

Le référencement

Google™

TFE présenté par **Nicolas De Boose** pour l'obtention du
« Diplôme d'Etudes spécialisées en Marketing et Communications »
Année académique 2013 - 2014

Je voudrais surtout remercier ma future épouse, Laura, d'avoir dit « oui » et de m'avoir soutenu durant la rédaction de ce travail de fin d'études. Je la remercie aussi pour la relecture.

Je remercie également :

Ma famille, ainsi que ma belle-famille, de m'avoir encouragé.

L'ensemble des formateurs de l'ICHEC qui m'ont appris énormément de choses.

Les étudiants du cursus 2013-2014 avec qui je me suis bien amusé. Sans eux, l'année aurait semblé un peu plus longue.

Table des matières

1	Introduction.....	5
1.1	SEA.....	5
1.2	SEO	5
2	Fonctionnement	7
2.1	Attitudes de l'internaute	7
2.2	Attitudes du webmaster.....	8
2.3	Fonctionnement des moteurs de recherche.....	8
2.3.1	Collecte d'informations	8
2.3.2	Indexation.....	9
2.3.3	Pertinence	9
2.3.4	SERP : Affichage des résultats	10
2.4	Un mot sur les annuaires.....	11
3	Préparation du référencement	13
3.1	Longue et courte traîne	13
3.2	Google suggest	14
3.3	Choix des mots-clés	14
3.4	Analyse des mots-clés	15
3.5	Gestion des évènements et référencement prédictif	15
3.6	Où se référencer ?	16
4	Optimisation des pages du site « in page ».....	17
4.1	Zone chaude 1 : la balise « title »	17
4.2	Zone chaude 2 : Le texte visible	18
4.3	Zone chaude 3 : L'adresse des pages	19
4.3.1	Nom de domaine	19
4.3.2	Nom des dossiers/pages.....	20
4.4	Zone chaude 4 : Les balises « meta »	20
4.4.1	Keywords	20
4.4.2	Description	20
4.4.3	Schema.org.....	21
4.5	Zone chaude 5 : L'attribut « alt »	21
5	Optimisation des pages du site : Critères « off page ».....	22
5.1	PageRank	22
5.1.1	Mode de calcul	23
5.2	Echange de lien.....	24
5.3	Paid linking	24

5.4	Link baiting	25
5.5	Les liens sortants	25
5.6	TrustRank.....	26
5.7	Temps de chargement.....	26
6	Référencement multimédia, multi support	27
6.1	Image.....	27
6.1.1	Le nom de l'image	27
6.1.2	L'attribut « alt ».....	27
6.1.3	Le texte autour de l'image.....	27
6.1.4	Le texte du lien	27
6.2	Vidéos.....	28
6.2.1	Meta données.....	28
6.2.2	Caractéristiques techniques.....	28
6.2.3	Les tags	28
6.2.4	Le texte du lien	29
6.2.5	Le texte autour de la vidéo.....	29
6.2.6	Indexage	29
6.3	Fichier word, powerpoint et pdf	29
6.4	Le référencement local.....	30
7	Technologies à éviter.....	31
7.1	Les frames.....	31
7.2	Le flash.....	31
7.3	Le javascript.....	31
7.3.1	Ajax	31
7.4	Formulaire	32
7.5	URL paramétrée	32
7.6	Cookie & session.....	32
8	Conseils à suivre	33
8.1	Redirection	33
8.2	HTTPS.....	33
8.3	Duplicate content	33
8.4	Multi langue	34
9	Accélérer le référencement d'un site.....	36
9.1	Soumission manuelle d'un site.....	36
9.2	Optimiser le temps d'indexation.....	36
9.3	Plan du site	36

9.4	Sitemap.xml.....	36
10	Mauvaises pratiques.....	38
10.1	Techniques à éviter	38
10.2	Détection	38
10.3	Types de pénalité	38
10.4	Sortir d'une pénalité.....	39
11	Instructions aux <i>spiders</i>	40
11.1	Robots.txt	40
11.1.1	Sitemap.....	40
11.1.2	User-agent	40
11.1.3	Disallow/Allow.....	40
11.2	Balise « meta »	40
12	Récapitulatif et derniers conseils	41
12.1	Conception	41
12.2	Code.....	41
12.3	Contenu	42
13	Assurer un suivi	43
14	Conclusion	44
15	Bibliographie.....	45

