquer à ce niveau que la valeur exacte de ce ratio dépend de nombreux paramètres comme par exemple de la façon de comptabiliser le nombre d'internautes français, mais l'ordre de grandeur de celui-ci reste inchangé, et nous ne rentrerons donc pas dans ce niveau de détail.

Afin de préciser ce résultat, il est donc nécessaire d'étudier plus en détail les habitudes des internautes français en matière de trafic web. Ceci permettra d'obtenir une moyenne du nombre de visiteurs français par site Internet français.

iii. *Extrapolation du nombre moyen de visiteurs par site web*

Étudions l'utilisation moyenne mensuelle d'Internet par un internaute français :

Utilisation moyenne mensuelle par internaute (source Nielsen // NetRatings)		
	Septembre 2003	Octobre 2001
Nombre de sessions	28	14
Nombre de domaines visités	65	16
Nombre de pages vues	1 016	571
Nombre de pages vues par session	37	41
Temps passé par mois	13:50:04	06:13:13
Temps passé par session	00:29:37	00:26:36

Dans le cas présent, une « session » représente le laps de temps durant lequel l'internaute parcourt le web entre l'ouverture d'un navigateur Internet et la fermeture de celui-ci.

Un « domaine » représente un ensemble de pages correspondant à une même adresse Internet de départ (du style [http://www.domaine.com/...](http://www.domaine.com/)). Nous pouvons considérer en première approximation qu'un domaine est l'équivalent d'un site Internet et de l'ensemble de ses pages.

En considérant donc qu'un internaute français visite en moyenne 65 sites web différents, et en considérant que ces 65 sites web font partie des 2,4 millions de sites web actifs français, on obtient une moyenne mensuelle de 374 visiteurs français par site.

Cependant, un internaute français ne visitant pas que des sites français, nous pouvons conclure que ce résultat de 374 visiteurs français mensuels par site est une borne supérieure à la moyenne réelle.

Nous pouvons également effectuer ce calcul en considérant l'ensemble des sites web actifs mondiaux comparés au nombre d'internautes français et à leurs habitudes et nous obtenons environ 21 visiteurs français par site et par mois. Ce résultat constitue une borne inférieure à la moyenne réelle.

La moyenne du nombre de visiteurs internautes français par site web actif français est donc comprise entre 21 et 374.

Cette fourchette, bien qu'assez imprécise, est à comparer avec l'audience mensuelle des principaux sites français :

Les 10 premières audiences mensuelles en France (source Nielsen // NetRatings)	
Domaines	Audience cumulée (visiteurs uniques en milliers)
wanadoo.fr	7 195
free.fr	5 800
microsoft.com	4 918
lycos.fr	4 648
google.fr	4 419
voila.fr	4 210
yahoo.fr	3 739
tiscali.fr	3 010
msn.com	2 955
pagesjaunes.fr	2 771

Ainsi observons-nous que même la borne supérieure de 800 visiteurs mensuels est extrêmement faible par rapport à l'audience des sites mentionnés ci-dessus. Si ces sites accaparent autant de visiteurs et que la moyenne est inférieure à 800, c'est qu'il reste encore moins de visiteurs potentiels pour les autres sites !

Il faut rappeler que les bornes de la fourchette que nous venons de calculer ne sont pas à prendre en tant que données précises, mais bien en tant que tendances. Ce ne sont pas les valeurs trouvées qui importent mais leur ordre de grandeur : moins de 1000 visiteurs par mois de moyenne contre plusieurs millions pour les principaux sites.

Ce qu'il faut retenir de cette fourchette, c'est que si l'on crée un site afin de le rendre visible au plus grand nombre, il n'est pas réaliste de le laisser vivre sans mesure particulière afin d'augmenter son trafic.

Et c'est en étudiant la façon dont les internautes se rendent sur un site Internet que nous en déduisons l'importance du référencement pour obtenir du trafic.

iv. *Étude du type d'accès aux sites Internet français : l'importance des outils de recherche*

Pour un site web donné, il faut distinguer 3 types d'accès différents à ce site :

- accès direct : l'internaute accède au site web en rentrant directement l'url de ce site, ou en utilisant les « favoris » de son navigateur.
- liens externes : l'internaute accède au site web en suivant un lien situé sur un autre site, en cliquant sur une publicité, sur un lien présent dans un mail, ou encore en cliquant sur un lien promotionnel présent dans la page des résultats d'un moteur de recherche.
- moteurs de recherche (ou outils de recherche) : l'internaute utilise un moteur de recherche ou un annuaire et clique sur un des liens non promotionnels qui lui sont proposés.

Répartition du trafic par type d'accès (sources @position - eStat)					
Type d'accès	Février 02	Juin 02	Octobre 02	Février 03	Janvier 04
Liens externes	54,64 %	51,07 %	46,58 %	36,40 %	35,21 %
Accès direct	25,20 %	27,92 %	28,37 %	36,26 %	29,52 %
Moteurs de recherche	20,16 %	21,01 %	25,05 %	27,34 %	35,27 %

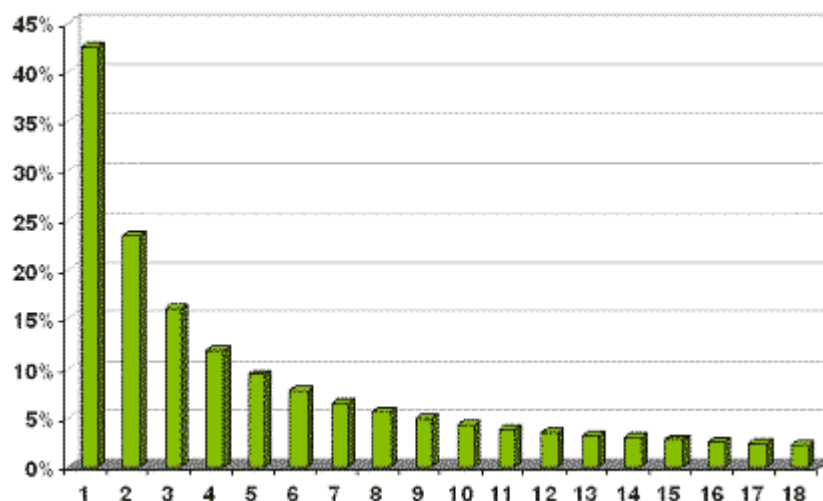
Ce tableau montre la forte diminution sur deux ans de l'accès par liens externes au profit de l'accès direct et de l'accès par un moteur de recherche : de moins en moins d'internautes accèdent à un site en cliquant sur une publicité, un lien promotionnel ou même en suivant un simple lien hypertexte d'un autre site.

Par contre, ils sont de plus en plus nombreux à connaître l'adresse du site auquel ils veulent accéder ou à se servir des « favoris » de leur navigateur.

De même, ils sont de plus en plus nombreux à utiliser un outil de recherche pour accéder au site voulu. Les outils de recherche sont la principale source de trafic pour un site Internet depuis le mois de décembre 2003.

Étudions à présent le taux de retour sur un même site pendant un mois donné :

Les taux de retour sur un même site (source @position)				
	1(2)	2(3)	5(6)	19(20)
Taux de retour	42,60 %	23,40 %	9,38 %	2,29 %

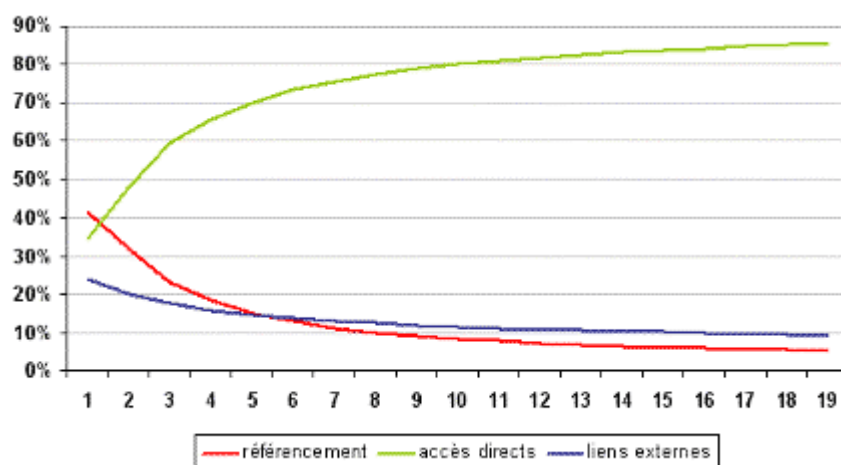


2(3) : 23,40 % des internautes venus deux fois sur un même site reviennent une troisième fois dans le mois.

Il est important de remarquer ici que moins d'un internaute sur deux revient sur le même site dans le même mois, et ils ne sont plus que 15 % à revenir une troisième fois. Les taux de fidélisation sont bas sur un mois, et l'acquisition régulière de nouveaux visiteurs pour un site est primordiale.

Plus encore que la répartition du trafic d'un site par type d'accès, c'est donc ce type d'accès en fonction du nombre de visites sur un même site qu'il faut étudier

Provenance en fonction du nombre de visites sur un même site (source @position)					
Type d'accès	1	2	5	10	19
Liens externes	23,93 %	19,96 %	14,80 %	11,64 %	9,29 %
Accès direct	34,59 %	48,25 %	70,16 %	80,02 %	85,40 %
Moteurs de recherche	41,47 %	31,79 %	15,04 %	8,34 %	5,30 %



Le tableau ci-dessus illustre l'importance du référencement pour l'acquisition de nouveaux visiteurs : pour leur première visite, les internautes proviennent en majorité (à 41 %) des moteurs de recherche.

Les courbes "référencement" et "accès directs" s'inversent de façon tout à fait normale. Néanmoins, seuls 60 % des internautes reviennent sur un site pour la troisième fois en direct et ils sont encore plus de 20 % à toujours utiliser un outil de recherche.

De la même façon, ils sont encore 15 % à revisiter le même site pour la 5^{ème} fois via un outil de recherche. La fidélisation par accès direct ne se marque de façon affirmée qu'à partir de la 10^{ème} visite sur un même site (80 %).

Enfin, les accès par liens externes (intégrant la publicité et les liens sponsorisés) ne représentent qu'une faible part de l'acquisition et de la fidélisation des internautes. Dès la seconde visite, moins d'un internaute sur cinq provient de liens externes alors qu'ils sont presque encore 1/3 à provenir d'un outil de recherche. Le trafic issu des outils de recherches est la principale source de trafic des sites Internet.

Nous avons déjà vu précédemment qu'il était essentiel de mettre en place une stratégie afin d'augmenter son trafic, nous venons maintenant de montrer l'importance de la part des moteurs de recherche dans cette stratégie : afin de toucher le plus grand nombre d'internautes, il est essentiel non seulement d'avoir des liens externes pointant vers son site, mais surtout d'être présent parmi les résultats des moteurs de recherche et des annuaires.

Le référencement est, pour un site Internet, le principal moyen d'obtenir des visiteurs, et notamment des visiteurs qui ne connaissent pas l'adresse Internet du site en question.

Le référencement est donc indispensable, et est d'autant plus important que la notoriété d'un site est faible.

b. Quel est l'intérêt d'une optimisation du référencement d'un site Internet ?

Nous venons de voir pourquoi il est nécessaire pour un site d'être présent sur les moteurs de recherche et annuaires. Il faut maintenant regarder les habitudes d'utilisation de ces outils par les internautes, afin de savoir si le simple fait d'être présent sur les moteurs et annuaires est suffisant.

i. Nombre de pages web visitées par un internaute au cours d'une recherche

Après avoir effectué une requête auprès d'un moteur de recherche ou d'un annuaire, ces derniers présentent à l'internaute une série de résultats, sous la forme d'une liste de liens vers différentes pages web. Ces liens sont généralement accompagnés d'une définition succincte de la page et apparaissent la plupart du temps par 10.

Chaque internaute utilisant un moteur de recherche ou un annuaire est identifié par un code unique qui permet par la suite d'analyser son comportement sur cet outil de recherche.

Ce code est bien sûr anonyme et ne permet en aucun cas de retrouver l'utilisateur en question en vertu du respect de la vie privée.

Ce système permet d'obtenir des résultats très intéressants comme notamment le nombre de pages web et donc de sites visités par l'internaute après présentation par le moteur de recherche ou l'annuaire des résultats de sa requête. Il n'existe que très peu de données sur le comportement des internautes français sur les outils de recherche, mais le peu de données disponibles montre qu'il est en tout point comparable à celui des internautes du monde entier.

Il est donc plus utile de se baser sur des données chiffrées internationales plus complètes.

Nombre de pages visitées par recherche (4th International Conference on Internet Computing - 2003)		
Nombre de pages visitées	Pourcentage des internautes	Cumul
1	54,3 %	54,3 %
2	18,9 %	73,2 %
3	9,4 %	82,6 %
4	5,5 %	88,1 %
5	3,3 %	91,4 %
6	2,2 %	93,6 %
7	1,5 %	95,1 %
8	1,0 %	96,1 %
9	0,8 %	96,9 %
10	0,6 %	97,5 %
>10	2,5 %	100 %

Ces premiers résultats montrent que plus d'un internaute sur deux ne visite qu'un site sur l'ensemble des sites que lui propose son outil de recherche, c'est-à-dire en d'autres termes, qu'il ne clique que sur un seul des liens dans toute la liste qui lui est présentée !

Ces résultats sont cependant à modérer puisque nous avons considéré ici le nombre de pages visitées « par recherche ». Or, très souvent, les internautes effectuent plusieurs tentatives de recherche afin d'obtenir celle qui se rapproche le plus de son idée première.

Nombre de recherches par session (4th International Conference on Internet Computing - 2003)	
Nombre de recherches	Pourcentage des internautes
1	53 %
2	18 %
> 2	29 %

En effet, seuls 53 % des internautes effectuent une seule et unique recherche, et environ 29% effectuent au moins 3 requêtes pour obtenir le résultat voulu. La moyenne est en fait de 2,94 recherches différentes par session.

Le nombre de pages vues par recherche n'est donc pertinent que pour les 53 % d'internautes qui n'effectuent qu'une seule recherche.

C'est donc aux sessions qu'il faut s'intéresser, c'est-à-dire à toutes les différentes recherches effectuées par un internaute au cours d'un processus de recherche.

Nombre de pages visitées par session (4 th International Conference on Internet Computing - 2003)		
Nombre de pages visitées	Pourcentage des internautes	Cumul
1	27,62 %	27,62 %
2	14,95 %	42,57 %
3	10,23 %	52,80 %
4	7,65 %	60,45 %
5	5,87 %	66,32 %
6	4,65 %	70,97 %
7	3,73 %	74,70 %
8	2,97 %	77,67 %
9	2,51 %	80,18 %
10	2,15 %	82,33 %
>10	17,67 %	100 %

Ces résultats rejoignent les premiers résultats obtenus avec le nombre de pages visitées par recherche : au cours d'une session entière, c'est-à-dire de l'ensemble des recherches effectuées par un internaute pour obtenir l'information qu'il souhaite, celui-ci ne visite finalement que moins de 3 pages web différentes dans près de 53 % des cas, et moins de 5 dans plus de 66 % des cas !

Il est là encore nécessaire de rappeler que ces chiffres peuvent légèrement différer selon leur source, mais les tendances qui en ressortent restent les mêmes.

Il apparaît clairement qu'être présent dans les résultats d'un moteur de recherche ou d'un annuaire n'est pas synonyme de trafic supplémentaire.

Si l'internaute qui effectue une recherche ne visite ensuite que très peu de sites web, il est indispensable de savoir quels sont les sites qu'il choisit de visiter.

ii. Nombre de pages de résultats visualisées par un internaute au cours d'une recherche

Nous avons déjà vu que la plupart des moteurs de recherches présentent leur résultat sous forme d'une liste de liens.

Le nombre de liens répondant à une requête pouvant atteindre plusieurs millions, les moteurs de recherche et annuaires présentent ceux-ci par pages de 10 liens en général.

L'internaute peut donc, s'il le souhaite, faire défiler les différentes pages de résultats pour obtenir plus de résultats.

L'ordre d'affichage est quant à lui prédéfini par le moteur de recherche en fonction d'algorithmes et de choix qui lui sont propres. Cependant, nous pouvons dire que le but de ces

moteurs de recherches et annuaires est de présenter en premier les sites qui leur paraissent les plus pertinents par rapport à la requête de l'utilisateur.

Le détails de ces algorithmes et surtout de celui de Google seront étudiés plus loin.

Nombre de pages de résultats vues (4 th International Conference on Internet Computing - 2003)		
Nombre de pages de résultats	Pourcentage des internautes	Cumul
1	54,1 %	54,1 %
2	19,3 %	73,4 %
3	9,6 %	83,0 %
4	5,5 %	88,5 %
5	3,3 %	91,8 %
6	2,1 %	93,9 %
7	1,5 %	95,4 %
8	1,0 %	96,4 %
9	0,7 %	97,1 %
10	0,5 %	97,6 %
>10	2,4 %	100 %

Plus de la moitié des internautes utilisant un moteur de recherche ne regardent que la première page de résultats, c'est-à-dire les 10 premiers liens proposés par les moteurs.

De même, plus de 73 % des internautes ne regardent que les 2 premières pages de résultats et seuls 26,6 % des internautes vont au-delà des 20 premiers sites proposés.

Regardons à titre indicatif l'évolution de ce phénomène depuis 1997 :

Nombre de pages de résultats vues (4 th International Conference on Internet Computing - 2003)				
Nombre de pages de résultats	1997	1999	2001	2002
1	28,6 %	42,7 %	50,5 %	54,1 %
2	19,5 %	21,2 %	20,3 %	19,3 %
> 2	51,9 %	36,1 %	29,2 %	26,6 %

De plus en plus d'internautes se contentent de la première page de résultats et parallèlement, de moins en moins d'internautes vont au-delà des 2 premières pages de résultats. Une explication possible peut être que les moteurs de recherche fournissent des résultats qui correspondent de mieux en mieux aux attentes des internautes.

Il est donc indispensable pour un site souhaitant générer du trafic non seulement d'être présent dans les résultats des moteurs de recherche et annuaires, mais surtout d'être présent dans les toutes premières pages de résultats présentées à l'internaute.

Plus que le référencement, c'est l'optimisation de ce référencement qui est importante.

iii. Études de quelques durées significatives

Afin de comprendre au mieux le comportement de l'internaute lorsqu'il effectue une recherche, attardons-nous un moment sur la notion de durée : en effet, de même qu'il est intéressant de savoir combien de pages web sont vues parmi l'ensemble de celles proposées par l'outil utilisé pour la recherche, il peut-être également pertinent de connaître par exemple la durée moyenne d'une session de recherche.

Durée d'une session (4 th International Conference on Internet Computing – 2003)		
Durée d'une session	Pourcentage des internautes	Cumul
< 5 minutes	26,2 %	26,2 %
de 5 à 10 minutes	6,2 %	32,4 %
de 10 à 15 minutes	19,7 %	52,1 %
de 15 à 30 minutes	9,1 %	61,2 %
de 30 à 60 minutes	14,5 %	75,7 %
de 1 à 2 heures	4,1 %	79,8 %
de 2 à 3 heures	10,3 %	90,1 %
de 3 à 4 heures	1,2 %	91,3 %
> 4 heures	8,7 %	100 %

S'il ne semble pas ressortir de tendance générale du tableau précédent, il faut quand même constater que plus d'un quart des internautes consacrent moins de 5 minutes à leur recherche, et plus de la moitié moins d'un quart d'heure.