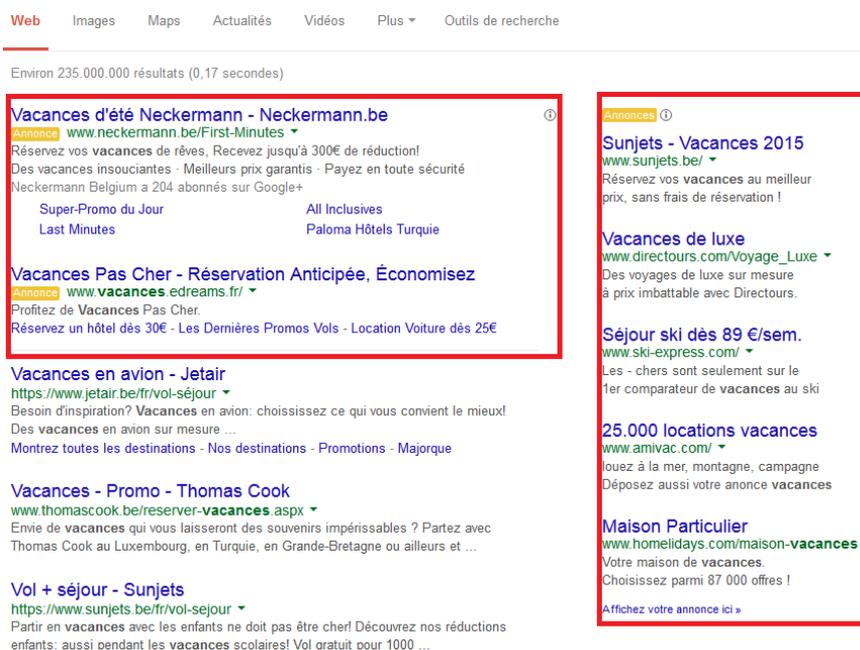
1 Introduction

« Je veux être tout en haut de *Google* ». Voilà une exigence que les agences web entendent de plus en plus souvent lors de la création d'un site internet. Comme pour tout produit dans un supermarché, y être référencé est une chose, mais y être bien positionné en est une autre. Afficher son site en tête de gondole chez *Google* est le rêve de chaque client, cela demande un budget et/ou du temps.

Il existe deux moyens de se situer en première place : Via le *SEA* ou le *SEO*.

1.1 SEA

Le *SEA*, pour *Search Engine Advertising*, est un système de liens sponsorisés. *Google* ainsi que tous les autres moteurs de recherche ont une zone publicitaire réservée aux annonceurs. En règle générale cette zone est idéalement située, c'est-à-dire au-dessus des résultats de la recherche ou sur la partie de droite.



Les zones SEA en rouge

Pour *Google*, les annonceurs s'inscrivent sur « *Adwords* »¹. L'interface leur permet d'ajouter des annonces et d'y lier des mots-clés. Une annonce s'affiche lorsque ses mots-clés correspondent à ceux insérés lors de la recherche. Le grand avantage est que l'annonceur ne paye pas aux nombres de vues, mais uniquement au nombre de clics.

Le *SEA* est l'équivalent du spot publicitaire ou tout autre action qui engendre des coûts directs comme : Se situer à hauteur des yeux dans les rayons, se trouver dans le flyer d'un supermarché, etc... Il n'a qu'un impact à court terme car une fois l'annonce arrivée à échéance, la visibilité sera nulle.

Il est très important de préciser que **le SEA n'influence en rien le SEO.**

1.2 SEO

Le *SEO*, pour *Search Engine optimisation*, correspond au référencement naturel (aussi appelé référencement organique) d'un site internet. Il entend toutes optimisations et pratiques nécessaires pour obtenir la meilleure visibilité sur un moteur de recherche. Le but étant de retrouver le site

¹ <http://www.google.com/adwords>

internet le plus haut possible dans la page des résultats. Contrairement au *SEA*, il est plus difficile de mesurer les coûts car une optimisation nécessite des moyens plus difficile à quantifier : Le temps.

Cet ouvrage abordera exclusivement le *SEO*, c'est-à-dire la partie qui pourra, au final, faire la différence par rapport aux concurrents.

2 Fonctionnement

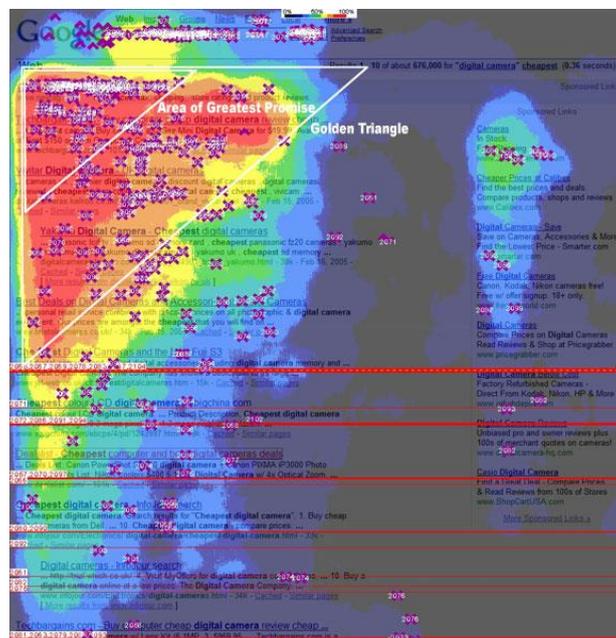
2.1 Attitudes de l'internaute

Pour comprendre ce que l'on entend par « visibilité », il est d'abord nécessaire de connaître les habitudes du chercheur.

L'internaute est, par définition, quelqu'un de paresseux. S'il ne trouve pas son bonheur sur la première page des résultats, ou pire, dans les 10 premiers liens, son reflex sera de modifier sa recherche. Il faut savoir que 60 à 80% des clics se font sur les 3 premiers liens², il est dès lors primordial de se trouver au plus haut de la page de résultats, appelée aussi *SERP*, pour *Search engine result page*.

L'idéal est que l'internaute n'ait pas à faire défiler la page des résultats. Un site se trouve sous la **ligne de flottaison**³ lorsqu'il n'est visible qu'après un défilement de la page. En règle générale, 5 sites sont affichés au-dessus de cette ligne mais, nous le verrons plus tard, ce n'est pas toujours le cas.