Regardons maintenant le temps passé par l'internaute sur les pages web issues des résultats du moteur de recherche ou de l'annuaire utilisé :

Durée d'une visite d'un site issu des résultats (4 th International Conference on Internet Computing – 2003)		
Durée d'une visite	Pourcentage des internautes	Cumul
< 30 secondes	14,6 %	14,6 %
de 30 à 60 secondes	15,4 %	30,0 %
de 1 à 2 minutes	5,1 %	35,1 %
de 2 à 3 minutes	15,2 %	50,3 %
de 3 à 4 minutes	4,8 %	55,1 %
de 5 à 10 minutes	14,9 %	70,0 %
de 10 à 15 minutes	4,4 %	74,4 %
de 15 à 30 minutes	13,0 %	87,4 %
de 30 à 60 minutes	2,9 %	90,3 %
> 60 minutes	9,7 %	100 %

Nous constatons grâce à ce tableau que plus de la moitié des internautes passent moins de 3 minutes à visiter un site web issu des liens présentés en résultat par le moteur de recherche, avant de revenir sur celui-ci.

De même, près de 15 % des internautes y passent moins de 30 secondes ! Autrement dit, l'internaute se fait une idée du site très rapidement : Il juge en moins de 3 minutes si le site web correspond à ce qu'il cherche ou non.

C'est la première impression du site web qui est donc la plus importante. Un internaute ne s'attardera pas à chercher sur le site l'information qu'il cherche si elle ne lui apparaît pas directement accessible.

De plus, à cela s'ajoute une autre statistique : dans environ 80 % des cas, c'est le petit résumé qui accompagne le lien vers le site web dans la page de résultats du moteur utilisé, qui dissuade l'internaute de visiter ce site.

Plus encore que la première impression lors de l'arrivée sur un site, c'est la description même de celui-ci dans le moteur de recherche qui est décisive.

Enfin, il faut savoir que des études récentes ont montré que les résultats des moteurs de recherche sont pertinents dans environ 52 % des cas et non pertinents dans 48 % des cas. Ce niveau de pertinence montre la relative efficacité des moteurs de recherche et annuaires actuels. De plus, cela confirme les résultats précédents : en effet, si un internaute a une chance sur deux de tomber sur un site pertinent parmi les réponses du moteur de recherche, il semble normal que près de 43 % des utilisateurs visitent un maximum de 2 sites parmi ceux proposés, et que plus de 52 % des internautes consacrent moins d'un quart d'heure à leur recherche.

Il est donc nécessaire d'être « bien » positionné sur les outils de recherche, c'est-à-dire dans les premiers résultats de ceux-ci, et avec une description efficace.

Nous pouvons conclure qu'il est indispensable pour un site d'être référencé, mais surtout d'optimiser son référencement. Cette optimisation doit aboutir à un positionnement dans les 2 premières pages de résultats des moteurs de recherche ou annuaires lorsqu'un internaute effectue une requête concernant le ou les thème(s) du site référencé.

Mais cette optimisation doit également aboutir à un résumé pertinent qui sera affiché en parallèle du lien vers le site sur les pages de résultats des moteurs, et enfin à un contenu facilement et rapidement accessible depuis la page référencée.

Afin d'en déduire au mieux la manière d'optimiser le référencement d'un site Internet, regardons d'abord l'état actuel de celui-ci et de ses acteurs.

c. Comment fonctionne le référencement ?

i. Les principaux moteurs de recherche et annuaires

La première chose qu'il faut constater, est que les internautes français préfèrent largement l'utilisation des moteurs de recherches à celle des annuaires :

Les outils utilisés par les internautes français (source @position – février 2003)	
Outils	Pourcentage des internautes
moteur de recherche	73 %
annuaire de recherche	27 %

Nous nous concentrerons donc naturellement sur l'étude des moteurs de recherches, mais sans pour autant laisser de côté les annuaires.

Nous savons que le trafic issu des moteurs et annuaires est indispensable pour les sites Internet, regardons maintenant la répartition de ce trafic entre les principaux outils de recherche :

Les moteurs et annuaires les plus utilisés dans le monde (source OneStat)					
Rang	Moteur / Annuaire	Juillet 02	Septembre 02	Juin 03	Novembre 03
1.	Google	53,2 %	55,1 %	55,2 %	56,1 %
2.	Yahoo	20,4 %	20,6 %	21,7 %	21,5 %
3.	MSN Search	9,1 %	9,4 %	9,6 %	9,4 %
4.	AOL Search	2,9 %	3,5 %	3,8 %	3,7 %
5.	Terra Lycos	3,7 %	3,0 %	2,6 %	2,3 %
6.	Altavista	2,8 %	2,4 %	2,2 %	1,9 %
7.	AskJeeves	2,2 %	1,7 %	1,5 %	1,6 %
	autres...	5,7 %	4,3 %	3,4 %	3,5 %

Deux tendances se dégagent de ce tableau :

- d'abord la prédominance de Google qui a lui seul réalise plus de la moitié du trafic des moteurs de recherches,
- ensuite, l'augmentation des parts des 4 principaux moteurs et annuaires au détriment de tous les autres : Google, Yahoo, MSN et AOL représentent 90,7 % du trafic généré par les outils de recherche en juin 2003 contre 85,6 % en juillet 2002.

Dans le cas de la France, les résultats sont légèrement différents :

Les moteurs et annuaires les plus utilisés en France (source eStat)					
Moteur / Annuaire	Février 02	Juin 02	Octobre 02	Février 03	Janvier 2004
google.fr	29,39 %	36,59 %	39,35 %	40,74 %	46,64 %
wanadoo.fr	13,78 %	11,08 %	8,82 %	8,65 %	6,76 %
google.com	12,66 %	13,90 %	14,80 %	15,32 %	18,86 %
yahoo.fr	10,48 %	9,47 %	9,35 %	8,84 %	7,17 %
voila.fr	7,52 %	5,86 %	4,52 %	3,32 %	1,82 %
lycos.fr	5,01 %	3,81%	2,86 %	2,18 %	< 1 %
yahoo.com	4,64 %	4,73 %	4,74 %	4,48 %	3,22 %
msn.fr	3,87 %	3,93 %	5,33 %	5,25 %	4,75 %
aol.fr	2,61 %	1,91 %	3,73 %	5,41 %	4,51 %
nomade.fr	2,25 %	2,51 %	1,80 %	1,37 %	< 1 %
autres...	7,77 %	6,20 %	4,71 %	4,45 %	> 4,27 %

Avant d'analyser ces résultats sous cette forme, nous remarquons que les moteurs google.fr et google.com sont les mêmes, de même que yahoo.fr et yahoo.com. Enfin, le moteur de recherche disponible sur le site de wanadoo.fr est en fait celui de voila.fr.

Nous pouvons donc regrouper ces 3 duos avant d'exploiter ces données.

Les moteurs et annuaires les plus utilisés en France (source eStat)					
Moteur / Annuaire	Février 02	Juin 02	Octobre 02	Février 03	Janvier 04
Google	42,05 %	50,49 %	54,15 %	56,05 %	65,50 %
yahoo	15,12 %	14,20 %	14,09 %	13,32 %	10,39 %
wanadoo / voila	21,30 %	16,94 %	13,34 %	11,97 %	8,58 %
aol.fr	2,61 %	1,91 %	3,73 %	5,41 %	4,51 %
msn.fr	3,87 %	3,93 %	5,33 %	5,25 %	4,75 %
lycos.fr	5,01 %	3,81%	2,86 %	2,18 %	< 1 %
nomade.fr	2,25 %	2,51 %	1,80 %	1,37 %	< 1 %
autres...	7,77 %	6,20 %	4,71 %	4,45 %	> 4,27 %

Les résultats français sont encore plus marqués par la très nette domination de Google, ainsi que par le déclin des autres moteurs et annuaires hormis AOL France et MSN France. Il faut également remarquer que les principaux acteurs français sont les mêmes qu'à l'échelle internationale sauf en ce qui concerne le couple Wanadoo / Voila qui représente en France environ 8,6 % du trafic moteur en janvier 2004.

Ces tendances nous permettent d'avoir une vue d'ensemble du paysage des outils de recherche.

Nous allons maintenant détailler les caractéristiques principales de ces 7 moteurs et annuaires les plus importants en France.

ii. Les caractéristiques principales de ces moteurs et annuaires

Beaucoup de rumeurs et de fausses informations circulent sur Internet à propos des outils de recherche et surtout de leurs fonctionnements respectifs. Aussi avons-nous choisi de ne reporter que des éléments vérifiés, issus pour la plupart des sites-mêmes de ces outils.

1. Google

Google est un moteur de recherche lancé fin septembre 1999, par Sergey Brin et Larry Page. Google est aujourd'hui accessible dans de très nombreuses langues avec, pour un très grand nombre de pays, un site dédié : google.fr pour la France.

Google propose également aux internautes un annuaire qui par contre ne lui appartient pas : l'annuaire Open Directory ou DMOZ.

Aujourd'hui on estime à plus de 4 milliards le nombre de documents web indexés par ce moteur qui a la particularité d'indexer également les groupes de discussions, les images, les documents de type Microsoft Word, Excel, Powerpoint, ou encore Adobe PDF. La base de données de Google est à ce jour la plus importante disponible.

La principale caractéristique de ce moteur réside dans sa façon de classer les résultats : il se base sur un algorithme mathématique qui évalue l'importance de chaque page web enregistrée sur la base d'un contrôle de plus de 500 millions de variables et de 2 milliards de termes. Cette technologie, appelée PageRank contrôle non seulement le contenu de la page web, mais vérifie également d'autres sites éventuels qui auraient un lien vers cette page. L'algorithme exact ainsi que l'ensemble des paramètres sont bien entendu tenus secrets par la société Google.

Comme la plupart des moteurs de recherche, Google parcourt le web et enregistre les pages grâce à un robot : le « Googlebot ». Ce petit programme parcourt les pages, les analyse et suit chacun des liens contenus dans celles-ci. La plupart des robots utilisés par les moteurs se comportent en quelque sorte comme un internaute qui utiliserait un navigateur web en mode texte, c'est-à-dire dans le mode le plus simple possible.

Google ne fait pas payer les sites pour qu'ils apparaissent dans sa base de données. Cependant, le moteur de recherche a mis en place un système de liens promotionnels : en échange d'une certaine rémunération, Google place l'adresse du site web client dans un encart spécifique juste au-dessus ou à droite de la liste normale des résultats.

2. Yahoo

Yahoo est à la base un annuaire créé à la fin de l'année 1993 par Jerry Yang et David Filo. Yahoo est devenu un véritable portail, et est non seulement l'annuaire le plus visité au monde, mais c'est également le site Internet le plus visité au monde. Par contre, il ne représente qu'environ un cinquième du trafic généré par les outils de recherche contre plus de la moitié pour Google.

On estime aujourd'hui à près de 2 millions le nombre de pages indexées par l'annuaire de Yahoo. Ces pages sont depuis le début ajoutées entièrement manuellement. Yahoo propose

tout comme Google un système de liens promotionnels mais passe pour cela par la société Overture, principale concurrente de Google dans ce domaine (avec la société Espotting). Il faut d'ailleurs noter que Yahoo vient de finaliser le rachat d'Overture.

Yahoo propose pour son annuaire un système de référencement Yahoo Express : pour 250 euros hors taxe, la société vous garanti que votre site sera analysé dans les 7 jours. Dans le cas d'un référencement classique gratuit, il n'y a aucune garantie de délai. Cependant, Yahoo Express est obligatoire pour les sites commerciaux.

Yahoo propose également sur sa page d'accueil un système de moteur de recherche qui a longtemps été fondé sur celui de Google : Yahoo utilisait la base de données et le classement de Google pour ensuite le modifier et y ajouter des références issues de son propre annuaire. Cependant, depuis mars 2004, Yahoo a abandonné la technologie moteur de Google au profit d'une technologie propre : la Yahoo Search Technology (YST). Dans le même temps, Yahoo propose pour ce nouveau moteur la possibilité de souscrire à une offre de référencement payant qui garantit le délai d'indexation des pages concernées. Cette offre est basée sur une double facturation : à l'entrée en fonction du nombre d'url (ou de pages) concernées par l'offre, puis à l'utilisation avec un système de coût par clic (CPC). Un webmaster voulant souscrire à cette offre doit donc payer lors de la soumission de son site, puis à chaque fois qu'un internaute cliquera vers ce site depuis les résultats de Yahoo.

3. Wanadoo / Voila

Le couple Wanadoo / Voila appartient à France Telecom et propose à la fois un système de moteur de recherche et d'annuaire. Dans les 2 cas, la technologie est une technologie propriétaire : le moteur « Voila » et l'annuaire « Le Guide Voila ».

Voila utilise son propre robot pour parcourir le web et son propre système de classement des résultats. On estime à 120 millions le nombre de documents indexés.

De même que Yahoo, Le Guide Voila propose également un système de liens payants, celui d'Overture, et un système de référencement gratuit pour les sites non commerciaux. L'offre proposée aux sites commerciaux est l'offre Soumission Express. Elle est facturée 199 euros hors taxe et garantie un délai d'évaluation du site de 7 jours.

Pour le moment, le moteur de recherche de Voila référence les sites gratuitement, mais un système de référencement payant est à l'étude.

4. AOL France

Alors que le portail AOL Time Warner utilise Google comme moteur de recherche et DMOZ comme annuaire, le portail AOL France utilise le moteur de recherche français Exalead et un annuaire propriétaire.

La base de données de sites web disponibles à travers AOL France est estimée à près de 5 millions de pages francophones.

Overture est là encore utilisé pour faire apparaître des liens promotionnels au dessus des résultats du moteur.

La soumission de sites commerciaux ou non est gratuite pour le moteur de recherche tout comme pour l'annuaire.

5. MSN France

Le portail MSN France utilise le moteur de recherche de la société Inktomi, rachetée en 2002 par Yahoo. Il utilise également un annuaire qui ne lui appartient pas : « La Mine » alors que le portail principal MSN Search utilise celui de la société australienne LookSmart.

Le moteur de recherche Inktomi référence environ 500 millions de pages web et l'annuaire La Mine environ 1 million.

Depuis mars 2002, MSN France intègre aux résultats de son moteur de recherche et de son annuaire 3 liens promotionnels fournis par Overture.

La soumission de sites commerciaux ou non est gratuite pour l'annuaire, par contre Inktomi propose une offre de soumission payante pour le moteur de recherche en garantissant un délai de 3 jours pour 25 dollars par an et par page dans la limite de 1000 pages.

6. Lycos France

Lycos France est la version française du moteur de recherche Lycos créé en 1994. Il a longtemps utilisé sa propre technologie mais utilise depuis mi-2000 le moteur de recherche « Fast - All The web » qui référence quelques 2,1 milliards de pages (la version propriétaire de Lycos en référençait 625 millions). Fast utilise son propre robot pour parcourir les pages web et son propre système de classement repris par Lycos.

Lycos France propose également un annuaire basé essentiellement sur l'annuaire DMOZ mais également sur ses archives propres. Enfin, Lycos France a mis en place un système de liens promotionnels en couplant sa technologie propriétaire avec celle d'Overture et celle d'Espotting.

Lycos référence gratuitement les sites non commerciaux dans son annuaire, mais facture 125 euros l'offre Soumission Express obligatoire pour les sites commerciaux. Dans ce cas, le délai d'évaluation du site est garanti à moins de 5 jours. De même, par le biais de son moteur Fast, Lycos propose un référencement payant pour 15 dollars par an et par page en garantissant une indexation du site dans le moteur de recherche sous 24 heures.

7. Nomade

Tout comme Lycos, Nomade est un moteur de recherche utilisant la technologie de Fast – All The web. Couplé avec Tiscali, Nomade propose également un annuaire propriétaire et un système de liens promotionnels utilisant une technologie propre en parallèle de celles d'Overture et d'Espotting.

Nomade propose le même référencement payant que Lycos pour son moteur Fast, et référence gratuitement les sites non commerciaux dans son annuaire. Pour les sites commerciaux, Nomade garanti un délai d'évaluation du site de moins de 2 jours pour 99 euros hors taxe.

iii. Aperçu des offres proposées par les professionnels du référencement

Nous venons de voir que même si un petit nombre d'acteurs représentent près de 90 % du trafic des moteurs de recherche et annuaires mondiaux et français, le paysage des outils de recherche reste complexe en raison des nombreux partenariats, des nombreux « sous-traitants », et en raison de la présence d'offres payantes plus ou moins obligatoires.

De plus, nous avons vu que chaque moteur ou annuaire utilise une technologie d'analyse et de classement qui lui est propre et qu'il est nécessaire de connaître pour construire un site Internet adapté.

A partir de ce constat est née une nouvelle profession : le métier du référencement. De plus en plus de sociétés se créent ou se spécialisent dans ce domaine afin de proposer leurs services aux propriétaires de sites.

Les premières sont apparues aux alentours des années 97-98 et sont issues de domaines très variés comme par exemple l'informatique, la publicité ou encore la communication. Ces sociétés facturent leurs clients généralement à partir de forfaits qu'ils adaptent en fonction du nombre de pages du site à référencer, du secteur d'activité du site, voire même en fonction du résultat obtenu : certains ne facturent leurs clients que si le site est effectivement en première page des outils de recherche après optimisation autour des mots-clés prédéfinis dans le contrat du client.

Cependant, il y a de grandes disparités entre les différentes sociétés de référencement, que ce soit en terme de tarification, d'offres ou de méthodes.

Les offres proposées par ces spécialistes du référencement peuvent être classées en 2 catégories : la mise en place d'une stratégie, et le suivi du référencement.

1. La mise en place d'une stratégie

Les « référenceurs » proposent la plupart du temps de mettre en place entièrement une stratégie globale de référencement qui peut comporter plusieurs phases :

- une étude du projet de site, du thème, du message à faire passer, de l'objectif du site
- une étude des sites des concurrents, des mots-clés importants de leurs sites
- une étude des moteurs de recherche sur ces mots-clés pour connaître les habitudes des internautes et le positionnement des concurrents par rapport à ces mots-clés
- une optimisation du contenu du site du client autour de ces mots-clés à partir des études précédentes
- une optimisation de la hiérarchie et du codage du site du client en fonction des études précédentes et des volontés du client
- une optimisation du nombre de soumissions à des moteurs et annuaires et du délai de ces soumissions
- la mise en place de stratégies particulières afin d'augmenter le classement.

2. Le suivi du référencement

En plus de ces offres ponctuelles, les spécialistes du référencement proposent généralement d'effectuer un suivi du site client et des sites concurrents : une sorte de veille permanente. Ce suivi permet de comparer jour après jour les classements du site du client sur les différents outils de recherche avec ceux de ses concurrents.

Cela permet également d'affiner bon nombre des analyses faites lors de la mise en place, notamment les analyses des mots-clés pertinents.

Enfin, ce suivi permet aux sites de suivre l'évolution des moteurs de recherche, de leurs techniques d'analyse et surtout de classement des résultats afin d'y être toujours adapté.

Nous venons de montrer que pour une entreprise souhaitant être présente sur Internet, il est primordial de prêter une attention particulière au référencement et à son optimisation.

Cependant, les moteurs et annuaires principaux présentent des caractéristiques assez variées, notamment en terme de tarifs et d'audiences.

De plus, le nombre de spécialistes du référencement, ou d'acteurs prétendant l'être, est en constante augmentation. Et chaque société met en valeur des services différents, garantit des résultats différents et utilise des techniques diverses.

Il est donc essentiel de connaître l'efficacité réelle des techniques les plus répandues, ou plutôt l'inefficacité réelle de certaines de ces techniques ainsi que l'utilité réelle de ces spécialistes dans une stratégie de référencement.

II. Analyse de l'efficacité et du coût des principales techniques d'optimisation du référencement : les professionnels sont-ils indispensables ?

Afin de juger de la nécessité et de l'efficacité des techniques principales d'optimisation du référencement, nous allons passer en revue les étapes apparemment essentielles de ce travail par ordre de difficulté croissante.

Ces étapes ont été identifiées par des recherches bibliothécaires et des entretiens avec différents acteurs du secteur.

Pour chacune de celles-ci, nous recenserons les choix qui s'offrent au gestionnaire du site à référencer, nous analyserons leur pertinence, leur coût et les connaissances nécessaires à leur mise en œuvre.