On évoque également la présence d'un triangle d'or qui symbolise les 3 premiers résultats. Selon plusieurs études⁴ d'« eye tracking », les 3 premiers liens de la *SERP* sont **vus** par 100% des internautes.



Le triangle d'or de Google : Où les internautes regardent-ils ?

² <http://www.abondance.com/actualites/20121012-11966-taux-de-clics-80-pour-les-3-premiers-resultats-naturels.html>

³ Calculé à partir d'une résolution de « 1024 x 768 ». En anglais : *above the fold*

⁴ <http://googleblog.blogspot.de/2009/02/eye-tracking-studies-more-than-meets.html> et <http://www.prweb.com/releases/2005/3/prweb213516.htm>

Par contre, cela diffère fort lorsque l'on parle de clics effectifs. En effet, le premier lien récolte 34% des clics contre respectivement 14 et 10% pour le 2^{ème} et 3^{ème} lien.



Le taux de clics en fonction de la position sur la page des résultats ⁵

2.2 Attitudes du webmaster

Nous venons de le voir, se placer premier apporte un réel avantage, mais se situer dans le top 3 reste un bon objectif. Évidemment, plus il y a de concurrents, plus il est difficile d'y arriver.

Il est possible de se placer assez facilement sur la première page si l'on décide de prendre des mots-clés rarement utilisés ou très ciblés. Une bonne stratégie est d'attaquer ces mots peu courants dans un premier temps, ce qui amènera assez peu de visiteurs.

Par contre, une fois que l'on touche à des mots forts demandés tels que « Bruxelles restaurant » ou « vacances pas cher », cela peut prendre (beaucoup) plus de temps. En effet, la plupart des sites actuels sont déjà bien optimisés pour ce genre de requête et il sera difficile de les détrôner. C'est, dès lors, une stratégie à moyen/long terme qu'il faut adopter ainsi que quelques campagnes SEA. La seule différence peut se faire sur les « bons » liens entrants sur le site, appelés aussi « *backlink* ».

Nous verrons qu'il existe plusieurs moyens d'optimiser son référencement. Ceux-ci sont répartis en deux catégories :

- *White hat* : Optimisation du code du site, du contenu de qualité et des partenariats.
- *Black hat* : Techniques interdites qui permettent de détourner les algorithmes des moteurs de recherche. Si elles sont détectées, les sanctions peuvent s'avérer graves : Rétrogradation du site, ou pire, sa suppression dans la base de données.

2.3 Fonctionnement des moteurs de recherche

2.3.1 Collecte d'informations

Les moteurs de recherche ont des outils, que l'on appelle « *spiders* » ou « *crawlers* », qui visitent (ou *crawlent*) le web en permanence. Ces araignées du web sont des visiteurs acharnés qui suivent tous les liens qu'ils trouvent. Ils sont assez intelligents que pour différencier le contenu de la charte graphique et reviennent régulièrement sur les sites pour vérifier une éventuelle mise-à-jour. Plus un site sera mis-à-jour, plus le *spider* passera sur les pages fréquemment. Un site d'actualité se verra donc « *crawlé* » plusieurs fois par jour⁶, contrairement à un site statique.

⁵ <http://marketingland.com/new-click-rate-study-google-organic-results-102149>

⁶ <https://www.mattcutts.com/blog/minty-fresh-indexing>

L'inconvénient du *spider* est qu'il ne visite pas de site n'ayant aucun lien entrant. Il n'interprète pas, non plus, le code *javascript*, ce qui veut dire que si un lien est généré dynamiquement, il ne le suivra pas.

2.3.2 Indexation

Une fois qu'une page est trouvée, le code source (en *HTML*) est indexé. Les mots trop communs (le, la, les, un, une, les, des, avec,...), appelés « *stop words* », sont ignorés afin de gagner en espace disque et en rapidité. On pourrait, dès lors, croire que « maison du monde » et « maison monde » renverraient le même résultat, mais il n'en n'est rien. En effet, *Google* retient où se situe les *stop words* qu'il a supprimé et sait reconnaître un résultat plus pertinent qu'un autre. Toutes les pages du web (approximativement 100 milliards) sont donc en base de données et indexées.

Il existe aussi un « index des termes » qui indique à quelle fréquence un mot revient sur chaque page. En fonction de l'endroit où se situe le mot, son poids sera également différent. En simplifiant, nous pourrions dire que cet index est celui utilisé lors d'une recherche : La page avec le plus d'occurrences (et beaucoup d'autres paramètres que nous verrons plus loin) sera celle située au-dessus des autres.