Enfin nous comparerons ces analyses avec les offres des professionnels afin d'identifier la valeur ajoutée apportée par ces derniers et donc leur utilité.

a. Dans quels moteurs de recherche et annuaires faut-il être présent ?

L'étape primordiale permettant d'être présent dans les bases de données des moteurs de recherche et des annuaires est la « soumission ».

Bien que les moteurs de recherche construisent leur base de données grâce à des « robots » qui parcourent le web afin d'indexer le plus de pages possible, il n'existe pas d'assurance que ce robot atteigne un jour le site à référencer.

Il existe donc une procédure permettant en théorie d'éviter cette attente : la procédure de soumission. La plupart des moteurs de recherche et même des annuaires proposent sur leur site une page permettant d'inscrire un site Internet, de le soumettre au moteur ou à l'annuaire.

Chaque moteur ou annuaire a alors un système qui lui est propre afin de répondre favorablement ou non à cette demande d'intégration dans leur base de donnée. Ce système peut être payant ou gratuit selon l'outil de recherche.

La procédure de soumission d'un site à ces différents outils est donc la première chose à étudier, c'est au cours de celle-ci que se font les premiers choix stratégiques.

i. Récapitulatif des principaux moteurs et annuaires français : tarifs et audiences

Nous avons déjà vu que beaucoup d'outils de recherche utilisent des bases de données qui ne leur appartiennent pas en propre. Or c'est dans ces bases de données qu'il faut placer le site à référencer.

De plus, nous avons également vu que nombre de ces outils font payer aux sites leur intégration dans ces bases.

Il apparaît donc nécessaire d'effectuer un récapitulatif des principales offres afin de s'y retrouver. Ce récapitulatif ne tient pas compte du changement technologique de Yahoo, car il est nécessaire d'attendre de pouvoir constater les effets de celui-ci avant d'en tirer des enseignements.

Les principaux moteurs et annuaires français				
Outil de recherche	Moteur utilisé	Offre de soumission	Annuaire utilisé	Offre de soumission
Google	Google	gratuit	Open Directory	gratuit
Yahoo	Google	gratuit	Yahoo	payant
Wanadoo / Voila	Voila	gratuit	Le Guide Voila	payant
Aol.fr	Exalead	gratuit	AOL	gratuit
Msn.fr	Inktomi	payant	La Mine	gratuit
Lycos.fr	Fast	payant	Lycos	payant
Nomade.fr	Fast	payant	Tiscali	payant

Dans le cas des moteurs de recherche, les offres de soumission payante existantes présentent en contrepartie des délais d'indexation dans les bases de données de moins de 3 jours.

S'il n'existe pas de possibilité de soumission gratuite pour les trois moteurs concernés (MSN, Lycos et Nomade), il est cependant tout à fait possible d'être présent dans leur base de données « naturellement », c'est-à-dire en attendant simplement le passage de leurs robots respectifs.

Pour les moteurs permettant une soumission gratuite (Google, Yahoo, Voila et AOL), le délai d'indexation est généralement de l'ordre de un mois, mais peut varier de sept jours à deux mois.

Dans le cas des annuaires, les offres de soumission payante sont des simples options permettant de garantir un délai d'indexation dans le cas de sites non commerciaux, mais elles sont obligatoires pour les sites commerciaux.

Lors d'une recherche sur un de ces outils, les résultats du moteur et de l'annuaire ne sont pas mélangés. Or nous savons que les internautes ne s'intéressent quasiment qu'aux résultats affichés en premier.

Il est donc essentiel de savoir pour chacun des outils de recherche s'il affiche d'abord les résultats de son moteur ou de son annuaire.

De même, il est également très intéressant de comparer la présence d'offres gratuites ou payantes avec l'audience de l'outil concerné.

Les principaux moteurs et annuaires français				
Outil de recherche	1 ^{ers} résultats par	Moteur ou annuaire	Offre de soumission	Audience de l'outil
Google	Google	moteur	gratuit	65,50 %
Yahoo	Google	moteur	gratuit	10,39 %
Wanadoo / Voila	Voila	moteur	gratuit	8,58 %
Aol.fr	Exalead	moteur	gratuit	4,51 %
Msn.fr	La Mine	annuaire	gratuit	4,75 %
Lycos.fr	Fast	moteur	payant	< 1 %
Nomade.fr	Tiscali	annuaire	payant	< 1 %

Ce tableau montre que les outils de recherche affichant en premier les résultats issus d'une offre de soumission payante représentent moins de 2 % du trafic généré par l'ensemble des outils de recherche... et moins de 1 % si l'on considère que l'offre du moteur Fast de Lycos n'est que facultative !

Nous en venons donc tout naturellement à nous interroger sur la nécessité de souscrire à une offre de référencement payante.

Il est nécessaire de rappeler que l'offre de référencement payante de Yahoo n'est pas présentée ici puisque elle est entièrement facultative et puisque l'audience de Yahoo est calculée sur la « période Google » : il convient donc d'attendre les effets de la YST sur l'audience de Yahoo avant d'étudier cette nouvelle offre.

ii. Faut-il payer pour être indexé par les moteurs de recherche ?

Regardons le cas des moteurs de recherche avec système de soumission payante : il s'agit de MSN, de Lycos, et de Nomade qui utilisent respectivement Inktomi et Fast comme bases de données. Il faut alors signaler deux choses importantes à leur propos :

Tout d'abord, il faut savoir que Inktomi et Fast appartiennent désormais tous les deux à Yahoo qui utilise maintenant les résultats de la « YST » pour son propre moteur de recherche. Cependant la stratégie de Yahoo vis-à-vis de ces 2 technologies n'est pas encore connue : il nous apparaît donc nécessaire de suivre les devenir d'Inktomi et de Fast dans ce contexte avant de souscrire à leurs offres payantes.

De plus, le cas de MSN est également particulier, puisque Microsoft prépare actuellement un nouveau moteur de recherche basé sur une nouvelle technologie qui devrait lui être propre, mais dont peu ou pas d'éléments sont encore connus. Là encore, il apparaît donc essentiel d'en savoir plus sur ce nouvel outil avant de décider d'y investir.

Enfin, même si ces 3 outils de recherche proposent des solutions payantes pour leur moteur respectif, ces solutions ne sont pas obligatoires pour figurer dans leurs bases de données : il suffit d'un lien issu d'un site déjà présent dans leurs bases et pointant vers le site à référencer pour que celles-ci l'intègrent automatiquement lors de leur prochain passage sur ce premier site ! Cette solution n'offre pas de garantie de délai mais elle fonctionne !

Signalons également que le trafic généré par ces trois moteurs ensemble ne représente que omis de 6,8 % du trafic généré par les moteurs de recherche : il n'est donc pas trop pénalisant d'attendre l'indexation automatique « naturelle » du site à référencer ou encore d'attendre que les situations de ces 3 moteurs ne deviennent plus claires.

iii. Faut-il payer pour être indexé par les annuaires ?

Regardons maintenant le cas des annuaires : la soumission est devenue payante pour les sites commerciaux sur la plupart des annuaires comme Yahoo, Voila, Lycos et Nomade. Comme nous l'avons vu précédemment, les tarifs sont loin d'être négligeables, d'autant plus que la tendance actuelle semble être le paiement annuel (déjà en place aux Etats-Unis).

Il est donc essentiel d'étudier la pertinence de cet investissement éventuel.

Il existe deux façons d'utiliser un annuaire Internet : la première est d'utiliser le portail de l'outil de recherche pour y saisir des mots-clés afin de trouver les sites de l'annuaire qui correspondent le mieux à ces mots-clés. Malheureusement, nous venons de voir que seuls MSN et Nomade affichent effectivement les résultats de leurs annuaires avant ceux de leurs moteurs. Or la soumission à l'annuaire de MSN étant gratuite, il ne reste que Nomade à afficher en premier des résultats payants.

Il apparaît donc clairement qu'on ne peut pas investir dans les annuaires dans cette seule optique.

La deuxième façon d'utiliser un annuaire Internet est de suivre les catégories de l'annuaire afin d'y trouver une liste (souvent alphabétique) des sites en rapport avec le sujet recherché. D'après le baromètre Ad'oc de septembre 2003, seules 9,1 % des requêtes sur les

outils de recherche se font par catégories. De plus, les outils proposant des annuaires à soumission payante génèrent moins de 27 % du trafic des outils de recherche : autrement dit, environ 2 % des utilisateurs des outils de recherche parcourront les catégories des annuaires à soumission payante...

Etre présent dans les annuaires peut avoir des conséquences plus indirectes sur le trafic d'un site. En effet, nous verrons plus tard que le classement d'un site dans les pages de résultats de Google dépend fortement du nombre de liens pointant vers ce site.

Un raisonnement immédiat permet donc de dire qu'en étant présent sur un annuaire, on augmente ce nombre de liens et donc son classement. Cet avantage est d'ailleurs très souvent mis en valeur par un certain nombre de professionnels du référencement.

Malheureusement, il n'en est rien ou presque ! En effet, des moteurs de recherche comme Google prennent effectivement en compte le nombre de liens qui pointent vers un site pour calculer sa « popularité », mais ils prennent également en compte la popularité de la page sur laquelle est située ce lien et le nombre de liens présents sur cette page : autrement dit, pour que la présence sur un annuaire améliore notablement le positionnement d'un site sur Google, il faudrait que le lien soit positionné « le plus près possible » de la page d'accueil de l'annuaire (en nombre de clics), et qu'il soit quasiment le seul lien cliquable de la page !

Ceci est évidemment contraire aux fonctionnements des annuaires pour lesquels il faut cliquer sur de nombreuses catégories et sous-catégories avant d'aboutir à une page comportant de très nombreux liens !

Etre présent sur un annuaire peut également accélérer la prise en compte de son site par le moteur de recherche associé à cet annuaire. En effet, en étant référencé par l'annuaire Open Directory, un site est automatiquement référencé par Google quelque temps plus tard sans aucune intervention supplémentaire. Il en est de même avec AOL dont le moteur Exalead intègre dans sa base de données tous les sites proposés à son annuaire. Seulement pour AOL comme pour Google, il s'agit d'annuaires à soumission gratuite ! Dans le cas de Yahoo, Voila, Lycos ou Nomade, rien ne semble montrer que la présence dans leurs annuaires entraîne une meilleure indexation dans les bases de données de leurs moteurs.

A part pour Yahoo qui est l'annuaire le plus utilisé au monde, il paraît superflu de souscrire aux offres de soumission payantes des annuaires. Pour le cas de Yahoo, l'appréciation est beaucoup plus nuancée et souscrire à l'offre payante de soumission peut par exemple être justifié par la présence dans ses catégories de plusieurs des concurrents du site à référencer : c'est l'effet salon.

La phase de soumission d'un site aux différents moteurs et annuaires est donc une étape facile et rapide à réaliser. Les seuls coûts de cette étape sont ceux des offres de soumission payantes qui sont à priori superflues sauf dans le cas de Yahoo.

On peut dès lors s'interroger sur la nécessité d'un professionnel du référencement pour cette étape.

iv. Quel est l'intérêt des offres proposées par les professionnels du référencement ?

Les prestations proposées par les professionnels du référencement sont extrêmement variables, tant par leurs contenus que par leurs tarifs. Cependant, elles comportent toutes au moins un élément en commun : la soumission du site du client aux différents outils de recherche. Mais là encore, des disparités sont présentes à ce niveau, et il est nécessaire de les étudier plus en détail.

La prestation la plus courante et la plus naturelle est la soumission du site client aux principaux outils de recherche. Cette soumission, comme nous l'avons déjà vu, est en effet primordiale sur les principaux moteurs de recherche, mais elle est surtout gratuite et facile : il suffit d'aller sur les sites de chacun de ces moteurs de recherche pour y soumettre son propre site en indiquant son adresse Internet (ou URL) et, si besoin est, une description succincte. Mieux, il existe aujourd'hui de nombreux sites Internet qui permettent à partir d'une seule page de soumettre un site à ces différents moteurs de recherche (<http://audit.abondance.com> par exemple).

Il en est de même avec les annuaires payants ou gratuits : il suffit soit d'aller sur leurs sites respectifs, soit d'aller sur un site regroupant les différents liens vers les procédures d'inscription de ces annuaires, pour souscrire, le cas échéant, à leurs offres payantes ou tout simplement pour inscrire gratuitement son site si cela est possible.

Ce que permet le professionnel du référencement dans ce cas, ce n'est que d'éviter d'avoir à trouver et à remplir les quelques formulaires d'inscription... Certains professionnels du référencement (appelons les des référenceurs) proposent à leurs clients des conditions particulières de soumission, avec notamment des délais garantis sur certains annuaires comme Yahoo, Voila, Lycos ou Tiscali, moyennant une certaine somme d'argent : il s'agit en fait purement et simplement des offres payantes proposées par ces annuaires, offres d'ailleurs obligatoires pour les sites commerciaux !

De la même façon, de nombreux référenceurs justifient leurs tarifs pour la phase de soumission par leurs partenariats avec les principaux outils de recherche, partenariats leur permettant d'obtenir des conditions plus favorables en terme de délai comme en terme de classement parmi les résultats. Or il n'existe à ce jour aucun partenariat de la sorte ! Les seules offres possibles des outils de recherche sont les offres connues du public.

Nombre de référenceurs mettent également en avant le nombre d'outils de recherche auxquels ils vont soumettre le site client, et ce moyennant un tarif augmentant avec ce nombre. Or nous avons déjà montré qu'à eux seuls, les 7 principaux outils de recherches utilisés en France représentent environ 96 % du trafic généré par les outils de recherche... Il n'est donc pas très utile en terme de trafic immédiat de soumettre un site à des centaines voire des milliers d'autres outils.

Cependant, nous avons déjà signalé également que le nombre de liens pointant vers un site agit sur le classement de ce dernier dans les résultats de Google par exemple. Être présent dans les bases de données de ces outils pourrait donc améliorer sensiblement ce classement. Dans le cas des annuaires, nous avons déjà vu que ce calcul n'était pas vraiment rentable, et encore moins avec des annuaires de moindre importance. De la même façon pour

les moteurs de recherche, être présent dans une page de résultats au milieu d'un grand nombre d'autres sites n'est pas très efficace, d'autant plus qu'il faudrait pour que ce lien soit comptabilisé par Google, que chaque page de résultats de ces moteurs de recherche soit archivée par leur soin, ce qui est loin d'être le cas !

Enfin, il faut signaler que les mêmes sites qui regroupent en une page l'ensemble des liens permettant d'accéder aux formulaires d'inscription des principaux moteurs et annuaires, permettent généralement également de s'inscrire automatiquement et gratuitement sur une centaine d'outils secondaires en entrant simplement l'adresse Internet (URL) de son site.

Les référenceurs proposent également parfois des logiciels de leur création qui sont censés permettre d'inscrire automatiquement un site à tous les moteurs de recherche ou presque (aux 100 principaux, aux 1000 principaux, etc. ...). Ces logiciels fonctionnent effectivement pour les moteurs secondaires, mais il faut savoir que la plupart des moteurs principaux ne sont pas inclus dans les listes de moteurs de ces logiciels. En effet, Google, Voila, Fast, Exalead ou Inktomi par exemple n'acceptent que les soumissions manuelles et cherchent à empêcher le plus possible les inscriptions automatiques à l'aide de scripts ou de logiciels.

Ces logiciels, payants pour la plupart, fonctionnent donc effectivement, mais seulement sur les moteurs secondaires dont l'audience cumulée est inférieure à 5 % ...

Enfin, les référenceurs proposent également des techniques qui sortent légèrement du cadre de la soumission afin de réduire les délais de prise en compte. En effet, il faut compter de l'ordre de 1 à 2 mois après la soumission de son site pour être à peu près sûr qu'il apparaisse dans les pages de résultats de Google par exemple. Or la fréquence de mise à jour de l'index de Google est dorénavant plus grande, puisque il faut compter en moyenne une quinzaine de jours entre 2 passages du robot de Google sur un site (dans certains cas il arrive même que la visite du robot soit quotidienne).

Un grand nombre de référenceurs proposent donc à leurs clients de créer des sites temporaires, de les inscrire dans les principaux moteurs de recherche et de les remplacer le moment venu par le site réel : le premier délai d'attente étant passé, le site réel devrait à priori apparaître plus rapidement dans les résultats du moteur. Si cette technique semble fonctionner du point de vue des délais de prise en compte du site final, elle demande néanmoins un travail supplémentaire certain : outre la technique de création de ce petit site temporaire, il faut surtout en définir le contenu, trouvant le compromis entre son image et sa simplicité.

Plus encore que pour réduire les délais d'insertion du site réel dans les moteurs, cette technique peut également amener un premier trafic à condition d'y adjoindre un contenu réel et mis à jour régulièrement. Mais nous arrivons là plus dans le cadre du travail du créateur du site, d'une « web agency » et non du « référenceur » à proprement parler : cette dernière option n'est d'ailleurs que rarement proposée par ces référenceurs.

Pour ce qui est de l'étape de la soumission du site à référencer aux différents outils de recherche, il semble que le professionnel du référencement n'apporte rien de convaincant.

Les seuls choix à faire sont des choix stratégiques concernant les offres de soumission payantes, et celles-ci ne semblent pas être utiles hormis pour l'annuaire de Yahoo.

b. Quelle est l'importance de l'optimisation par mots-clés d'un site Internet ?

Si la soumission d'un site est une étape indispensable du référencement de celui-ci, elle est cependant très rapide et ne demande pas de travail particulier autre que quelques choix stratégiques.

Mais nous avons déjà vu auparavant qu'il ne suffisait pas d'être positionné parmi les résultats des moteurs de recherche, il faut surtout être bien positionné pour les résultats qui nous concernent : cela passe par une optimisation du site par rapport à un certain nombre de mots-clés.

i. Rappel du fonctionnement des moteurs de recherche : l'importance des mots-clés

Afin de comprendre l'importance de cette notion de mots-clés, étudions plus en détail le fonctionnement d'un moteur de recherche comme Google.

Lorsqu'un internaute veut utiliser Google pour effectuer une recherche sur Internet, il se rend généralement sur la page d'accueil de celui-ci (www.google.com ou encore www.google.fr pour la version française). Ensuite, dans le cas d'une recherche normale (il existe aussi la possibilité d'effectuer une « recherche avancée » en bénéficiant d'options supplémentaires), l'utilisateur entre une série de mots-clés décrivant sa recherche dans le formulaire prévu à cet usage et lance cette recherche en cliquant sur le bouton approprié. Google va alors chercher dans sa base de données l'ensemble des pages contenant les mots clés demandés, puis il va les classer selon un algorithme qui lui est propre, avant de les afficher sur la page de résultats au rythme généralement de 10 par pages.

Si le détail de l'algorithme de Google est tenu secret, de nombreux éléments sont connus, notamment en ce qui concerne le « PageRank » qui est une sorte de note donnée par Google aux différentes pages de son index.

Sans rentrer dans les détails, Google va chercher à savoir quelles sont les pages dans lesquelles les mots-clés sont les plus présents, mais aussi dans lesquelles ils sont le plus mis en valeur, dans lesquelles ils sont inclus dans l'adresse Internet même de la page (URL), dans son titre, etc. ... Puis il va ensuite étudier l'ensemble des liens situés sur d'autres sites qui pointent vers la page en question, leurs nombres, leurs positionnements, et leurs pertinences par rapport aux mots-clés donnés au départ : ainsi, les mots-clés ne seront pas uniquement cherchés dans les pages données en résultat, mais également dans les pages qui pointent vers celles-ci.

Si cette description très sommaire montre bien la complexité de l'algorithme mis en œuvre par Google pour classer les résultats des requêtes des internautes, elle montre surtout l'importance accordée par l'algorithme aux mots-clés entrés par l'utilisateur.

Il va donc de soi qu'un site Internet doit être optimisé autour de ces mots-clés afin d'être correctement référencé.

Une démarche pour obtenir un trafic maximal sur un site serait par exemple de regarder quels sont les principaux mots-clés entrés par les internautes sur quelques-uns des plus importants moteurs de recherche et d'optimiser ce site par rapport à ces mots-clés afin d'être le plus visible possible... oui mais si vous avez un site spécialisé dans le commerce en ligne de vin, vous n'aurez que faire des internautes souhaitant s'informer sur les dernières frasques d'Eminem...

Ce qu'il ne faut pas perdre de vue, c'est que plus que du trafic, c'est du trafic utile qu'il faut obtenir grâce au référencement : le choix des mots-clés autour desquels optimiser votre site est primordial.

ii. Comment choisir les mots-clés importants ?

Le choix des mots-clés doit se faire autour de trois étapes très distinctes :

- une analyse du site lui-même s'il est déjà créé, ou du projet de site sinon, en s'efforçant de faire ressortir l'objectif réel du site (informer, vendre, etc...), c'est-à-dire en essayant de savoir pourquoi nous voulons que les internautes viennent sur ce site en particulier,
- une analyse des sites des concurrents, de leurs positionnements sur les moteurs, des mots-clés sur lesquels ils se sont appuyés pour construire leur site,
- une analyse des comportements des internautes sur les moteurs de recherche, par rapport aux mots-clés dégagés à la suite des deux premières analyses.

La première étape, qui est assez rapide, n'est pas à négliger : il est important de définir correctement l'objectif principal du site, mais également l'objectif de chaque page et surtout son intérêt pour l'internaute.

Une unique page comportant de multiples informations sans liens entre elles n'a pas de sens général, de cohérence, alors qu'une suite de pages ayant chacune un objet bien précis est beaucoup plus efficace et pertinente : ces pages seront cataloguées plus efficacement par les outils de recherche.

Il ne faut en effet pas perdre de vue que seuls un titre et une description de 3 ou 4 lignes seront visibles dans les pages de résultats des moteurs : il vaut donc mieux avoir plusieurs pages concernant des sujets précis qu'une seule n'ayant qu'une description vague et incohérente car traitant de divers sujets simultanément !

Enfin, les internautes jugeant généralement extrêmement vite un site Internet, il est essentiel d'avoir une description générale du site qui correspond réellement au contenu du site, celui-ci devant ressortir immédiatement dès la page d'accueil.

Cette phase a pour objectif de dégager une première série de mots-clés ou d'expressions qui caractérisent le site à référencer.

L'analyse des sites concurrents doit se faire autour de 2 axes principaux : tout d'abord il faut identifier le vocabulaire utilisé par ces concurrents, leur façon de définir leur métier et leur site, et même la structure de leur site.

Ensuite, il faut faire ressortir plus précisément les expressions et les mots-clés qui ressortent de ces sites, avant de constater leur positionnement sur les différents moteurs de recherche par rapport à ces mots-clés ou par rapport à ceux définis dans la première phase s'ils sont différents.

Pour faire ressortir les mots-clés et les expressions importantes, il est pratique de se servir d'analyseurs syntaxiques comme le « keyword density analyser » du site www.searchengineworld.com ou encore celui du site français www.outiref.com qui propose une analyse automatique très détaillée de l'optimisation d'un site pour son référencement. Ce dernier permet également d'étudier le positionnement d'un site sur les principaux moteurs de recherche pour un mot-clé donné.

De nombreux logiciels permettent d'effectuer de telles analyses de façon plus automatisée, mais ceux-ci sont payants et coûtent généralement entre 200 et 1000 euros...

L'objectif de cette seconde phase est d'affiner la liste d'expressions et mots-clés issue de la première étape, afin d'en obtenir la liste la plus complète et la plus adaptée possible.

Enfin, il reste à confronter cette liste avec les habitudes des internautes : prenons l'exemple d'un site d'auto-école, il est primordial de savoir si « auto école » est plus souvent demandée sur les outils de recherche que « autos écoles » ou encore « permis conduire » !

Il existe là encore des outils gratuits qui permettent de répondre à cette interrogation, avec par exemple :

- <http://fr.espotting.com/popups/keywordgenbox.asp>
- <http://www.fr.overture.com/d/FRm/adcenter/tools/frfrindex.jhtml>

A partir d'un mot clé saisi par l'utilisateur, ces sites donnent le nombre de requêtes associées à ce mot-clé durant les 30 derniers jours, mais également une liste d'expressions similaires avec leurs scores respectifs.

Pour ce faire, Espotting comptabilise les requêtes effectuées sur les moteurs et annuaires partenaires d'Espotting, et Overture fait de même avec ses propres partenaires.

En passant ainsi en revue l'ensemble de la liste formée au cours des deux premières étapes, nous obtenons le classement des mots-clés et expressions les plus importants pour le site à référencer.

C'est finalement autour de ce classement que le site doit être construit, afin de mettre en valeur les expressions les plus recherchées, sans pour autant ignorer les autres formulations utilisées par les internautes.

Cette liste permet en effet de dégager les principaux mots-clés (entre 10 et 15 au maximum) qu'il faudra utiliser et mettre en valeur le plus possible, ainsi qu'un véritable champ lexical à employer dans l'ensemble du site.

iii. Quel est l'intérêt des offres proposées par les professionnels du référencement ?

L'analyse précédente fait généralement partie de ce que les professionnels du référencement appellent « l'audit » d'un site : une analyse du contenu et des objectifs du site, une analyse des concurrents, des requêtes sur les outils de recherche, mais également une analyse de l'état actuel du référencement du site, tant au niveau des résultats que des techniques utilisées pour le cas où le site à référencer est déjà existant.

Le coût de cet audit dépend naturellement de l'ampleur du travail à effectuer, c'est-à-dire principalement de la taille et de la complexité du site. Cependant, le tarif minimum constaté pour ce genre de prestation est d'environ 350 euros en moyenne pour un site de quelques pages et pour moins de 10 mots-clés.

En ce qui concerne les prestations mêmes des référenceurs, il n'y a pratiquement pas d'apport supplémentaire par rapport à ce qu'il est possible d'effectuer gratuitement, si ce n'est une plus grande expérience des sites Internet et des comportements des internautes sur les outils de recherche : une connaissance accrue d'Internet et du référencement n'est pas indispensable à ce niveau. Cependant, cette analyse « syntaxique » du site à référencer, demande un travail certain et méthodique, et donc un fort investissement en temps : environ un à deux jours de travail pleins pour une personne pour un site de taille modeste.

En réalité, ce temps nécessaire dépend plus du nombre de concurrents importants et de la richesse du champ syntaxique correspondant au site qu'au nombre de pages de celui-ci.

Si une certaine expérience des outils d'analyse syntaxique et des outils de recherche peut être avantageuse en terme de rapidité d'exécution de cette phase, c'est surtout une expérience de contenu, de rédaction qui est utile.

Le seul coût de cette analyse syntaxique est le temps d'occupation de la personne en charge de celle-ci : les professionnels du référencement n'apportent aucune autre valeur ajoutée qu'une externalisation de ce coût.

c. Quelle est l'importance de l'optimisation de l'organisation d'un site Internet ?

Pour un référencement optimal, il est indispensable d'optimiser le fond du site à référencer, mais également sa forme, et notamment son organisation générale.

Pour comprendre l'importance de l'organisation générale du site à référencer, prenons le cas de Google et de son algorithme « PageRank » que nous avons déjà évoqué.

Google se sert de cet algorithme afin de calculer une sorte de popularité des pages web présentes dans sa base de données, cette popularité permettant ensuite de classer les résultats d'une requête : en fait, pour afficher ses résultats, Google mixe le résultat du calcul du PageRank avec un calcul de pertinence par rapport aux mots-clés entrés par l'internaute.

Dans un souci de simplicité, nous allons décrire plus précisément le calcul de ce PageRank en faisant abstraction du calcul de pertinence qui l'accompagne.

i. Cas de Google : qu'est-ce que le PageRank ?

Le PageRank de Google est basé sur l'étude des liens entre les pages des sites Internet. L'idée principale repose sur le fait que lorsqu'une page A fait un lien vers une page B, cela signifie que la page A juge la page B comme suffisamment importante pour mériter d'être citée et d'être proposée aux visiteurs. De façon logique, ce lien de A vers B augmente le PageRank de B.

Deux idées supplémentaires essentielles viennent compléter cette définition : premièrement, l'augmentation de PageRank de la page B est d'autant plus importante que le PageRank de la page A est élevé. En d'autres termes, il est bien plus efficace d'avoir un lien depuis la page d'accueil de Google que depuis une page d'un petit site familial.

Deuxièmement, l'augmentation de PageRank de la page B est d'autant plus importante que la page A fait peu de liens. En d'autres termes, si la page A juge qu'il n'y a qu'une page qui mérite un lien, alors il est normal que le PageRank de la page B augmente plus que dans le cas où de nombreuses pages obtiennent un lien.

Voyons maintenant la formulation mathématique de ces 3 caractéristiques :

- soient A_1, A_2, \dots, A_n : n pages pointant vers une page B,
- $PR(A_k)$ le PageRank de la page A_k ,
- $N(A_k)$ le nombre de liens sortants présents sur la page A_k ,
- d un facteur compris entre 0 et 1, fixé en général à 0,85 par Google.

Alors le PageRank de la page B se calcule à partir du PageRank de toutes les pages A_k de la manière suivante :

$$PR(B) = (1 - d) + d \times \left(\frac{PR(A_1)}{N(A_1)} + \dots + \frac{PR(A_n)}{N(A_n)} \right)$$

La seule difficulté de ce calcul est donc sa récursivité : pour calculer le PageRank d'une page, il faut avoir calculé celui de toutes les pages pointant vers elle. Mais cela est également vrai pour ces dernières pages... Il faut donc savoir comment commencer le calcul !

En fait, il suffit de commencer avec des valeurs arbitraires de PageRank (par exemple 1) : le choix de cette valeur n'a pas d'influence sur le résultat final, du moment que toutes les pages commencent avec la même valeur.

Une application de la formule permet de calculer un nouveau PageRank pour chaque page, plus proche de la réalité que la valeur par défaut que nous avons choisie au début. Ensuite, nous recommençons à calculer les PageRank de toutes les pages en appliquant de nouveau la formule, mais en partant cette fois des valeurs que nous avons calculées précédemment.

Après plusieurs itérations, le système converge et les valeurs de PageRank de chaque page ne bougent plus entre deux itérations.

Pour résumer, le PageRank d'une page B ne dépend que du nombre de pages A_k faisant un lien vers B, du PageRank de chaque page A_k , et enfin du nombre de liens sortants de chaque page A_k .

ii. Cas de Google : comment optimiser le PageRank ?

A la vue des éléments précédents, il est clair que pour augmenter le PageRank d'un site, il faut augmenter le nombre de liens pointant vers lui. L'idée la plus évidente est donc d'obtenir des liens venant d'autres sites : ce sont des liens externes, mais nous reviendrons sur ces liens plus tard.

La deuxième méthode pour optimiser la PageRank d'un site, c'est de travailler sur les liens internes. En effet à partir de l'étude de la formule du PageRank, nous pouvons considérer que chaque page possède une réserve de PageRank dont elle peut faire profiter d'autres pages en leur en "transmettant" une partie.

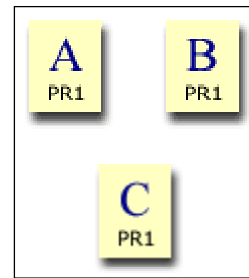
Cela signifie qu'un site Internet possède intrinsèquement une capacité totale de PageRank : si nous considérons que chaque page est créée avec un potentiel de 1, alors le potentiel total d'un site est égal au nombre de ses pages.

De plus, en prenant un site Internet seul, c'est-à-dire sans aucun lien externe, alors son PageRank ne pourra dépasser ce potentiel initial, mais il pourra lui être inférieur dans bien des cas.

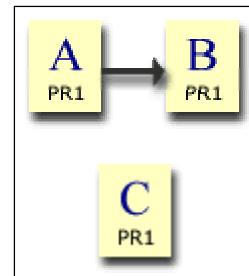
L'optimisation des liens internes d'un site a pour principal objectif d'atteindre ce maximum. L'objectif secondaire mais néanmoins important de cette optimisation est de répartir intelligemment ce PageRank entre les différentes pages.

Etudions pas à pas le mode de calcul du PageRank afin de mieux comprendre l'intérêt de cette optimisation.

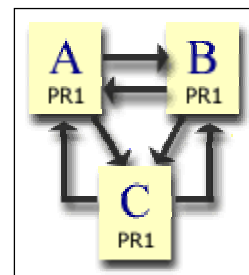
Prenons tout d'abord le cas de 3 pages seules, sans aucun lien entre elles. En leur attribuant une valeur initiale de 1, on obtient une valeur maximale de 3 pour l'ensemble de ces pages. L'application de la formule de calcul sur chacune des pages donne un PageRank de 0,15 pour chacune, c'est-à-dire un PageRank total de 0,45 bien inférieur au maximum théorique. Il y a dans ce cas extrême un important « gâchis » de PageRank.



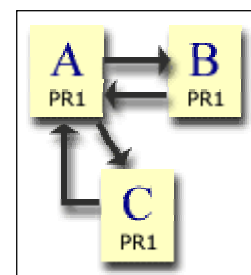
Créons maintenant un lien de la page A à la page B, en attribuant à nouveau 1 comme valeur initiale aux 3 pages. Contrairement au cas précédent où une seule itération de la formule a suffi à obtenir les valeurs définitives, il faut dans ce cas plusieurs dizaines d'itérations pour arriver à des valeurs stables : $PR(A) = 0,15$ $PR(B) = 0,2775$ et $PR(C) = 0,15$. Le PageRank total de 0,5775 est supérieur à celui du cas précédent mais reste bien en deçà du maximum. Cet exemple permet de vérifier un des postulats précédents : en pointant vers la page B, la page A lui a permis d'augmenter son PageRank, augmentant ainsi le total.



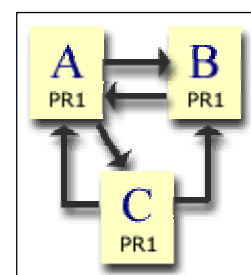
Regardons alors le cas extrême inverse en reliant toutes les pages entre elles et en gardant la même valeur initiale. Dans ce cas, une seule itération suffit puisque les 3 pages gardent leur valeur de départ, à savoir 1. Le PageRank total est alors de 3, c'est-à-dire le maximum possible pour trois pages. Nous venons donc de montrer qu'avec un nombre insuffisant de liens on risquait de « gâcher » du PageRank, alors qu'avec suffisamment de liens on pouvait le maximiser. Cependant, il n'est pas toujours intéressant d'obtenir un PageRank identique pour chaque page : il paraît logique de vouloir favoriser par exemple la page d'accueil d'un site au détriment des autres.



En reliant la page A aux pages B et C, puis les pages B et C à la page A, nous obtenons une sorte de structure hiérarchique. Après plusieurs itérations, nous obtenons les valeurs suivantes : $PR(A) \approx 1,46$ $PR(B) \approx 0,77$ et $PR(C) \approx 0,77$. Si nous obtenons à nouveau la valeur maximale de PageRank total, c'est-à-dire 3, nous avons par contre une page A dont le PageRank vaut presque le double de celui des pages B et C. Dans le cas précédent, les pages B et C distribuaient leur PageRank aux pages A et C, et A et B, alors que dans le cas présent elles ne le distribuent plus qu'à la page A. Nous avons ainsi réussi à mettre en valeur notre page d'accueil.



Etudions un dernier exemple afin de bien comprendre ce phénomène de flux de PageRank, et reprenons le cas précédent en ajoutant un lien de la page C à la page B. Nous obtenons alors $PR(A) \approx 1,3$ $PR(B) \approx 1$ et $PR(C) \approx 0,7$ et donc toujours un total de 3. Le nouveau lien de C vers B provoque la diminution du PageRank de A au profit de celui de B. Cependant cette diminution est atténuée puisque l'augmentation de B permet d'augmenter A en retour. De plus, la baisse du PageRank de A provoque également une diminution de B et C, et ainsi de suite ...



A partir de ces différentes constatations, nous pouvons tirer quelques conclusions sur l'organisation interne d'un site Internet :

- plus un site a de pages, plus il peut avoir un fort PageRank,
- toutes les pages d'un site doivent être reliées,
- une page qui reçoit un lien alors qu'elle n'en émet pas « gâche » du PageRank,
- les pages importantes doivent être mise en valeur par des structures hiérarchiques : toutes les pages pointent vers la page d'accueil, celle-ci pointe sur quelques sous pages seulement afin de ne pas diluer son PageRank, ces sous pages reçoivent les liens des pages qui en dépendent, etc. ...

Le contrôle des liens internes permet également de réduire volontairement la valeur de certaines pages : cela est notamment très intéressant pour les pages de « sortie », c'est-à-dire les pages de liens vers d'autres sites, des liens externes.

Par exemple, faire un lien vers un autre site depuis la page d'accueil va fournir beaucoup de notre PageRank à ce site au détriment de nos propres pages.

Par contre, faire un lien vers un site externe peut être positif si celui-ci fait un lien en retour vers notre page d'accueil...

Parce que de plus en plus de sites proposent à leurs visiteurs la possibilité d'accéder à une page représentant le plan du site, et parce que cette page a des avantages mais aussi des inconvénients, l'étude de celle-ci paraît très intéressante.

En effet, proposer un plan d'un site Internet permet d'abord aux utilisateurs de se retrouver plus facilement dans l'arborescence du site si besoin est. Mais il peut permettre également de remplacer les sempiternelles pages d'erreur habituelles et incompréhensibles qui apparaissent lorsqu'un lien est erroné par exemple (il faut simplement pour cela configurer le serveur pour qu'il fasse ce remplacement).

Moins visible de l'internaute mais tout aussi utile, le plan du site est un moyen sûr de vérifier que l'ensemble des pages sont reliées et peuvent donc être indexées par les outils de recherche. De plus, c'est une page supplémentaire, et donc une augmentation du PageRank possible....

Mais le plan de site peut aussi avoir des conséquences très désavantageuses ! Prenons un site qui soit parfaitement optimisé, avec un fort PageRank au niveau de la page d'accueil, et avec une page de liens externes qui n'a qu'un faible PageRank... et admettons que ce site ait ajouté au dernier moment un plan du site accessible comme cela est souvent le cas par un lien en bas de toutes les pages et notamment sur le page d'accueil...

Alors automatiquement celui-ci va bénéficier d'un très fort PageRank, qu'il va redistribuer à toutes les pages du site présentes sur ce plan, et donc à la page de liens externes pourtant si bien isolée auparavant !

Si la présence sur un site Internet d'une page présentant un plan de celui-ci apparaît comme une bonne idée, il faut néanmoins l'adapter légèrement. Il suffit par exemple d'adapter la structure hiérarchique du site sur celui-ci en créant tout d'abord un plan ne montrant que les grandes catégories, puis des plans plus détaillés accessibles par des clics sur ces dernières.

Nous venons ainsi de montrer qu'il est très important d'accorder une attention particulière à l'organisation interne d'un site à référencer. De plus, si le travail nécessaire semble compliqué pour un néophyte, il ne s'agit que de règles très simples à respecter lors de la création du site. Il va de soi que dans le cas d'un site déjà créé, le travail nécessaire à sa modification est très lourd et peut prendre beaucoup de temps et donc peut être facturé très cher par un professionnel du référencement.

Cependant, la nature même de ce travail fait qu'il doit en fait être effectué principalement à la création du site, et plutôt par un professionnel de la création de sites et non par un référenceur...

L'organisation d'un site ne dépend pas seulement de ses liens internes, mais également de l'organisation de la structure des fichiers qui le composent. Cette structure est, la plupart du temps, choisie par le créateur du site pour faciliter et organiser son travail, et non pour faciliter la lecture du site par les outils de recherches ou par les internautes.

Il est donc nécessaire d'étudier l'influence de cette structure sur le référencement d'un site.

iii. La structure de fichiers a-t-elle un rôle dans le référencement ?