2.3.3 Pertinence

Comment est calculée la pertinence d'un document ? Aujourd'hui, seul *Google* le sait réellement. Comme Coca-Cola, le moteur de recherche garde secrètement sa recette (son algorithme), car c'est ce qui fait sa plus grande force. Nous allons rapidement parcourir les techniques de tri les plus courantes :

- Tri par pertinence :
 - o Où se situe le mot : Dans le titre ? En bas de page ?
 - o A quelle fréquence apparaît un mot par rapport au nombre total de mots?
 - o Le mot est-il mis en avant (en gras par exemple) ?
 - o Si une expression est recherchée, le moteur donnera priorité aux documents qui ont exactement la même expression plutôt que des mots séparés dans le document.
- Tri par popularité :
 - o La page avec le plus de liens entrants a sûrement plus d'importance que les autres.
- Tri par mesure d'audience (*Pogosticking*) :
 - o Les pages les plus cliquées depuis le moteur de recherche sont retenues.
 - o Le moteur de recherche enregistre également si un internaute revient rapidement sur la page de résultat. Si c'est le cas, cela veut dire que le site proposé n'était pas assez pertinent. Si, par contre, l'internaute ne revient plus, c'est qu'il a trouvé chaussure à son pied.
- Tri par catégories
 - o Classement automatique en dossiers à la manière d'un annuaire.

2.3.4 SERP : Affichage des résultats

2.3.4.1 Informations complémentaires

Dès 2012, Google a changé ses SERP et la ligne de flottaison est devenue imprévisible. En effet, le moteur de recherche peut, à tout moment, ajouter un emplacement appelé « one box » au-dessus des liens naturels.

The screenshot shows a Google search result for 'Météo Bruxelles'. At the top, there is a navigation bar with 'Web', 'Maps', 'Images', 'Actualités', 'Vidéos', 'Plus', and 'Outils de recherche'. Below this, it says 'Environ 23 200 000 résultats (0,26 secondes)'. The main content is a weather widget for 'Bruxelles' on 'vendredi 00:00' with 'Temps clair'. The temperature is 4°C | 4°F. Precipitation is 0% and humidity is 81%. Below the weather widget, there are search results for 'Météo bruxelles (BEL) - 1er site meteo pour Bruxelles et sa ...' and 'Météo Bruxelles - 14 jours'.

La ligne de flottaison est chamboulée pour la recherche « météo Bruxelles »

On le voit très clairement, l'emplacement réservé est de taille variable et même être premier dans les résultats ne garantit plus que l'on soit au-dessus de la ligne de flottaison. Quant au 4^{ème}, il pourrait se retrouver sur la seconde page, ce qui n'est pas négligeable.

Le *knowledge graph*⁷ est aussi apparu en 2012. Cette fonctionnalité a permis à Google d'afficher des informations vérifiées lors d'une recherche sur un sujet concret : Une célébrité, un bâtiment, un animal, etc... Ces données sont affichées sur la droite des résultats.

The screenshot shows a Google search result for 'Benjamin Franklin'. At the top, there is a navigation bar with 'Web', 'Images', 'Maps', 'Vidéos', 'Lignes', 'Plus', and 'Outils de recherche'. Below this, it says 'Environ 82 700 000 résultats (0,36 secondes)'. The main content is a knowledge panel for 'Benjamin Franklin' with a 'Plus d'images' button. Below the knowledge panel, there are search results for 'Benjamin Franklin - Wikipédia', 'Benjamin Franklin - L'Internaute', 'Benjamin Franklin - Biography - Writer, Inventor, Scientist...', and 'Quick Biography of Benjamin Franklin - USHistory.org'.

Recherche sur "Benjamin Franklin"

⁷ <http://www.google.com/insidesearch/features/search/knowledge.html>

2.3.4.2 Le second index de Google

Nous avons vu qu'il y avait un index qui contenait toutes les pages du web. Ce n'est pas tout à fait exact. En fait, *Google* range les pages qui lui sont considérées comme non-pertinentes dans un second index. Cet index ne sera jamais appelé, à moins de le lui indiquer explicitement, via le lien « afficher les pages ignorées ». Il y a donc tout intérêt à ne pas s'y retrouver.

Voici les principales raisons pour lesquelles une page peut se retrouver « ignorée » :

- La page ne comporte pas suffisamment de texte : il faut 200 mots minimum.
- Le contenu est une copie d'une autre page. On appelle cela du *duplicate content*. Nous verrons comment y échapper au chapitre 8.3 : Duplicate content.
- La page n'a aucun lien entrant. Toutes les pages doivent au moins être accessibles via une autre page. Il faut donc penser à ajouter des liens internes (appelés *deeplinking*) vers d'autres pages dans un article par exemple.

2.3.4.3 Sitelink

Le *sitelink*⁸ est l'espace que l'on peut apercevoir en dessous du premier lien du *SERP*. Il contient 4 à 6 liens additionnels.



Le sitelink du restaurant "S'eat"

Personne ne sait concrètement pourquoi et comment *Google* choisi d'afficher ces liens. Cependant nous pouvons nous faire une idée :

- Le premier site proposé doit être non-ambigu. En général, il correspond à 100% à ce que l'internaute cherche, comme le site officiel d'un chanteur, le site d'un restaurant, etc...
- Ce ne sont pas forcément les premiers liens du menu de navigation.
- Ce ne sont pas les liens qui ressortent dans l'ordre lorsque l'ont fait un recherche exclusive sur le site (avec « site : monsite.com »).
- Certains pensent que ce sont les liens les plus cliqués sur la page d'accueil d'un site. A priori, *Google* pourrait le savoir grâce à tous ses « mouchards » tels que *Google Analytics*, *Google Chrome*, *Google Toolbar*, etc...