Comme nous l'avons déjà signalé à plusieurs reprises, un fort PageRank ne suffit pas pour être correctement référencé, puisque son calcul s'accompagne toujours d'une analyse syntaxique basée sur les mots-clés entrés par l'internaute : par exemple, au niveau des liens internes à un site, un moteur comme Google pondérera le PageRank calculé précédemment en fonction du nombre de liens contenant le ou les mots-clés dans le texte du lien, mais aussi en fonction du rapport entre le nombre de liens et le nombres de liens contenant le ou les mots-clés dans le texte du lien.

De nombreux outils de recherche, Google en tête, attribue une certaine importance à l'URL (adresse Internet) des pages indexées. Cette URL ne dépendant que des choix du créateur de la page, est très facilement modifiable et peut donc être optimisée. Une adresse d'une page Internet se décompose généralement de la manière suivante :

$$\underbrace{\text{http://}}_{\text{protocole}} \quad \underbrace{\text{www}}_{\text{sous-domaine}} \cdot \underbrace{\text{votresite}}_{\text{nom de domaine}} \cdot \underbrace{\text{com}}_{\text{extension}} / \underbrace{\text{repertoire1/repertoire2/.../}}_{\text{répertoires}} \underbrace{\text{accueil}}_{\text{nom de fichier}} \cdot \underbrace{\text{htm}}_{\text{extension}}$$

Pour chaque page de sa base de données et pour chaque requête effectuée par un internaute, Google analyse l'adresse de la page et plus particulièrement la présence du mot-clé dans l'URL, la position du mot-clé dans l'URL (distance par rapport à la racine), le nombre de caractères de l'URL et le nombre de niveaux de répertoires. Autrement dit, idéalement, il faudrait avoir une adresse comportant le mot-clé, que celui-ci soit le plus proche possible du « http:// », que l'adresse soit la plus courte possible et que la page en question soit directement à la racine du site (sans répertoire)... c'est-à-dire une adresse du type <http://mot-clé.com> !

Le protocole est la partie immuable de l'adresse Internet d'une page d'un site web.

Le nom de domaine et l'extension qui lui est attachée sont au choix du créateur du site, modulo la disponibilité et la légitimité de ceux-ci : en effet, pour obtenir un nom de domaine et son extension, il est nécessaire des les acheter. Mais cet achat ne peut se faire qu'à condition qu'ils soient « libres », c'est-à-dire que personne ne les ait achetés auparavant, auquel cas il faudra alors négocier avec cette personne.

Cependant, le choix du nom de domaine peut être refusé si il n'existe aucune justification d'un droit sur ce nom (secteur d'activité, marque ou nom de famille par exemple) pour ce qui est des extensions en « .fr ».

La deuxième condition est de satisfaire aux conditions d'attribution des extensions : être une association pour obtenir un « .asso.fr », être une entreprise française pour obtenir un « .fr », être une école supérieure ou une université pour avoir un « .edu », etc. ...

S'il est donc difficile d'optimiser l'extension de son nom de domaine pour le référencement, il est au contraire très facile de choisir un nom de domaine « explicite », c'est-à-dire un nom qui représente le plus possible l'activité du site Internet, voire même un nom composé autour des principaux mots-clés issus de « l'audit » du site, afin de se rapprocher de l'idéal <http://mot-clé.fr> par exemple !

Ensuite, la longueur de l'URL et le nombre de répertoires qui la composent étant aussi pris en compte, il est intéressant d'essayer de les réduire le plus possible. Même si créer de nombreux répertoires peut-être utile pour classer les différentes pages, il est préférable d'en réduire leur nombre, par exemple en multipliant les répertoires de mêmes niveaux si besoin est, afin de diminuer le nombre de niveaux de sous répertoires.

De même, les noms de ces répertoires ne doivent pas être choisis pour faciliter le travail des informaticiens qui gèrent le site, mais plutôt pour les rendre explicites et naturels aux yeux des internautes et par conséquent des moteurs de recherche : un dossier « images » dans lequel on place toutes les images du site est bien moins pertinent qu'un dossier du nom d'un produit dans lequel on place les pages qui concernent ce produit avec les images qui s'y rapportent.

De la même façon, il est préférable de nommer les noms des pages en fonction des internautes, et surtout des mots-clés qu'ils sont susceptibles de soumettre aux outils de recherche, plutôt que de les nommer en fonction de la logique informatique : peu d'internautes voulant des informations sur un produit vont chercher le mot-clé « description » au lieu du nom du produit !

L'extension qui termine l'adresse URL d'une page correspond aux types de fichiers de la page (.htm, .html, .php, .asp par exemple). Or ces extensions font rarement partie des mots-clés importants d'un site, sauf cas exceptionnel. Il s'agit donc d'une partie totalement inutile du point de vue de l'optimisation du référencement.

Cependant, il est possible de configurer facilement le serveur hébergeant le site Internet à référencer, afin que celui-ci reconnaisse automatiquement d'autres extensions. Par exemple, il est possible de demander au serveur de considérer que les fichiers se terminant par « .voiture » soient interprétés comme des fichiers « .htm ».

Il est donc tout à fait possible de créer un ensemble d'extensions qui est propre à chaque site, afin de profiter également de cette partie de l'adresse pour le référencement.

Enfin, la dernière partie de l'adresse à étudier est le sous domaine. Généralement, le sous domaine le plus commun est le fameux « www ». Cependant, celui-ci n'apporte rien non plus au référencement. Or il est encore possible dans ce cas de modifier ce sous domaine, soit

en le supprimant purement et simplement, soit en le remplaçant par un autre grâce à la règle suivante : une adresse du type <http://sousdomaine.site.com/> est équivalente à l'adresse plus classique <http://www.site.com/sousdomaine/> . Cette technique permet non seulement de supprimer 4 caractères inutiles de l'URL, mais également un niveau de répertoire.

Elle peut permettre également de contourner certaines contraintes de disponibilités au niveau des noms de domaines.

Le dernier point à avoir en tête lors de l'optimisation des URL est qu'il est inutile de répéter plusieurs fois un même mot-clé dans celles-ci. Par contre, il est très intéressant de profiter de ces techniques pour pouvoir combiner plusieurs expressions et mots-clés dans une même adresse.

Si toutes ces techniques semblent compliquer fortement le travail des créateurs du site et de ceux qui le gèrent au quotidien, il ne s'agit en fait que d'un simple jeu de règles d'écriture à mettre en place dès le début de la création.

De plus, il ne faut pas oublier que rendre les adresses des pages plus explicites pour les outils de recherche, c'est aussi les rendre plus compréhensibles et donc mémorisables pour les internautes.

L'organisation interne du site a une influence directe sur la valeur que lui attribuent les outils de recherche comme Google, et donc sur le positionnement de ce site parmi les résultats de ces derniers. Les liens internes représentent la partie la plus importante de cette organisation interne, mais la structure de fichiers n'est pas à oublier.

Dans le cas d'un site à créer, ces optimisations ne demandent pas un surplus de travail mais une simple adaptation du travail de création habituel. Par contre, dans le cas d'un site déjà existant, les modifications à apporter peuvent demander un lourd travail, mais qui de toute façon sera effectué plus facilement par les créateurs du site que par des référenceurs qui n'ont pas participé à la création de celui-ci.

d. Quelle est l'importance de l'optimisation du code ?

Les différentes pages des sites Internet que nous voyons lorsque nous parcourons le web sont écrites dans différents langages de programmation qui reposent tous sur un élément commun : le html, qui est le langage de base d'une page web. Si celui-ci n'est pas directement lisible par les internautes lorsqu'ils visualisent une page, il n'en est pas de même pour les outils de recherche.

En effet, leurs robots qui parcourent le web afin d'indexer les différentes pages rencontrées ne les visualisent pas comme un internaute d'aujourd'hui, mais plutôt comme un internaute du début de l'Internet, à savoir en mode texte : ils ne voient pas les images, les animations, ni même les différentes mises en page, ils ne voient que le code html de la page ! C'est donc à partir de ce code qu'ils analysent la page, et c'est donc à travers ce code que l'on peut optimiser une page pour son référencement.

Nous avons déjà montré qu'il fallait optimiser l'organisation interne du site ainsi que le texte de celui-ci autour de certains mots-clés soigneusement choisis. Et cette optimisation du texte, c'est essentiellement au niveau du code qu'elle doit se faire.

i. Quelles sont les « balises HTML » prises en compte par les outils de recherche ?

Afin de pouvoir effectuer des mises en page plus ou moins complexes, les créateurs de sites se servent de « balises HTML » : balises formées de caractères spéciaux servant à repérer des parties de texte devant recevoir un traitement particulier. Une balise HTML est constituée d'une directive sous forme de mot-clé (Font, Strong, Table, etc.) encadré par les signes inférieur à (<) et supérieur à (>), qui permet de mettre en forme un texte et qui indique au navigateur web comment devrait être affiché un document.

Dès l'apparition des premiers outils de recherche, des balises spécifiques ont été créées : ce sont les balises « META description » et les balises « META keywords ». Au départ, la balise « META description » permettait au créateur d'un site d'entrer une description succincte de celui-ci, description qui devait être affichée sous le lien du site parmi les résultats d'une recherche. De même, la balise « META keywords » devait permettre au créateur d'un site de rentrer une série de mots-clés pour lesquels le site apparaîtrait dans les réponses des moteurs.

Ces deux balises ayant été très vite détournées de leurs usages par des emplois abusifs (mots-clés n'ayant rien à voir avec le site en question, mots-clés par milliers, etc...) et les technologies employées par les outils de recherche ayant fortement évolué, ces deux balises ont vu leur importance diminuer très fortement, mais elles restent néanmoins utilisées.

Continuons notre étude de l'algorithme de Google : il analyse le contenu de la balise « META description » en testant la présence du ou des mots-clés à l'intérieur du texte de cette balise, il analyse la position de ce ou ces mots-clés en fonction de leur distance par rapport au début du texte de la balise, et enfin il calcule le rapport entre le nombre de mots-clés présents effectivement et le nombre de mots de la balise !

Autrement dit, à trop vouloir mettre de mots dans cette balise, on la rend totalement inefficace... C'est donc un véritable résumé intelligent qu'il faut placer à cet endroit, en essayant d'utiliser le plus de mots-clés et d'expressions possibles parmi les plus importants que nous avons déjà identifiés.

En règle général, on estime que cette balise doit comporter environ entre 150 et 200 caractères espaces compris. Il est également conseillé de l'utiliser en y intégrant une ou plusieurs phrases et non pas seulement une suite de mots.

Pour la balise « META keywords », Google teste à peu près les mêmes éléments, à savoir la position du ou des mots-clés parmi ceux de la balise, ainsi que le rapport entre le nombre de mots-clés présents et le nombre total de mots de la balise. Les mêmes règles s'appliquent donc quant à son utilisation. On estime pour cette balise qu'elle doit comporter au maximum 100 mots ou 1000 caractères espaces compris.

Ces deux balises n'ont aucune influence sur l'apparence d'une page Internet et ne servent qu'aux moteurs de recherche et pas aux annuaires. De plus, leur prise en compte par les moteurs de recherche est de plus en plus faible : il n'est donc pas utile d'y consacrer trop de temps.

De nouvelles balises HTML sont prises en compte par les outils de recherche, et une des plus importantes est la balise « TITLE » qui sert à mettre un titre à chaque page. Celui-ci apparaît alors dans la barre de titre située en haut du navigateur web.

Là encore, Google prend en compte divers éléments de cette balise pour son algorithme de pertinence, à savoir la présence du ou des mots-clés dans le titre, la position du ou des mots-clés par rapport au début du texte, le nombre de caractères dans le titre, la distance entre les mots-clés et le rapport entre le nombre de mots-clés présents dans le titre et le nombre total de mots de celui-ci.

Le titre ne doit pas dépasser les 100 caractères, et ne doit pas contenir si possible plusieurs fois le même mot. Au contraire, il est par contre très utile d'utiliser plusieurs formes d'un même mot-clé : le singulier et le pluriel par exemple.

Cette balise est prise en compte par l'ensemble des outils de recherche (moteurs et annuaires) et elle est évidemment également visible par l'internaute au travers du titre de la page. Il est donc important de s'y intéresser.

Les outils de recherche portent également une importance croissante aux balises de mise en page classiques comme par exemple les balises « <H1>, <H2>, ..., <H6> » qui permettent de faire varier la taille du texte selon des formats prédéfinis (titre 1, sous-titre, etc. ...). Pour ces balises comme pour la plupart des balises actuelles, les outils comme Google testent la présence du ou des mots-clés ainsi que leur position et leur densité parmi les mots qui composent ces balises.

Cependant, en plus de ces derniers éléments, Google attribue un coefficient à ces balises en fonction de leur signification : plus la balise correspond à une marque d'importance des éléments qui y sont rattachés, et plus Google attribue de valeur à la présence du ou des mots-clés dans celle-ci.

On pourrait donc penser que prendre en compte tous ces éléments lors de la création du site à référencer représente une charge de travail supplémentaire conséquente, mais il n'en est rien : il suffit d'avoir une mise en page cohérente, avec notamment les mots-clés et expressions importants mis en valeur avec du gras, des titres, etc. ...

De nombreuses balises ont été détournées afin d'améliorer le classement de sites sur les outils de recherche, comme par exemple en modifiant les couleurs de texte et de fond afin qu'on ne puisse plus lire le contenu du texte et y placer une liste de milliers de mots-clés.

Ces techniques, bien que toujours répandues, sont heureusement de moins en moins efficaces puisque détectées par la plupart des outils de recherche.

Les sites utilisant ce genre de pratique sont même de plus en plus souvent bannis des résultats des moteurs de recherche.

Enfin, le dernier élément important du code d'une page web est son contenu, c'est-à-dire le corps du texte qui la compose. En effet, l'algorithme de Google prend en compte notamment la présence et la position du mot-clé dans le texte par rapport au début de celui-ci, le nombre de caractères dans le texte, le nombre d'occurrences du ou des mots-clés dans le texte, la densité du ou des mots-clés dans le texte, ou encore l'ordre des mots-clés...

Autrement dit, plus le texte est pertinent par rapport à la série de mots-clés recherchés, plus la page sera classée dans les premiers parmi les résultats de la recherche !

Le travail à effectuer sur la page n'est donc plus réellement un travail de référencement, mais plus un travail sur la pertinence du contenu : il faut éviter le contenu inutile et mettre en valeur les mots importants.

L'optimisation du code des pages d'un site Internet doit se faire dans un but de pertinence de la mise en page et du contenu l'un par rapport à l'autre, et non réellement dans un but de référencement.

Cependant, il existe des cas où une optimisation particulière du code semble nécessaire pour le référencement : il s'agit par exemple du cas des frames ou de celui des pages dynamiques.

ii. Faut-il une optimisation du code particulière : le cas des frames

Le principe des « frames », ou cadres, repose sur la possibilité d'afficher une page composée de plusieurs pages : par exemple, la page peut se composer d'une première page située à gauche de l'écran et prenant environ le quart de celui-ci, et d'une deuxième occupant le reste. Utiliser ainsi 2 pages simultanément permet de ne modifier qu'une d'entre elles lors d'un clic sur une des deux pages.

Prenons l'exemple d'un site marchand construit à l'aide de « frames » autour de 2 cadres de pages : le cadre de gauche montre une page de menu où sont répertoriées les différentes rubriques. Ce cadre est constant et restera visible en permanence quelque soit le comportement de l'internaute. Le cadre de droite est quant à lui occupé par la page correspondant à la rubrique choisie par l'internaute dans le cadre de gauche. Ainsi, un clic dans l'un ou l'autre des cadres n'aura d'effet que sur celui de droite, évitant ainsi le rechargement complet de la page.

Les frames sont construites autour d'une page spécifique qui indique au navigateur qu'il doit afficher plusieurs pages en même temps, leur position respective et le nom des pages à afficher. Cette technique est cependant assez récente, et généralement seuls les navigateurs de moins de 2 ans ont la capacité d'afficher correctement ces pages.

Malheureusement, nous avons déjà signalé que les robots des outils de recherche se comportent la plupart du temps comme des navigateurs de première génération (navigateur en mode texte) : il existe encore de nombreux outils de recherche qui ne sont pas capables d'indexer correctement les sites construits autour des frames. Arrivés sur la page principale, ils ne verront pas de liens vers d'autres pages et arrêteront là leur indexation du site !

Il ne faut toutefois pas renoncer pour autant à utiliser des cadres pour la mise en page de son site car non seulement la plupart des outils de recherche apprennent actuellement à gérer les frames, mais il existe de toute façon des techniques permettant de contourner ce problème.

Nous venons de voir que ces frames sont construites autour d'une page principale qui, à l'aide de la balise HTML « <FRAME> » signale au navigateur qu'il va devoir afficher plusieurs cadres. Il est alors possible d'ajouter en bas de cette page principale une balise « <NOFRAMES> » dont le contenu ne sera visible qu'au cas où le navigateur utilisé n'a pas la capacité de gérer les frames.

Il suffit donc d'ajouter grâce à cette balise une version sans cadre de la page, voire tout simplement les deux liens vers les pages censées être contenues dans les cadres. De cette

façon, lorsque les robots des outils de recherche arriveront sur la page en question, ils pourront suivre ces liens pour ensuite indexer l'intégralité du site.

Malheureusement, indexer l'intégralité des pages d'un site de cette manière peut avoir des inconvénients : si un internaute accède grâce à un outil de recherche à une page autre que la page principale définissant les frames, alors il obtiendra une page qui aurait du en théorie se trouver dans un cadre, par exemple à côté d'un menu, sans aucun cadre autour... la navigation dans le site peut alors être très difficile puisque sans ce menu ! Mais là encore il existe une solution simple pour contrer ce problème, en utilisant un « javascript », c'est-à-dire une petite application qui, dans ce cas, ajoute automatiquement les frames nécessaires autour d'une page si celle-ci est appelée directement.

Ce petit script n'est pas pris en compte par les outils de recherche, et ne gêne donc pas leur parcours du site.

Là encore, il peut paraître compliqué d'utiliser des techniques spécifiques pour le référencement d'un site par les outils de recherche, mais il ne faut pas perdre de vue que ces techniques ne servent en fait qu'à rendre le site visible par des internautes n'ayant pas la configuration nécessaire pour visualiser correctement le site en question : grâce à cela, même si le navigateur utilisé n'est pas à même de prendre en compte les frames du site, l'internaute pourra le visiter sans difficulté.

Plus qu'une optimisation particulière du code pour le référencement, il s'agit donc encore une fois d'une optimisation ayant un impact réel sur ce que voient les internautes, même si elle ne concerne qu'une partie de ceux-ci.

iii. Faut-il une optimisation du code particulière : le cas des pages dynamiques

Avec le développement de l'Internet, de nouvelles techniques sont apparues pour la création de pages spécifiques, notamment avec la possibilité de créer des pages « dynamiques ». On appelle un site « dynamique » un site web dont les pages HTML se construisent au fur et à mesure de leur consultation par un internaute en sollicitant des bases de données et des mises en forme prédéfinies.

La plupart des sites marchands sont des sites dynamiques construits autour de leur base de données de produits, ce qui a pour but de faciliter leur mise à jour, tant sur le fond que sur la forme : ajouter un produit à la base de données aura pour effet immédiat sur le site de créer un nouvel espace présentant ce produit en utilisant la mise en page prédéfinie. De même, changer la mise en page prédéfinie des produits aura un effet sur l'ensemble des pages produits du site.

Pour fonctionner, ce genre de page utilise des variables qui permettent d'aller chercher les données correspondantes dans leur base de données. Ces variables sont généralement incluses dans l'adresse URL de la page à créer (article.php) sous la forme : « article.php?page=12&id=2 ».

Ce type d'adresse a longtemps posé des problèmes aux outils de recherche qui en détectant la présence d'un symbole comme « ? » dans l'URL choisissait de ne pas indexer la page concernée.

Aujourd'hui, la plupart des principaux moteurs et annuaires acceptent ce type de page, mais restreignent souvent le nombre de variables prises en compte. Dans le cas de Google par exemple, le robot ignore volontairement toutes les pages dynamiques ayant plus de 3 variables.

Certains sites étant entièrement construits autour de base de données, il est essentiel pour eux de contourner cette limitation. Ceci est rendu possible par une technique appelée « URL Rewriting » proche de celle utilisée pour modifier les extensions de pages. Cette technique permet par exemple de transformer l'URL se terminant par « article.php?page=12&id=2 » en une URL se terminant par « article_12_2.html » et donc prise en compte normalement par les robots des outils de recherche.

Nous ne rentrerons pas dans les détails de la manipulation, mais cette technique se fait, comme pour le changement des extensions des pages, par une modification de la configuration du serveur qui héberge le site Internet concerné.

Un autre type de pages a vu le jour ces dernières années avec l'apparition de la technologie Flash développée par la société Macromedia. Cette technologie permet de créer tout ou partie d'un site autour d'animation graphique. L'animation en question est alors une sorte de programme qui se lance quand un internaute arrive sur la page web le contenant et qui se joue comme un petit film interactif.

Tout comme les frames, cette technologie n'est pas compatible avec les navigateurs trop anciens, et donc avec les robots des moteurs et annuaires. En effet, seuls quelques moteurs très spécialisés sont aujourd'hui capables d'indexer le contenu de ces animations dans leur base de données.

Cependant, il à nouveau possible de créer une version classique du site pour les internautes n'étant pas équipés des derniers navigateurs Internet. Cette version sera alors prise en compte sans difficulté par l'ensemble des outils de recherche.

Nous retrouvons là encore dans le cas du Flash une optimisation qui a un impact visuel sur l'internaute et non uniquement sur le comportement des outils de recherche. Le seul cas où une optimisation particulière semble nécessaire est donc pour les pages dynamiques type « .php ». Mais de plus en plus de moteurs prennent en compte naturellement ce type de pages, et l'optimisation qui est aujourd'hui nécessaire ne le sera bientôt plus.

Rendre un site compatible avec les robots des outils de recherche, c'est donc avant tout le rendre compatible avec des navigateurs plus anciens, et donc le rendre accessible à une plus grande part des internautes.

e. Quel est l'intérêt des professionnels du référencement ?

Nous avons déjà vu qu'en ce qui concerne la soumission d'un site aux différents outils de recherche et le travail d'identification des mots-clés importants, les professionnels du référencement n'apportent qu'un gain d'expérience du référencement et d'Internet qui n'est en général pas indispensable.

Il reste donc à savoir ce qu'il en est en ce qui concerne les optimisations de l'organisation interne du site à référencer et du code de celui-ci.

i. L'utilité des professionnels pour l'optimisation de l'organisation interne et du code d'un site

Que ce soit pour l'organisation des liens internes du site à référencer ou pour l'optimisation de la structure de fichiers utilisée, le travail à effectuer dépend considérablement de l'avancée du site. En effet, si celui-ci est encore à créer, les modifications apportées par le souci de l'optimiser pour le référencement ne sont pas très importantes, puisque ce sont comme nous l'avons déjà dit de simples règles facilement respectables. Par contre, il va de soi qu'en cas de site déjà entièrement créé, les modifications en terme de structure des liens comme de structure de fichiers sont extrêmement lourdes à mettre en place puisque nécessitent une refonte quasi-totale du site !

De plus, si cette refonte totale doit être effectuée par des professionnels du référencement qui n'ont à priori pas participé à la création même du site en question, le travail à effectuer sera beaucoup plus difficile, et par conséquent beaucoup plus long et plus cher, que si cette refonte est faite par les créateurs du site qui le connaissent déjà.

Il en est de même pour l'optimisation du code : la volonté des outils de recherche de se rapprocher du mode de vision des internautes en terme de calcul de pertinence des pages web rend quasiment inutile la connaissance accrue des éléments de code pris en compte par ces outils. Il suffit en effet de rechercher la pertinence, tant en terme de contenu que de mise en page, aux yeux de l'internaute pour obtenir une pertinence aux « yeux » des outils de recherche.

Les deux seules règles à respecter sont de ne pas créer de texte inutile sous peine de diluer cette pertinence, et de profiter de toutes les possibilités offertes de mise en page pour améliorer la présence des mots-clés les plus importants (liens explicites, titres des pages, légendes d'images, etc. ...).

Si la soumission d'un site aux différents outils de recherche, ainsi que l'analyse des mots-clés peuvent être effectués en interne, les optimisations de l'organisation du site et du code de celui-ci demandent des connaissances informatiques plus poussées.

Cependant, ces compétences correspondent parfaitement à celles des professionnels de la création du site Internet, et il semble donc quelque peu inutile de faire appel à des référenceurs uniquement pour ce travail.

ii. L'utilité des professionnels pour leurs propres techniques d'optimisation

Les professionnels du référencement communiquent également sur leur capacité à améliorer le positionnement d'un site sur les différents outils de recherche sans toucher directement à celui-ci, grâce à des techniques que l'on pourrait qualifier d'« externes ». Les 3 principales sont l'échange de liens, les sites satellites et le cloacking.

L'échange de liens est fondé sur le constat que plus un site a de liens pointant vers lui, plus il est considéré comme populaire, et mieux il est positionné parmi les résultats des outils de recherche. Si ce constat est vrai, il est néanmoins nécessaire de le modérer : le positionnement n'est ni directement lié à ce nombre de liens, ni uniquement. En effet, comme nous l'avons déjà remarqué, plus le PageRank de la page faisant un lien vers le site à référencer est grand, plus l'impact de ce lien sur le positionnement de ce site est important. De même, moins il y a de liens issus du site, plus son impact est fort. Enfin, c'est le PageRank dont est issu le lien qui est pris en compte et non le PageRank du site auquel appartient cette page. Autrement dit, il est bien plus efficace d'avoir un lien issu d'une page de PageRank 3 sur laquelle il n'en existe pas d'autres, que d'avoir un lien venant d'une page de PageRank 10 qui propose une dizaine de liens !

Il faut également rappeler que la pertinence entre en compte dans la valorisation de ces liens externes : un lien issu d'une page comportant le ou les mots-clés recherchés, voire même un lien dans lequel sont présents le ou les mots-clés recherchés a un impact beaucoup plus fort qu'un lien issu d'une page n'ayant aucun rapport avec le domaine concerné.

En conclusion, un lien issu d'une page n'ayant qu'un très faible PageRank et n'ayant pas de rapport réel avec le site à référencer n'a quasiment aucune influence.

La technique d'échange de liens proposés par de nombreux référenceurs s'appuie sur le réseau de sites clients de ces référenceurs, ainsi que sur leurs propres sites. Ils proposent alors la possibilité à leur client d'ajouter un lien vers le site à référencer sur l'ensemble des sites de leur réseau, en échange de liens réciproques depuis le site client vers les autres sites du réseau.

Cette technique a en effet l'avantage de générer extrêmement rapidement un certain nombre de liens. Cependant, elle a également des inconvénients qui peuvent être conséquents.

Tout d'abord, la plupart des sites acceptant que leur référenceur ajoute des liens sur leur site sont généralement relativement modestes et n'ont que de faibles PageRank.

De plus, le lien ne sera pas créé sur les pages d'accueil, mais sur des pages de liens, donc des pages de moindre importance comportant de nombreux liens sortants, et généralement sans rapport avec le site initial. L'efficacité de ces liens est donc relativement faible, mais cela reste toujours bon à prendre.

Ensuite, ces liens ne sont possibles qu'en échange de liens depuis le site à référencer, ce qui a pour effet immédiat de multiplier les liens sortants de celui-ci. Et même si la plupart de ces liens sont issus d'une page de faible PageRank, il n'est pas rare que le référenceur demande à ce que le lien le concernant soit visible depuis la page d'accueil, voire même depuis l'ensemble des pages.

Le plus gros gain de cette technique semble être pour le site du référenceur lui-même.

Enfin, le dernier inconvénient est de ne pas réellement contrôler les sites qui vont ainsi être reliés au site à référencer... il peut par exemple être dommageable pour un site sur la prévention routière d'être reliée à un site de vente d'alcool !

Si le principe de multiplication de liens vers le site à référencer est primordial, il est tout de même préférable de l'entreprendre soi-même afin de choisir au mieux les sites partenaires et de négocier avec eux les emplacements les plus adéquats pour les liens réciproques.

La deuxième technique mise en place par de nombreux référenceurs est la création de sites satellites. Ce sont des sites généralement d'une seule page qui ne sont créés qu'autour d'un seul mot-clé ou d'une seule expression et qui ne proposent qu'un lien vers la page d'accueil du site à référencer ainsi que des liens entre eux.

En achetant des noms de domaines pour chaque mot-clé important, puis en créant à chaque fois un site d'une page optimisée autour de ce seul mot-clé, et enfin en soumettant ce site au maximum de moteurs et d'annuaires gratuits, on espère ainsi optimiser la page d'accueil du vrai site pour chacun des mots-clés pour lesquels on a créé un site satellite.

La faiblesse du contenu de ces sites satellites et le fait qu'aucun site ne pointe vers eux rend cette technique relativement peu efficace à moins d'en créer un très grand nombre. Mais cela signifie également acheter un très grand nombre de noms de domaine !

Par contre, dans la plupart des cas, les référenceurs qui proposent des sites satellites à leurs clients pensent également à rajouter sur chacun de ces sites un lien vers leur propre page d'accueil, contribuant à nouveau à leur propre optimisation...

Outre sa faible efficacité (pour le client), le principal inconvénient de cette technique malheureusement assez répandue, est qu'elle est contraire aux différentes chartres de bon usage des outils de recherche qui commencent à bannir les sites utilisant ces techniques.

Enfin, la dernière technique parfois proposée est appelée « cloaking », ou « dissimulation » en français. Il existe plusieurs variantes de cette technique, mais toutes reposent sur la même idée : créer des pages différentes pour les internautes et pour les outils de recherche.

Il existe en effet un moyen de « parler » aux robots des moteurs de recherche, et par exemple de leur interdire l'accès à certaines pages. Cette possibilité se fait grâce notamment à un fichier nommé « robots.txt » que l'on place sur le serveur hébergeant le site en question. Ce fichier peut contenir par exemple des instructions permettant au robot de Google d'indexer le site, mais interdisant à celui de MSN de le faire. L'utilité principale de ce fichier est de signifier aux robots que l'on désire qu'une partie du site ne soit pas référencée, comme par exemple une base de données confidentielle, etc. ...

Pour ce faire, le fichier indique les différentes permissions d'accès accordées en fonction des noms des robots ou de leurs adresses IP. Il est alors possible de détourner quelque peu ce fichier de son usage principal en indiquant par exemple que les internautes ont la possibilité d'accéder à certaines pages et les robots à d'autres !

Quel intérêt cela a-t-il ? Cela permet de ne pas se préoccuper de l'aspect graphique de la version du site destinée aux outils de recherche, et donc de l'optimiser intégralement autour du texte de celui-ci. Les différentes pages seront ainsi notablement allégées et nettoyées de

tout code superflu, ce qui devrait améliorer la pertinence de leur contenu, et donc leur positionnement parmi les résultats des moteurs de recherche.

Tout comme dans le cas de la création de sites satellites, cette technique va à l'encontre des chartes de bon usage des outils de l'Internet puisque ce qui se trouve référencé par les moteurs est différent de ce qui est visible par les internautes ! Heureusement, les principaux outils de recherche développent des techniques de plus en plus perfectionnées afin de détecter ce genre de pratique, et de bannir de leurs bases de données les sites incriminés.

L'utilisation de sites satellites et du cloaking s'apparente à des techniques de « hackers » s'escrimant à piéger les algorithmes des différents outils de recherche afin d'améliorer le positionnement des sites concernés sur ces outils. Ces techniques tendent à disparaître, mais d'autres plus perfectionnées font également leur apparition.

Dans le même temps, les outils de recherche cherchent à faire disparaître ces techniques en améliorant leur propres algorithmes et en essayant de les rendre de moins en moins vulnérables. Cette course, proche de celle que connaissent les logiciels anti-virus avec les différents virus qui parcourent le web, semble ralentir aujourd'hui, surtout grâce à l'effet des bannissements des sites, même si ces derniers sont encore en nombre insuffisants. Ces trois techniques proposées par certains professionnels du référencement sont donc à éviter toutes les trois.

Par contre, il est vrai qu'une politique d'échange de liens réalisée en interne peut-être extrêmement efficace et améliorer considérablement le positionnement d'un site parmi les résultats des outils de recherche.

Mais cette politique concerne alors plus un département de communication en ligne, qu'un prestataire extérieur de référencement.

L'impact de la présence dans les bases de données des outils de recherche sur le trafic d'un site est prouvé et conséquent.

De même, l'impact de l'optimisation du site au niveau de son organisation interne, de son code et de son texte sur le positionnement de celui-ci, et donc sur son trafic, est également réel et prouvé.

Les seules connaissances nécessaires à la mise en œuvre de ces différentes optimisations sont des connaissances en rédaction pour le contenu et en création de site Internet pour le code, et ne nécessitent pas de faire appel à des professionnels du référencement.

Si une entreprise considère qu'elle n'a pas en interne les ressources humaines nécessaires à la mise en place de cette optimisation, c'est plus vers le prestataire qui a créé son site qu'elle doit se tourner, et non vers un nouveau prestataire.

III. Mise en pratique, comparaison des coûts et modèle mathématique

a. Mise en pratique des recommandations

Les délais de prise en compte d'un site Internet et de ses modifications par les outils de recherche ne permettent pas de comparer une à une les différentes techniques évoquées avec celles des professionnels du référencement.

Cependant, une expérimentation de l'ensemble de mes recommandations a pu être réalisée sur la base d'un site Internet ayant les caractéristiques suivantes :

- site institutionnel (non marchand) d'information générale et d'information plus spécifique sur un certain nombre de produits ;
- environ 400 pages ;
- architecture basée sur des frames : la page principale est encadrée à gauche par une page comportant les différents menus du site, et en haut par une page de titres comportant notamment des liens vers différentes rubriques. Cette structure est assez typique de l'utilisation des frames ;
- environ 120 pages dynamiques (de type « php ») construites autour de bases de données permettant la mise à jour d'informations quotidiennes, et également l'affichage de données concernant la région de l'internaute grâce à des techniques de géolocalisation ;
- secteur très concurrencé.

Les caractéristiques de ce site en font un exemple assez complet et représentatif des sites Internet français. L'ensemble des techniques recommandées ont été mises en œuvre sur mes conseils par l'équipe en charge de la création du site : la gestion des frames, des pages dynamiques, la structure de fichiers, l'organisation des liens internes, les url, la pertinence du contenu, etc....

Le surcoût engendré par la prise en compte du référencement et de son optimisation revient à 3 jours de travail pour un développeur : un jour d'apprentissage des règles de base à respecter et deux jours d'« audit » du futur contenu et des principaux sites concurrents.

Le reste des optimisations n'a pas représenté de surcoûts puisque ne consistait qu'à appliquer des simples règles d'écriture, comme nous l'avons déjà vu. En estimant à 3.000 € le salaire moyen d'un tel développeur et à 6.000 € son coût pour l'entreprise, le coût de l'optimisation du référencement d'un tel site est d'environ 900 euros.

Le résultat d'une telle optimisation est conforme aux attentes puisque le site est apparu dans la première page de résultats des principaux moteurs.

b. Étude du tarif des professionnels du référencement

En raison de l'impossibilité de faire faire le même travail par plusieurs professionnels du référencement, des devis détaillés décrivant le travail effectué et le prix des prestations ont servis de base d'étude.

A partir de la liste des adhérents à l'IPEA (Internet Positioning European Association) qui regroupe les principaux professionnels européens du référencement autour d'une charte de déontologie, du réseau Abondance (réseau dédié au référencement), et du principal annuaire du référencement (www.annuaire-referencement.com), 35 sociétés de référencement ont été contactées et 11 ont accepté de répondre.

Elles peuvent être classées en 3 catégories en fonction du niveau du travail proposé : celles qui proposent des prestations de base, des prestations moyennes et enfin des prestations supérieures.

i. Prestations de base

4 sociétés de référencement ont proposé une prestation de base, à savoir :

- une phase d'audit du site et éventuellement d'un ou deux concurrents ;
- la création de pages satellites optimisées pour chaque mot-clé ;
- et parfois des recommandations générales.

A titre d'exemple, la prestation proposée par l'une de ces sociétés comprend la recherche des mots-clés importants et la création puis l'optimisation d'une page par mot-clé trouvé.

Selon cette société :

« Cela permet de positionner ces pages, et donc le site Internet, dans la première page de résultats des outils de recherche pour chacun de ces mots-clés. Pour le cas d'un travail sur une dizaine de mots-clés, il suffit donc d'optimiser en tout et pour tout 10 pages pour positionner convenablement le site. Le nombre de pages du site ainsi que sa structure (frames et pages dynamiques) n'ont alors aucune influence sur le travail à effectuer et donc sur le prix de la prestation. Le prix ne dépend en effet que du nombre de mots-clés pour lesquels une page doit être créée, ainsi que, dans une moindre mesure, du niveau de concurrence existant sur ces mots-clés. Le tarif proposé est de 4.500 € pour dix mots-clés. »

Il faut également remarquer qu'une telle prestation est présentée par cette société comme étant « la prestation complète de référencement ».

Pour commenter cette prestation, nous avons déjà vu que les pages ou sites satellites n'ont qu'une efficacité très réduite, voire nulle, et se positionnent à la limite des comportements tolérés par les moteurs de recherche.

De plus, nous avons également montré que plusieurs optimisations qui ne sont pas incluses dans cette offre sont indispensables pour un référencement correct de l'intégralité du site, comme par exemple la prise en compte des frames.

Ce type de prestation peut à la rigueur convenir dans le cas de secteurs très faiblement concurrentiels et pour lesquels aucune optimisation n'a été faite par des concurrents, mais il reste inefficace dans tous les autres cas.

ii. Prestations moyennes

5 sociétés de référencement ont proposé une prestation moyenne, à savoir :

- une phase d'audit du site et des principaux concurrents ;
- éventuellement la création de pages satellites optimisées pour chaque mot-clé ;
- des recommandations générales ;
- des optimisations techniques visant à permettre l'indexation de l'ensemble des pages du site par les outils de recherche (gestion des frames et des pages dynamiques notamment)

A titre d'exemple, une de ces sociétés propose, en plus des prestations de base, le travail suivant :

« Une refonte totale du site en exportant les bases de données du site Internet vers nos propres serveurs qui se chargeraient alors de dupliquer automatiquement les pages dynamiques créées autour des bases de données en pages statiques lisibles par les moteurs ; et une soumission manuelle dans environ 500 outils généralistes francophones, puisque les petits moteurs se référencent eux mêmes dans les gros : donc, si vous êtes inscrits chez les petits, vous pouvez y gagner doublement en présence chez les gros. »

Il faut là encore remarquer que cette société présente son offre comme étant « la prestation complète de référencement »...

Cette prestation amène deux commentaires :

- Tout d'abord, elle a le mérite de rendre accessible l'ensemble des pages du site aux moteurs de recherche, ce qui est très positif et marque une réelle différence avec les prestations de base. Cependant, et c'est regrettable, cette société se contente de les rendre accessibles, c'est-à-dire de les faire indexer par les moteurs, mais ne les optimise pas une à une ... Or c'est en optimisant chacune des pages d'un site que celui-ci est le mieux optimisé.

- Ensuite, concernant la soumission manuelle dans 500 outils francophones qui est justifiées par la société par le fait que les « petits » se référencent dans les « gros », nous avons déjà montré que cela n'entraînait pratiquement aucune amélioration du classement dans les principaux moteurs.