2.4 Un mot sur les annuaires

En plus des moteurs de recherche, il existe encore quelques sites « annuaires ». L'annuaire est une hiérarchie dans lesquels se trouvent des sites rangés dans des catégories. Leurs informations sont très succinctes, puisqu'en général, pour un site, on ne retrouve que son adresse, son titre et une courte description. Pour trouver un site, il faut donc parcourir l'annuaire en naviguant de catégorie en sous-catégorie, afin de tomber sur une page contenant plusieurs sites ayant le même thème.

⁸ <https://support.google.com/webmasters/answer/47334?hl=fr>

Pour figurer dans cet annuaire, il faut remplir un formulaire. Une personne physique est chargée d'accepter ou refuser le site, ce qui alourdit encore plus le référencement et le rend très subjectif.

Les annuaires comme « dmoz.fr » ou « dir.yahoo.com » ont pratiquement été tués par les moteurs de recherche. *Yahoo!*, qui originellement un annuaire, en est un bon exemple. Depuis 2009, il sous-traite ses recherches à *Bing*, le moteur de recherche de *Microsoft*, et se concentre sur son portail et ses services.

Seuls quelques annuaires spécialisés ont survécus au tsunami *Google*.

3 Préparation du référencement

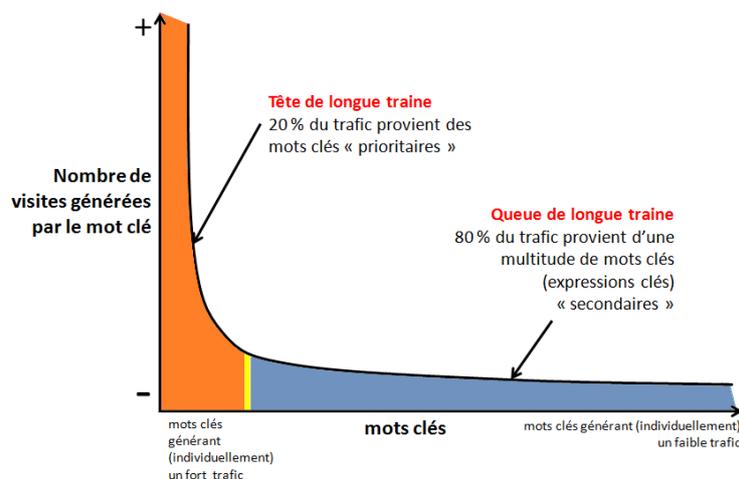
L'idéal est de pouvoir préparer son référencement lors de la création d'un site ou éventuellement lors de sa refonte. Voici une liste d'actions à mener dans l'ordre :

- Choix des mots-clés.
- Choix des moteurs et annuaires.
- Création des pages en fonction des mots clés.
- Ajout des pages dans les moteurs (référencement).
- Vérification du positionnement et du trafic.
- Suivi et corrections pour de meilleurs résultats.

3.1 Longue et courte traine

La traine est un concept très intéressant qui se rapproche assez du principe de Pareto. Josh Petersen, employé d'Amazon décrit⁹ la longue traine comme ceci : « Aujourd'hui, nous avons vendu plus de livres qui ne se sont pas vendus hier que de livres que nous avons aussi vendu hier ». En d'autres termes, il y a beaucoup d'articles peu vendus, et peu d'articles très vendus.

En référencement, c'est exactement le même principe : la courte traine représente les quelques mots-clés qui apportent 20% du trafic global et la longue traine correspond à un grand nombre de mots-clés qui apportent, individuellement un faible trafic.



La longue traine¹⁰

Cette longue traine est d'autant plus intéressante lorsque le coût de stockage des articles vendus est quasi nul comme sur « iTunes », « Netflix » ou les vendeurs d'« iBook ». Le but est d'avoir une librairie/ludothèque immensément grande en supposant qu'un article aura, au moins une fois, trouvé un acquéreur.

Il faut donc partir sur deux stratégies :

- Pour la courte traine : Trouver les mots-clés qui représentent bien l'activité et optimiser les pages. Ceux-ci représenteront 20% du trafic. Les internautes qui viendront sur le site via ces mots seront ceux qui veulent vraiment y venir.

⁹ http://longtail.typepad.com/the_long_tail/2005/01/definitions_fin.html Commentaire du 10 Janvier 2005 à 08h25.

¹⁰ <http://www.olivier-corneloup.com/ressources/dico-referencement/definition-longue-traine>

- Pour la longue traîne : Mettre en avant le texte et la structure du site. Certaines pages seront visitées via des mots-clés qui n'ont pas été pensés. On appelle cela un trafic opportuniste.

3.2 Google suggest

Google suggest a été lancé en 2008¹¹ afin de suggérer d'autres requêtes de recherche aux internautes.

La liste est calculée en fonction des recherches les plus populaires et de l'historique de recherche de l'internaute. Elle gère également la correction orthographique¹².

Google Suggest est malheureusement impossible à anticiper, même pour Google. Il suffit de taper un nom de film pour se rendre compte que les suggestions sont illégales : *streaming*, téléchargement, etc... Certaines sociétés ont même porté plainte¹³ contre cet outil car le nom de leur société était associé à des mots peu envieux tels que « arnaque », « voleur », etc...



Avec Google Suggest, le CNFDI est lié au mot "arnaque"

La suggestion est un bon outil pour le marketeer pour « prendre la température » de sa marque. Le site *Uber Suggest*¹⁴ propose même une liste beaucoup plus longue que celle proposée par Google.