Ce type de prestation apporte un élément supplémentaire certes important, mais reste incomplète, et donc loin de l'optimisation maximum annoncé.

iii. Prestations supérieures

2 sociétés de référencement ont proposé une prestation supérieure, c'est-à-dire comprenant l'optimisation de chacune des pages du site à référencer.

La première d'entre elles propose une offre originale puisque facturée à la page : l'audit d'une page et son optimisation coûtent un minimum de 6 € (tarif dégressif). Cette prestation apporte la certitude d'un réel travail personnalisé sur chaque page, mais est au total extrêmement coûteuse : 24.000 € pour notre site de 400 pages...

La seconde société est celle dont l'offre est la plus complète :

- une phase d'audit du site et des principaux concurrents ;
- des recommandations générales ;
- des optimisations techniques visant à permettre l'indexation de l'ensemble des pages du site par les outils de recherche (gestion des frames et des pages dynamiques notamment) ;
- des optimisations de chaque page du site ;
- des optimisations avancées concernant la structure de fichiers et les « url » sont envisageables.

Pour cette société :

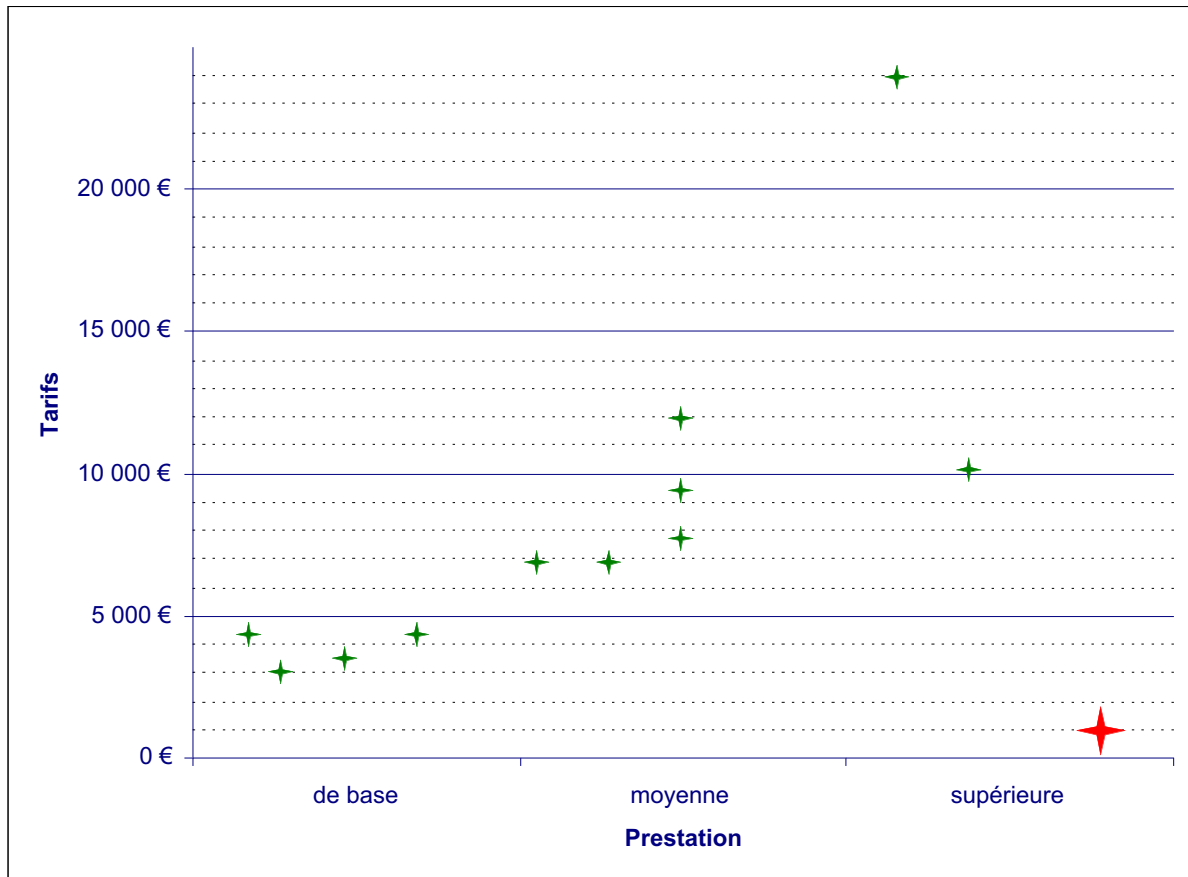
« Le premier niveau de référencement d'un tel site comprend une analyse des objectifs de celui-ci, une analyse du secteur et des concurrents, ainsi que le référencement et l'optimisation de l'ensemble des pages statiques du site. Cela passe par la prise en compte des frames et du javascript éventuel, et par l'optimisation individuelle de chaque page (titre, balises, contenu, etc. ...). Une prestation plus avancée comprend la prise en charge des pages dynamiques de type « php » permettant de les rendre accessible par les moteurs. On obtient alors un tarif d'environ 10.000 €. Enfin, il est également possible de travailler sur la structure de fichiers, le nom des pages et les « url ». Il s'agit alors de la prestation maximum, prestation que nous sommes les seuls à notre connaissance à proposer à ce jour. Le tarif est évidemment supérieur... »

L'offre de ce prestataire est certainement l'offre la plus professionnelle de celles qui nous a été de rencontrer. Néanmoins nous notons que le travail sur les fichiers et les « url » n'est pas encore systématiquement proposé, et demeure à ce jour, pour cette société, l'optimisation

ultime à venir. Cette dernière entraînera naturellement un coût supplémentaire au 10.000 € pour le client.

Cette prestation se rapproche des recommandations que nous avons décrites tout au long de ce document et confirme aussi bien leur efficacité que leur importance.

iv. Comparatif des offres des professionnels du référencement



★ : société de référencement

★ : réalisation en interne selon mes recommandations

Le graphique ci-dessus résume les différentes prestations et coûts exprimés par les sociétés de référencement ayant accepté de nous répondre.

Il confirme à la fois l'hétérogénéité des niveaux de prestations et celle des niveaux de prix, alors qu'elles sont toutes annoncées comme optimale par chacun des prestataires.

Pour une prestation réellement maximum, nous voyons très clairement l'écart en terme de tarifs entre la solution en interne avec l'application de mes recommandations et la réalisation par un prestataire externe.

c. Création d'un modèle mathématique

Cette partie a pour but d'essayer de construire un modèle mathématique mettant en relation la variation de chiffre d'affaires provoquée par le référencement d'un site Internet d'entreprise et les différents éléments composant ce référencement, puis d'étudier la validité effective de ce modèle dans la pratique. Je montrerais en fait que cette relation ne me semble modélisable ni dans le cas général ni dans le cas particulier des sites des clubs de sports français pris comme exemple de famille de sites.

Mon point de départ pour cela est l'hypothèse suivante : pour un site Internet d'entreprise, la variation de chiffre d'affaires due au référencement est directement liée à la variation de son audience, elle-même directement liée au positionnement de ce site dans les résultats des outils de recherche. Je m'appuie pour ce dernier point sur les différentes statistiques présentées dans ma première partie.

Je chercherai donc à exprimer la variation de chiffre d'affaires en fonction des différents éléments définissant le référencement du site Internet, à un coefficient multiplicatif près. Je passerai en revue les différents aspects analysés lors de la comparaison des techniques d'optimisation et chercherai à déterminer leur impact sur le positionnement et donc sur la variation de chiffre d'affaires induite, en traduisant cet impact par une fonction mathématique s'en approchant le plus possible.

Dans un deuxième temps, je chercherai à comparer ce modèle avec la réalité de nombreux sites Internet afin d'en prouver ou non sa validité dans le cas général, et, le cas échéant, de déterminer son périmètre de validité.

i. Formulation mathématique des différentes composantes

1. Présence dans les outils de recherche

La présence dans les outils de recherche est le premier paramètre à étudier. L'influence de celui-ci est liée à deux éléments distincts : le nombre d'outils de recherche dans lequel le site est inscrit, et l'audience respective de ces outils.

En classant ces derniers par ordre d'audience du plus important au plus faible, on obtient une courbe d'influence très fortement croissante au début, puis de moins en moins au fur et à mesure que le nombre d'outils augmente.

Cette courbe est du type $x \rightarrow \sqrt{x}$.

Cependant, il ne me paraît pas pertinent d'utiliser une telle fonction, en raison de la nécessité de classer les outils de recherche par ordre d'importance. Une solution est donc d'utiliser une fonction composée non explicite : la fonction audience cumulée des outils dans lequel le site est indexé, en pourcentage de l'audience totale des outils.

Une telle fonction prend ses valeurs dans l'intervalle $[0;1]$.

Appelons-la $Aud_{cumulée}(\text{outils})$.

2. Optimisation autour des mots-clés

L'optimisation autour des mots-clés est le deuxième facteur clé du référencement, comme nous l'avons déjà vu.

Il comprend le nombre de mots-clés sur lesquels repose l'optimisation, la pertinence de ceux-ci par rapport au sujet du site et surtout par rapport à l'objet et la cible de celui-ci, l'adéquation de ces mots-clés avec l'usage des internautes et enfin la concurrence existante sur ces mots-clés.

Nous avons déjà remarqué qu'une optimisation devait se faire autour de plusieurs mots-clés ou expressions pour être optimale, mais que cependant l'excès est évidemment à éviter, puisqu'un trop grand nombre de mots-clés diminuent la pertinence du site sur chacun d'entre eux.

Il est en général recommandé d'avoir entre un et dix mots-clés mis en relief, avec une valeur idéale de cinq. J'ai donc choisi de modéliser ceci par une courbe gaussienne atténuée et centrée sur 5, c'est-à-dire la fonction :

$$x \rightarrow c_1 \times e^{-\left(\frac{x-5}{3}\right)^2} \text{ où } c_1 > 0.$$

c_1 est un coefficient permettant de faire varier l'importance de cette fonction dans le modèle global. La partie gaussienne de la fonction peut être considérée comme une efficacité et varie de 0 à 1.

La pertinence des mots-clés a un effet positif et croissant sur le référencement et donc indirectement sur l'audience puis sur la variation de chiffre d'affaires induite d'après mon hypothèse de départ. Elle peut également avoir pour résultat d'annihiler toute optimisation.

En considérant simplement que plus les mots clés sont pertinents, plus l'optimisation est efficace et plus l'apport en trafic et en chiffre d'affaires est important, on obtiendrait une fonction linéaire de type $x \rightarrow a \times x + b$.

Cependant, ni le niveau de pertinence, ni son effet sur le positionnement ne peuvent croître à l'infini : il est nécessaire de prendre une fonction qui prend un argument borné et qui renvoie des valeurs également bornées.

Pour modéliser au mieux l'effet de la pertinence des mots-clés, une portion de courbe sinusoïdale variant entre 0 et 1 me paraît idéal : une courbe aplatie au début, puis de plus en plus fortement croissante jusqu'à un point d'inflexion se situant au niveau de la pertinence moyenne, puis ensuite s'aplatissant de plus en plus.

Une telle fonction se construit facilement : la fonction $x \rightarrow \frac{1 - \cos(x)}{2}$ pour $x \in [0; \pi]$ prend ses valeurs entre 0 et 1 et présente les caractéristiques voulues.

$x \in [0; \pi]$ est alors le degré de pertinence de la liste des mots-clés choisis, et peut donc être modélisé par une fonction $m \rightarrow Pert_1(m)$ qui prend une liste de mots en argument et renvoie un degré de pertinence compris entre 0 et Pi.

Cette fonction n'étant pas directement modélisable, je me contenterai de la laisser sous sa forme implicite.

La modélisation de l'effet de la pertinence des mots-clés est donc la suivante :

$$m \rightarrow \frac{1 - \cos(Pert_1(m))}{2} \text{ avec } m \text{ la liste des mots-clés.}$$

L'adéquation de cette liste de mots-clés avec les usages des internautes est comme nous l'avons déjà montré, tout aussi importante que la pertinence de ceux-ci.

L'effet de cette adéquation est similaire à celui de la pertinence et peut donc être modélisé de la même manière :

$$m \rightarrow \frac{1 - \cos(Pert_2(m))}{2}$$

avec $m \rightarrow Pert_2(m)$ représentant cette fois le degré d'adéquation de la liste de mots-clés avec les usages des internautes.

Le degré de concurrence des mots-clés est lié à la taille des autres acteurs présents, à leur niveau d'optimisation, mais aussi indirectement aux « scores » de ces mots-clés dans les outils de recherche (nombre de résultats proposés par ces outils en réponse à ces mots-clés).

Son effet est négatif : plus la concurrence est forte et plus l'effet d'une optimisation est atténué, et au contraire, dans le cas d'une absence de concurrence, la moindre optimisation suffira à avoir un résultat très positif.

La fonction inverse $x \rightarrow \frac{1}{x}$ correspond parfaitement à cette définition.

Afin de lisser cette fonction et de correspondre mieux à la réalité de l'effet de la concurrence, j'ai choisi d'affecter un coefficient $c_2 > 1$ à cette fonction. On obtient donc la modélisation suivante :

$$\frac{c_2}{d_{concurrence}(m)} \text{ avec } c_2 > 1 \text{ et } d_{concurrence}(m) \text{ le degré de concurrence des mots-clés.}$$

Ces différents éléments multipliés entre eux permettent d'obtenir la modélisation de l'effet de l'optimisation autour des mots-clés sur la variation de chiffre d'affaires (à un coefficient multiplicatif près) :

$$\frac{c_1 \times c_2 \times (1 - \cos(Pert_1(m))) \times (1 - \cos(Pert_2(m)))}{4 \times d_{concurrence}(m)} \times e^{-\left(\frac{n-5}{3}\right)^2}$$

avec $c_1 > 0$, $c_2 > 1$, n le nombre de mots-clés sur lesquels est basée l'optimisation, $Pert_1$, $Pert_2$, et $d_{concurrence}$ les fonctions attribuant respectivement à ces mots-clés leur degré de pertinence, d'adéquation et de concurrence.

3. Optimisation de l'organisation interne

L'optimisation de l'organisation interne repose principalement sur le nombre de pages finalement indexées par les moteurs de recherche et par l'organisation des liens entre ces pages.

Le nombre de pages a un effet positif, proportionnel et pouvant aller à l'infini. Par contre, les liens internes pondèrent ce nombre de pages (à travers le PageRank de Google notamment), et ne peuvent pas l'amplifier. Ce paramètre doit donc être modélisé par une fonction d'efficacité prenant ses valeurs entre 0 et 1.

Celle-ci ne pouvant être modélisé explicitement, nous la désignerons par $Eff_1(liens_{internes})$. Nous obtenons donc la modélisation suivante : $n_{indexées} \times Eff_1(liens_{internes})$.

4. Optimisation de la structure de fichiers

L'optimisation des fichiers repose à la fois sur la pertinence des noms de fichiers et de domaines, et sur l'organisation même de ces fichiers (structure de répertoires).

Il s'agit en fait d'un coefficient pondérateur de l'organisation interne dont l'optimisation peut voir son effet réduit par une mauvaise optimisation au niveau des fichiers.

J'utiliserai donc là encore une fonction efficacité à valeurs dans $[0;1]$: $Eff_2(fichiers)$.

5. Optimisation du code

L'optimisation du code se fait autour de deux éléments distincts. Les balises html ont pour rôle de rendre la mise en forme pertinente avec le contenu du site, et la prise en compte des frames et des pages dynamiques de rendre toutes les pages accessibles par tous les moteurs.

Ce deuxième élément est en fait déjà pris en compte au travers du nombre de pages effectivement indexées, et il ne reste donc plus qu'à modéliser l'effet de la pertinence de la forme par rapport au fond.

Il s'agit à nouveau d'une fonction efficacité : $Eff_3(fond, forme)$ qui prend ses valeurs dans $[0;1]$ et pondère l'effet de l'organisation interne.

6. Liens externes

Les liens externes ont un effet aussi important que le nombre de pages indexées et cet effet s'articule autour de trois points : le nombre de liens, la force de ces liens (importance de la page source) et la pertinence de ces liens (rapport entre le sujet source et le sujet du site).

Le nombre de liens a un effet positif et potentiellement infini, par contre la force et la pertinence des liens ont un effet plus similaire à la pertinence des mots-clés, et sera par conséquent modélisé de manière identique :

$$\frac{1 - \cos(Pert_3(\text{liens}_{\text{externes}}))}{2}$$

avec $Pert_3$ la fonction degré de force et de pertinence des liens externes.

L'effet des liens externes est donc modélisé par : $n_{\text{liens}_{\text{externes}}} \times \frac{1 - \cos(Pert_3(\text{liens}_{\text{externes}}))}{2}$.

7. Pages satellites et cloaking

La mise en place de pages satellites et l'utilisation du cloaking ont parfois un effet positif sur le référencement d'un site Internet, mais dans la plupart des cas cet effet est négatif avec notamment la possibilité de bannissement du site par les outils de recherche.

L'effet est donc aléatoire, limité dans le positif et infini dans le négatif.

En modélisant par le coefficient $c_3 > 1$ l'effet positif maximum possible, j'ai choisi de modéliser l'effet des pages satellites et du cloaking par la fonction :

$$x \rightarrow c_3 - \frac{1}{\text{rand}(x)}$$

avec $x \rightarrow \text{rand}(x)$ la fonction aléatoire à valeurs dans $[0;1]$.

ii. *Formulation du modèle mathématique théorique*

Nous avons déjà vu que les fonctions représentant les effets de l'optimisation des fichiers et du code pondèrent celle de l'organisation interne ; de la même façon, les fonctions modélisant la présence dans les outils de recherche et le nombre de mots-clés sur lesquels repose l'optimisation ont une influence des éléments d'optimisation du référencement sur la variation du chiffre d'affaires sous la forme :

$$\text{présence} \times \text{motsclés} \times \left(\text{org}_{\text{Interne}} \times \text{org}_{\text{fichiers}} \times \text{optim}_{\text{code}} + \text{liens}_{\text{externes}} + \text{satellites \& cloaking} \right)$$

En remplaçant les différents éléments de la formule précédente par leurs modélisations respectives, on obtient le modèle suivant :

$$\Delta_{CA} = K \times Aud_{cumulée}(outils) \times \frac{c_1 \times c_2 \times (1 - \cos(Pert_1(m))) \times (1 - \cos(Pert_2(m))) \times e^{-\left(\frac{n-5}{3}\right)^2}}{4 \times d_{concurrence}(m)} \\ \times [n_{indexées} \times Eff_1(liens_{internes}) \times Eff_2(fichiers) \times Eff_3(fond, forme) \\ + n_{liens_{externes}} \times \left. \frac{1 - \cos(Pert_3(liens_{externes}))}{2} + c_3 - \frac{1}{rand(x)} \right]$$

où : $K > 0$, $c_1 > 0$, $c_2 > 1$ et $c_3 > 1$ coefficients multiplicatifs

n nombre de mots - clés

m liste des mots - clés

$n_{indexées}$ nombre de pages effectivement indexées

$n_{liens_{externes}}$ nombre de liens externes

$Pert_1$, $Pert_2$, $Pert_3$ fonctions pertinence prenant en arguments une liste d'éléments

(mots - clés ou liens) et retournant un degré de pertinence entre 0 et Pi

$d_{concurrence}$ fonction prenant en argument une liste de mots - clés et retournant un degré de concurrence entre 0 et $+\infty$

Eff_1 , Eff_2 , Eff_3 fonctions efficacité prenant en argument des listes d'éléments et calculant leur degré d'efficacité entre 0 et 1

$rand$ fonction aléatoire à valeurs dans $[0;1]$

iii. Validation du modèle par des cas concrets

Ce modèle mathématique ainsi créé est construit à partir de mes propres analyses et comparaisons des différentes techniques d'optimisation du référencement. Il s'agit donc d'un modèle relativement théorique qu'il reste à confronter avec la réalité de sites Internet concrets.

Pour ce faire, j'ai interrogé de nombreuses sociétés dont le seul point commun est d'avoir à l'heure actuelle un site Internet, dans le but de connaître leur attitude face au référencement et le résultat de celle-ci, tant en terme de visites qu'en terme de variation de chiffre d'affaires.

Une vingtaine de sociétés ayant accepté de me répondre, j'ai ensuite analysé chaque site de ces sociétés, afin de comparer, corriger et compléter leurs dires concernant les optimisations effectuées et surtout visibles.

En effet, de nombreuses sociétés m'ont indiqué avoir faire appel à un référenceur pour optimiser leur site en vue du référencement, alors qu'aucune trace d'une quelconque réelle

optimisation (à travers le site comme à travers les résultats dans les outils de recherche de la soi-disant optimisation) n'est visible.

Cependant, mon objectif à travers ces études de cas réels n'est pas de juger de la qualité du référencement, ni de la véracité ou de la réelle application des dires des prestataires de référencement ayant vendu leurs services à ces sociétés, mais uniquement de comparer l'impact des différents points-clés du référencement sur la variation d'audience et de chiffre d'affaires, avec le résultat de mon modèle mathématique.

Dans un premier temps, j'ai choisi d'étudier des sites de sociétés très diverses, de tailles et de poids économiques différents, et surtout ayant des objectifs divers : site marchand avec une boutique en ligne, site offrant un service gratuit, site servant de vitrine institutionnelle ou de plaquette de communication, etc. ...

Pour chacun de ces sites, j'ai complété les témoignages reçus par une analyse :

- du nombre d'outils de recherche sur lesquels le site est positionné ;
- du nombre de mots-clés sur lesquels le site semble être optimisé (notamment grâce à une analyse par densité des mots-clés sur les pages les plus importantes) ;
- de la pertinence de ces mots-clés et de leur adéquation avec les usages des internautes (en comparant avec le contenu du site, avec les objectifs des responsables, ou encore avec les statistiques de Google, d'Overture et d'Espotting concernant les usages) ;
- de la concurrence de ces mots-clés (en étudiant le nombre de résultats proposés par les outils de recherche et le niveau d'optimisation des concurrents) ;
- du nombre de pages réellement indexées (principalement sur Google, Aol et Lycos) ;
- de l'organisation interne et de la structure de fichiers (à travers le site Internet lui-même) ;
- de l'optimisation du code (à travers l'étude de la source des pages principales du site) ;
- du nombre de liens externes pointant vers le site (principalement sur Google et Lycos) ;
- de la valeur de ces liens externes (en les examinant dans les bases de résultats de Google et Lycos) ;
- et de la présence de pages satellites et de cloaking.

Les résultats de ces analyses combinés aux différents témoignages des personnes contactées sont résumés dans le tableau ci-après.

A la demande de ces personnes et afin de préserver la confidentialité des éléments fournis, le nom des sociétés et des sites Internet n'apparaissent que dans mes remerciements et ne sont pas directement reliés aux analyses des sites.

N°	Prise en compte du référencement				Audience cumulée des outils	Nombre de mots-clés	Pertinence	Adéquation	Concurrence	Nombre de pages indexées	Organisation interne	Organisation fichiers	Optimisation code	Nombre de liens externes	Valeur des liens externes	Satellite & cloaking	Résultat du modèle	Augmentation trafic	Augmentation CA
	Prestataire extérieur	Prestataire créateur	Budget (€)	intégré															
1	Oui	Non	Non	intégré	97%	5	95%	50%	2000	29000	85%	60%	80%	190	75%	0	289	positif	100%
2	Oui	Oui	Non	6 860	97%	1	30%	10%	0,1	60	40%	75%	50%	3	20%	0	8	N/A	100%
3	Oui	Oui	Non	intégré	77%	3	45%	75%	400	40	30%	75%	75%	1	60%	0	0	0%	0%
4	Oui	Oui	Non	15 000	97%	1	50%	75%	5300	20	85%	35%	65%	20	65%	0	0	N/A	10%
5	Oui	Oui	Non	15 000	82%	3	80%	80%	350	380	85%	40%	65%	5	50%	0	11	N/A	3%
6	Oui	Oui	Non	50 000	97%	3	65%	80%	80	3350	85%	40%	75%	5000	80%	0	2 749	30%	0%
7	Oui	Oui	Non	5 000	75%	6	90%	95%	25	51	75%	60%	65%	1	20%	0	39	N/A	15 prospects/mois
8	Oui	Oui	Non	5 000	97%	3	75%	95%	15	80	75%	80%	50%	2	60%	0	89	N/A	4 prospects/mois
9	Non	Non	Non	0	75%	1	50%	50%	0,5	23	75%	80%	25%	0	0%	0	22	0%	0%
10	Non	Non	Non	0	97%	2	65%	75%	474	23	65%	75%	50%	8	60%	0	1	0%	0%
11	Oui	Oui	Non	6 860	83%	3	65%	80%	200	20	70%	75%	75%	33	18%	1	0	200%	0%
12	Oui	Non	Non	0	97%	4	80%	90%	60	626	75%	45%	75%	300	25%	0	258	0%	0%

Le calcul du résultat du modèle mathématique a été fait en attribuant les valeurs suivantes aux constantes : $K = 1$, $c_1 = 1$, $c_2 = 100$, et $c_3 = 1$; et en multipliant par π les pourcentages correspondant au degré de pertinence et d'adéquation des mots-clés, et à la valeur des liens externes.

La première vérification que j'ai effectuée a été de comparer les résultats du modèle avec le positionnement des sites dans les résultats des outils de recherche.

Deux constatations sont alors possibles :

- tout d'abord, les résultats sont cohérents par rapport aux positionnement, puisque plus le résultat est important et plus le site est présent et bien positionné dans les outils de recherche ;
- ensuite, quatre niveaux de sites se distinguent :
 - o les sites ayant un résultat nul sont présents mais très mal positionnés dans les résultats des outils de recherche, lors d'une recherche par le nom du site ;
 - o les sites ayant un résultat compris entre 1 et 100 sont positionnés dans les 5 premiers résultats des principaux outils lors d'une recherche par le nom du site et apparaissent quelques fois dans les 30 premiers résultats sur des mots-clés liés au contenu ;
 - o les sites ayant un résultat supérieur à 100 sont positionnés dans la première page de résultats lors de recherche sur plusieurs mots-clés liés au contenu, et dans les 3 premières positions lors d'une recherche sur le nom du site.

Ces remarques montrent la cohérence des rapports des effets des différents points étudiés les uns par rapport aux autres avec les résultats réellement observables de ces effets dans les résultats des moteurs de recherche.

Les sites n°1, 2 et 12 sont tous les trois des sites marchands permettant d'acheter en ligne. De plus, tous trois ont prévu très tôt l'optimisation de leur référencement puisque le n°1 y a travaillé avant même la création du site, le n°12 juste après la création, et le n°2 à peine trois mois après.

Pourtant, si l'on compare les résultats du modèle avec les résultats réels donnés par les responsables des sites en question, on s'aperçoit qu'il n'y a aucun rapport entre eux : deux scores de plus de 250 d'après le modèle n'ont pas du tout le même impact au niveau du trafic, ni au niveau du chiffre d'affaires ; et un score de 8 pour le modèle a un impact similaire à un score de 290 en terme de chiffre d'affaires. Il n'est donc pas possible de valider le modèle mathématique avec ces éléments.

De même, au vu des disparités existant entre ces trois sites au niveau de l'impact de leur positionnement dans les résultats des outils de recherche sur leur trafic et leur chiffre d'affaires, il ne semble pas possible d'établir de relation simple entre ces trois éléments.

Les sites n°4, 5 et 6 ne proposent pas de boutique en ligne directement, mais proposent aux internautes un service gratuit dans le but de faire vendre indirectement les produits non électroniques des sociétés éditrices de ces sites.

Là encore, il n'y a pas de rapport entre les différents résultats du modèle et les résultats donnés lors des interviews : le site ayant le plus haut score du modèle a vu son trafic augmenter de 30 % mais n'a pas eu d'impact sur son chiffre d'affaire, alors qu'un site ayant un score de 11 a gagné 3 % de chiffre d'affaires et un site ayant un score nul de 10 %.

Pourtant, le positionnement de ces sites correspond totalement aux scores donnés par le modèle !

Enfin, les autres sites (n°3, 7, 8, 9, 10 et 11) ne proposent ni boutique en ligne ni service gratuit, mais uniquement des informations sur la société concernée. Ce sont des sites « institutionnels » ou « plaquettes » comme ils se désignent eux-mêmes.

Les seuls résultats communs à ces sites sont leur score inférieur à 100 et leur positionnement très faible dans les différents outils de recherche. La variation de trafic varie, elle, de 0 à 200 % et le chiffre d'affaire de « aucun apport » à « 15 prospects par mois », sans d'ailleurs que le maximum de l'un soit le maximum de l'autre, bien au contraire.

Si le modèle paraissait cohérent avec le positionnement des sites dans les outils de recherche, il ne l'est absolument pas avec la variation de chiffre d'affaires due à ce référencement ou encore avec l'augmentation du trafic due à celui-ci. De même, il ne semble y avoir aucune règle générale liant cette augmentation de trafic à celle du chiffre d'affaires..

Il convient également de noter qu'il n'y a pas non plus de rapport entre le budget investi dans le référencement et les résultats de celui-ci sur le positionnement, le trafic ou le chiffre d'affaires : si le plus gros budget correspond effectivement au meilleur score du modèle et au meilleur positionnement dans les résultats des outils de recherche, les deux meilleurs scores et positionnements suivant sont obtenus par deux sociétés qui ont optimisé leur référencement en interne et pour un budget nul ou presque.

De plus, le plus gros budget investi correspond à une augmentation positive mais non exceptionnelle de l'audience et surtout à aucune augmentation du chiffre d'affaires. Enfin, il faut remarquer que sauf pour un ou deux cas, les sites ayant fait appel à des prestataires extérieurs obtiennent les moins bons scores du modèle, et les moins bons positionnements dans les résultats des différents outils de recherche.

Cette analyse confirme donc les conclusions de l'étude des tarifs et prestations des référenceurs.

La diversité des sites Internet comparés pouvant être une des causes de cette absence de règle générale, j'ai choisi de renouveler cette analyse sur un secteur précis, et j'ai choisi pour cela le secteur des clubs de sport français (football et rugby principalement).

La méthodologie appliquée pour cette seconde analyse est rigoureusement identique à la première et les résultats sont représentés dans le tableau suivant :

N°	Prise en compte du référencement				Audience cumulée des outils	Nombre de mots-clés	Pertinence	Adéquation	Concurrence	Nombre de pages indexées	Organisation interne	Organisation fichiers	Optimisation code	Nombre de liens externes	Valeur des liens externes	Satellite & cloaking	Résultat du modèle	Augmentation trafic	Augmentation CA
	Prestataire extérieur	Prestataire créateur	Budget (€)																
1	Oui	Non	Non	0	97%	2	75%	80%	800	4800	85%	75%	75%	700	95%	0	103	0%	0%
2	Oui	Oui	Oui	intégré	97%	1	90%	90%	3	2	30%	0%	0%	21	95%	0	109	5900%	0%
3	Oui	Oui	Oui	intégré	97%	2	70%	75%	3	100	70%	70%	80%	28	90%	0	536	10%	0%
4	Non	Non	Non	0	97%	1	85%	75%	77	15300	80%	45%	55%	800	95%	0	657	0%	0%
5	Oui	Oui	Oui	intégré	92%	3	90%	95%	2	60	55%	30%	25%	60	90%	0	1 745	15%	0%
6	Oui	Non	Non	0	97%	2	90%	90%	150	1500	45%	10%	60%	2000	95%	0	459	5%	0%
7	Oui	Oui	Oui	intégré	97%	3	95%	95%	30	2300	85%	35%	65%	200	95%	0	1 318	limité	considérable
8	Oui	Oui	Oui	30 000	90%	3	92%	90%	15	140	85%	30%	75%	120	95%	0	539	au mieux	0%

Pour les 8 sites analysés, le résultat du modèle est supérieur à 100, et on vérifie à nouveau que les sites sont positionnés dans la première page de résultats lors de recherche sur plusieurs mots-clés liés au contenu, et dans les 3 premières positions lors d'une recherche sur le nom du site (du club) correspondant.

La cohérence du modèle avec le positionnement dans les outils de recherche est donc là encore vérifiée.

Les cibles respectives des sites sont tout à fait similaires, et les différents contenus suivent à peu près les mêmes modèles. Cinq de ces sites proposent une boutique en ligne permettant d'acheter des objets à l'effigie des clubs concernés, et sept sites proposent une billetterie.

Pourtant, les résultats du modèle comme ceux annoncés par les clubs eux-mêmes sont extrêmement variés et ne semblent suivre aucune logique.

Le deuxième score le plus faible du modèle correspond à une augmentation de 5000 % du trafic du site due au référencement de celui-ci, mais à aucune augmentation de chiffre d'affaires malgré la présence à la fois d'une boutique en ligne et d'une billetterie.

A l'opposé, un site présentant les mêmes fonctionnalités et un contenu similaire obtient le deuxième plus gros score du modèle, pour une augmentation « limitée » du trafic et un impact « considérable » sur le chiffre d'affaires.

La seule véritable différence entre ces deux sites est dans l'optimisation du second qui est basée entre autre sur un mot-clé correspondant à l'article le plus recherché de la boutique en ligne. L'adéquation entre l'optimisation du site et les attentes ou recherches réelles des internautes ciblés semble donc être l'explication de cette différence.

De même, le site obtenant le meilleur score du modèle, qui correspond effectivement au meilleur positionnement dans les outils de recherche, a vu son trafic n'augmenter que de 15% et n'a bénéficié d'aucune hausse de chiffre d'affaires due à son référencement ; et le troisième score n'a bénéficié ni d'une hausse de trafic, ni de chiffres d'affaires !

Même dans le cas d'un secteur très précis avec des sites Internet extrêmement similaires, il apparaît qu'il n'existe aucune relation directe entre les différents éléments d'optimisation et la variation de chiffre d'affaires, ou même d'audience due au référencement.

Néanmoins, les résultats du modèle mathématique créé à partir des analyses et comparaisons de ces techniques concordent avec le positionnement réel des sites dans les résultats des outils de recherche, c'est à dire avec leur visibilité à travers ces outils. C'est donc bien la transformation de cette visibilité en audience ou en chiffre d'affaires qui n'est pas modélisable en fonction de ces techniques.

Cette transformation n'est en effet pas seulement basée sur une optimisation interne du site, mais plus sur l'adéquation entre le contenu (produits vendus, service proposé ou informations publiées) et les attentes réelles de la cible visée, attentes sur lesquelles ne peut agir directement l'entreprise éditrice du site, et qui ne sont par conséquent pas modélisables.

Ce modèle mathématique permet d'établir un jugement fiable sur l'état du référencement d'un site Internet, et surtout sur le positionnement de celui-ci dans les principaux outils de recherche.

Concernant l'effet des différents éléments du référencement sur la variation de chiffre d'affaires, il apparaît que celui-ci n'est pas modélisable pour le cas général ou pour le cas particulier des sites de clubs de sport français.

Dans le cas d'autres familles de sites Internet comme celui des sites de vente en ligne par exemple, la question de la validité de ce modèle reste entière.

La question de l'existence de familles de sites Internet, comme celui des sites de vente en ligne par exemple, pour lesquelles ce modèle est valide ou facilement modifiable pour l'être, reste entière.

Conclusion

Nous avons posé au départ deux questions :

- est-il nécessaire pour un site web d'effectuer un travail particulier sur le référencement ?
- Quel est ce travail et qui doit l'effectuer ?

L'analyse de quelques statistiques, et surtout du comportement des internautes, montre qu'il est indispensable, pour une entreprise souhaitant être présente sur Internet à travers un site Web, d'optimiser le référencement de ce site.

De même, l'analyse des principales techniques existantes à ce jour montre qu'il n'est pas forcément nécessaire de faire appel à un prestataire extérieur pour ce travail. Dans les cas où une externalisation est utile, les professionnels du référencement ne sont pas obligatoirement les prestataires les plus appropriés.

En résumé, deux cas sont possibles : celui d'un site entièrement à créer et celui d'un site déjà existant.

Dans le cas d'un site encore à créer, c'est l'ensemble de la création qui doit se faire en gardant à l'esprit l'objectif d'optimisation du référencement.

La première étape est l'analyse du but réel du site, des mots-clés associés et des sites des concurrents. Aucune connaissance particulière n'est nécessaire, mais seulement un investissement en temps.

Viennent alors les choix stratégiques sur la forme à utiliser (design, présence de frames, de pages dynamiques, etc. ...). Ces choix doivent être faits en connaissance des implications en terme de référencement : il suffit de veiller à ce que l'agence de création les connaisse, voire de les connaître soi-même.

Les optimisations des liens internes, de la structure de fichiers et du code interviennent lors de la création même des pages du site en respectant les règles d'écritures nécessaires et en évitant toutes données inutiles (titre, description, mots-clés, etc. ...). Là encore, il suffit de vérifier le respect de ces règles par l'agence de création.

La soumission du site ne nécessite ni connaissance autre que la liste des principaux outils de recherche, ni appel à un quelconque spécialiste.

La diffusion des liens externes est à faire également en interne. Cette étape rejoint par sa nature l'activité de publicité en ligne et peut être effectuée par le département marketing ou communication.

Enfin, le suivi nécessaire de son propre référencement et de celui des principaux concurrents doit se faire en même temps et par les mêmes personnes que la maintenance du site, et donc généralement par le service informatique.

Dans le cas d'un site déjà existant, la volonté d'optimisation du référencement peut nécessiter la refonte complète de celui-ci.

La première étape est de re-identifier l'objectif du site, puis d'analyser l'état actuel du référencement de celui-ci et des concurrents, afin d'établir la liste des travaux à effectuer. Cette étape nécessite seulement du temps et la connaissance des règles de base du référencement, voire des outils gratuits disponibles.

Si les modifications à apporter sont trop nombreuses, il est préférable de recréer une nouvelle version du site et on se place ainsi dans le cas de la création.

Optimiser la structure de fichiers et les liens internes implique de refaire l'ensemble des liens : cela nécessite des compétences en création de sites web et la connaissance des quelques règles d'écritures à respecter en vue du référencement.

De même, les modifications du code et du texte ne nécessitent pas non plus l'intervention d'un spécialiste du référencement : un référenceur pourrait faire ce travail, mais sa non connaissance du site à modifier rend le travail plus long et donc plus cher que pour l'agence de création qui l'a créé.

Enfin, le travail de vérification et de mise à jour des liens externes demande également du temps et la connaissance des règles de bases les concernant, mais aucune autre qualification particulière n'est nécessaire.

Les connaissances nécessaires à l'optimisation du référencement sont donc les mêmes dans les deux cas, et se limitent en fait aux règles de base décrites et analysées dans ce document. Ces règles de base du référencement peuvent être résumées en deux idées principales :

- un contenu riche, complet et pertinent ;
- un site accessible par tous les navigateurs même anciens (surtout anciens).

Il est inutile de faire appel à un professionnel du référencement pour des aspects techniques que les agences de création maîtrisent totalement.

L'analyse des tarifs des référenceurs rapportées aux prestations proposées confirme les conclusions de l'analyse technique s'il en était besoin.

Enfin, les résultats du modèle mathématique créé montrent qu'il n'existe aucun retour sur investissement garanti lors du passage par un professionnel du référencement. La solution la plus économique et la plus sûre est donc bien d'internaliser ce travail.

L'intérêt des référenceurs est dans l'apprentissage des règles de base, voire dans leur expérience de pertinence du contenu : il s'agit plus d'une activité de conseil, et notamment de conseil en contenu.

Une évolution logique de la profession pourrait être l'abandon de la partie « programmation effective » au profit des agences de créations, et la transformation en « agence de contenu pertinent »...