3.3 Choix des mots-clés

Pour choisir les mots-clés, il faut suivre deux principes :

- Ils doivent être souvent tapés par l'utilisateur.
- Si un internaute utilise ce mot-clé, il doit être techniquement possible, dans des délais raisonnables, de se positionner dans les premiers de la liste.

Il est également important de prendre en compte le fait que les requêtes lancées par l'internaute se rallongent¹⁵ pour arriver à une moyenne de 4¹⁶ mots-clés.

Pour trouver ces mots clés, il y a plusieurs stratégies :

- Se fier à l'intuition.

¹¹ <http://googleblog.blogspot.be/2008/08/at-loss-for-words.html>

¹² <https://support.google.com/websearch/answer/106230>

¹³ <http://www.abondance.com/actualites/20090727-9982-google-suggest-nouveau-jugement-cette-fois-favorable.html>

¹⁴ <http://ubersuggest.org>

¹⁵ <http://searchengineland.com/search-queries-getting-longer-16676>

¹⁶ <http://www.abondance.com/actualites/20120126-11246-plus-de-4-mots-cles-en-moyenne-dans-les-requetes-moteurs.html>

- Chercher dans une base de données : Avec *Google hot trends*¹⁷ ou *Word Tracker*¹⁸.
- User de l'auto complétion de *Google suggest* pour voir ce qu'il propose.
- Lancer des sondages publiques ou en interne.
- Vérifier les résultats de *Google* : En cherchant avec un mot connu, d'autres mots pourront peut-être apparaître.
- En bas de la *SERP*, *Google* indique une liste de « recherches associées » bien pratique pour trouver des idées.
- Faire un audit de la concurrence en allant voir les « meta » des concurrents. Nous parlerons de leurs significations au chapitre 4.4 : Zone chaude 4 : Les balises « meta ».
- Penser aux fautes de frappe/orthographe qu'il faudra placer « intelligemment » sur le site pour ne pas être décrédibilisé. En effet, il serait judicieux de créer une page intitulé « Comment nos visiteurs nous ont trouvé ce mois-ci » et y placer les fautes d'orthographe les plus courantes. Une autre possibilité est de se faire passer pour un visiteur lambda, et d'insérer des fautes d'orthographe dans des commentaires.

3.4 Analyse des mots-clés

Une fois les termes trouvés, il faut analyser si, pour chacun d'eux, il est techniquement possible d'être bien positionné. En les utilisant sur *Google* il est possible de se faire une idée en fonction du nombre de résultats :

- En dessous de 100.000 : Cela ne devrait pas poser problème.
- Entre 100.000 et 500.000 : C'est possible à moyen terme.
- 500.000 et plus : Le résultat sera très aléatoire mais peut être réalisable sur le très long terme et avec un peu de chance. L'optimisation à elle seule ne suffira peut-être pas et il faudra sans doute avoir une bonne qualité de *backlinks* pour se démarquer des concurrents.

Tout dépend également de l'agressivité des concurrents et de l'expression visée. En effet, il est plus difficile d'être le premier pour « hôtel » que pour « hôtel de luxe à Auderghem ». Cependant, comme nous l'avons vu, les requêtes s'allongent et il y a fort à parier que personne ne cherche juste « Hôtel ».

Bref, sur des mots peu concurrentiels, l'optimisation des pages devrait suffire. Par contre, pour les plus concurrentiels, la différence se fera sur les *backlink*, et d'autres critères « off page » que nous verrons au chapitre 5 : Optimisation des pages du site : Critères « off page ».

3.5 Gestion des évènements et référencement prédictif

Pour des évènements fixes comme Noël ou la saint-Valentin, il est clair que la stratégie est différente. Il faut s'y préparer à l'avance, mais combien de temps ?

Grâce à *Google trends*¹⁹, il est possible d'analyser les fréquences de recherche pour un ou plusieurs mots-clés.

¹⁷ <http://www.google.com/trends/>

¹⁸ <http://www.wordtracker.com/>

¹⁹ <http://www.google.be/trends/?hl=fr>



Résultat pour le mot clé "Saint-Valentin"

On voit clairement les pics à chaque début d'année. En zoomant, on s'aperçoit que les internautes cherchent en janvier. C'est, dès lors, à partir de cette période qu'il faut mettre les articles en ligne pour ne pas rater ceux qui recherchent des cadeaux très tôt. Cela permet aussi à *Google* d'apprécier/de digérer le site, de voir qu'il est à jour et qu'il correspond bien à la tendance du moment.

Il ne faut pas pour autant trop pousser dès le début. D'abord, il suffit de lancer une page avec quelques informations. Ensuite, la mettre à jour régulièrement pour que les *spiders* prennent l'habitude de souvent y revenir. Enfin, augmenter la cadence petit à petit et ajouter de plus en plus de contenus.

3.6 Où se référencer ?

Il est primordial de se référencer sur les moteurs de recherche et, à moindre mesure, sur les annuaires.

Il faut savoir que 90% des internautes qui arrivent d'un moteur de recherche, proviennent de *Google*. Il est donc primordial de se focaliser sur lui. Au final, *Bing* utilise quasi le même algorithme, donc si le site est optimisé pour *Google*, il le sera sans doute pour ses concurrents.

Concernant les annuaires, s'ils peuvent apporter un bon *backlink*, s'y inscrire n'est jamais perdu. D'autant plus que les visiteurs provenant des annuaires sont généralement de qualité.

Lors de l'inscription d'un site à un annuaire, il faudra vérifier plusieurs choses:

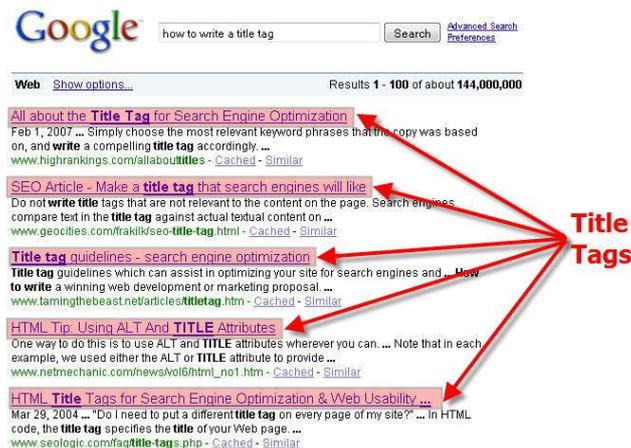
- L'annuaire doit faire un lien **direct** vers le site et non pas un lien de redirection tel que « <http://annuaire.com/redirection?site=16547> », car cela ne comptera pas comme un *backlink*.
- Si l'annuaire demande un lien en retour, tout le bénéfice sera perdu.
- Si le site est dans une page avec énormément de liens, le bénéfice sera nul. En effet, *Google* compte le nombre de liens sur la page pour se faire une idée de l'importance de chacun. Nous reparlerons de ce point au chapitre 5.1 : PageRank.
- *Google* ne voit pas d'un bon œil l'utilisation des annuaires car certains sont juste créés pour essayer d'augmenter les *backlinks*.

4 Optimisation des pages du site « in page »

Maintenant que les mots-clés ont été choisis, il faut les insérer dans les zones considérées comme chaude par *Google*.

4.1 Zone chaude 1 : la balise « title »

La balise « title » est insérée dans l'en-tête (ou *head*) du code *HTML* de la page. C'est elle qui est affichée dans les onglets du navigateur et c'est aussi elle qui sera affichée dans les résultats de *Google*.



Les titres sont les liens vers le site²⁰

Le titre doit contenir des mots clés importants car c'est la zone la plus utilisée par *Google*. Il ne s'agit donc pas d'y mettre « accueil » ou « bienvenue sur mon site » comme on peut le voir encore trop souvent. On peut, en effet, douter du nombre de personnes ayant déjà cherché « bienvenue » sur un moteur de recherche.

Voici quelques recommandations générales pour le titre d'une page :

- Il doit contenir entre 7 et 10 mots descriptifs, sans compter les *stopwords*.
- Il doit être différent sur chaque page.
- S'il est trop long (+/- 70 caractères), *Google* le coupe. Il faut donc placer les mots les plus importants en premier lieu.
- Il faut utiliser des mots uniques et ne pas trop les répéter de page en page. Autrement cela pourrait être interprété comme du *spam*.
- Le titre est aussi ce qui va être utilisé dans les favoris du navigateur. Il est donc intéressant d'y mettre ce que la page représente vraiment.
- Il doit être écrit en une seule langue.
- Parfois, il est intéressant d'y insérer une localisation, pour les restaurants ou hôtels par exemple.

Pour une page d'accueil, il est de bonne pratique d'y mettre le nom de la société et quelques mots sur son activité. Le slogan de la marque est plus que dispensable s'il ne décrit pas réellement l'activité. À titre d'exemple, la société « MMA » n'a heureusement pas utilisé « Mma - zéro tracas, zéro blabla » comme titre pour son site <http://www.mma.fr>.

Pour les pages internes, il est beaucoup plus optimisé d'utiliser l'ordre « titre éditorial – rubrique – nom du site » que l'inverse pour deux raisons. D'abord, comme nous l'avons déjà vu, le titre peut être

²⁰ <http://labs.linkeo.com/seo-la-balise-title-dans-les-serps-de-google/>

coupé s'il est trop long. De plus, Il semblerait que seuls les 12 premiers caractères soient pris en compte²¹. Enfin, *Google* rajoute parfois, lui-même, le nom du site à la fin du titre²², il n'est donc pas nécessaire de l'indiquer.

Plus on descend dans l'arborescence du site, plus les mots-clés devront être précis. Aussi, il faut avantager les pages plus petites avec des titres précis, plutôt qu'une grosse avec trop de contenus « différents » et un titre plus vague.

4.2 Zone chaude 2 : Le texte visible

Par « texte », on entend tout texte qui peut être copié/collé. Ce n'est ni du flash, ni une image, ni une animation, ni une vidéo.

Par « visible », on entend qu'il est affiché dans des conditions normales. Un texte blanc sur fond blanc, par exemple, est non-visible et considéré comme une fraude. *Google* a pour habitude de pénaliser les sites ayant recourt à ce genre de pratique. Il existe plusieurs techniques pour « ne pas afficher du texte », mais, si certaines fonctionnent parfois encore, il est tout de même déconseillé de les utiliser à moins de vouloir jouer avec le feu.

Pour pouvoir comprendre comment le *spider* lit les informations d'une page, il faut se mettre à sa place, pour cela il existe des simulateurs²³. Le texte est repris sans sa mise en page et les liens sont répertoriés. Pour faire ressortir des mots, il existe plusieurs méthodes.

Tout d'abord, les mots en gras (avec la balise *HTML* « *strong* ») ont plus de poids qu'un texte normal. Il faut les utiliser avec parcimonie et donc bien cibler les mots à mettre en évidence.

Google considère aussi que les liens sont importants. Leur libellé a beaucoup de valeur, c'est pour cela qu'il ne faut surtout pas utiliser de termes comme « cliquez ici ». Bien entendu, si un lien est en gras, son poids est encore plus grand.

Les titres éditoriaux (balise *HTML* « *h1*, *h2*, *h3* », etc...), qui représentent des titres sur plusieurs niveaux (comme sur *Word*), ont aussi un impact. Le premier titre (« *h1* ») doit être unique sur la page et représente le contenu, comme un titre éditorial par exemple. Ce dernier contient souvent le même texte que la balise « *title* » vue précédemment. Bien sûr, au fur et à mesure que l'on descend dans l'arborescence, au moins les mots ont du poids. Pour un contenu éditorial, on utilise le plus souvent cette hiérarchie :

- Balise « *h1* » : Le titre éditorial
- Balise « *h2* » : Le chapeau
- Balise « *h3* » : Les sous-titres de l'article

Enfin, il est nécessaire de « penser *Google* » lors de la rédaction d'articles. Il existe pour cela plusieurs astuces, en plus de toutes celles que nous venons de voir :

- Grouper des mots. Si dans le contenu, il est écrit « Bougie pour les fêtes de Noel », il sera opportun de changer la phrase en « Bougie de Noel », car ce sera le terme qui sera probablement le plus recherché dans *Google*.

²¹ <http://alekseo.com/longueur-balise-title/>

²² <http://www.abondance.com/actualites/20120524-11493-comment-eviter-les-modifications-de-titre-par-google.html>

²³ <http://www.webconfs.com/search-engine-spider-simulator.php> ou <http://smallseotools.com/spider-simulator/>

- L'ordre des mots à son importance : « monde maison » ne renvoie pas le même résultat que « maison monde »
- Le pluriel et les synonymes ne sont pas ou peu gérés par *Google*, il faut donc essayer de les utiliser autant que possible²⁴.
- Le contenu éditorial doit être de 200 mots minimum, autrement *Google* placera la page dans son second index.
- Il ne faut aborder qu'une seule thématique par page.
- Comme pour les titres, il faut utiliser une seule langue.
- Le fait de faire des liens vers des pages internes du site (appelés liens internes ou *deeplinking*) « dans la même thématique » aidera aussi *Google* à mieux comprendre la page.

4.3 Zone chaude 3 : L'adresse des pages

Plus communément appelé « Url » pour « Uniform Resource Locator », l'adresse doit être soignée et bien pensée.

4.3.1 Nom de domaine

Pour avoir un bon référencement, il est indispensable d'avoir son propre nom de domaine²⁵. Les adresses gratuites sont très mauvaises pour le référencement et elles ne valent pas l'économie de quelques euros par an. Les mots compris dans le nom de domaine ont un certain poids, mais il ne faut pas en abuser. En effet, les noms de domaines qui ressemblent plus à une suite de mots-clés qu'à une adresse (« formation-soir-entreprise-bruxelles.be ») sont détectés par *Google* via son filtre « EMD » (pour « *Exact match domain* ») et sont sujets à des sanctions telles que le bannissement ou la rétrogradation.

Google prend également en compte l'extension (« .com », « .be », etc...) et la localisation de l'hébergeur du site. Pour s'en rendre compte, il suffit de taper « actualité » sur *Google.fr* et sur *Google.be* pour s'apercevoir que les résultats sont liés au pays. Il est conseillé d'acheter le « .com » en plus de l'extension de son pays, ne serait-ce que pour être sûr que quelqu'un ne le « vole » pas.

Aussi étrange que cela puisse paraître, l'âge du nom de domaine et sa date d'expiration a aussi un impact. *Google* fait plus confiance aux noms de domaine qui sont « vieux » et qui ont une date d'expiration lointaine. Il vaut donc mieux réserver son nom de domaine sur plusieurs années plutôt que de le renouveler chaque année (qui plus est, en dernière minute).

Si le nom de domaine est composé, le tiret est à privilégier car *Google* le considère comme séparateur de mots-clés. Cependant, pour communiquer une adresse longue aux médias classiques, il est plus simple de communiquer une adresse sans tiret car elle est plus facile à indiquer oralement (ex : <http://www.pourquoiuroulaissivite.be/>). Dans ce cas-là, l'achat des deux versions est opportun avec l'un qui redirige vers l'autre. Il est toutefois important de ne communiquer que sur un seul nom de domaine, que ce soit en communication online ou offline (carte de visite, trac, posters, etc..) pour deux raisons :

- Cela risque d'embrouiller une partie de la cible.
- *Google* n'est pas friand de ce genre de pratique et peut considérer cela comme du *spam*.

²⁴ <http://blog.axe-net.fr/le-referencement-les-accents-le-pluriel-et-lorthographe> ou <https://guyot16.wordpress.com/2012/12/23/en-ligne-seo-marketing-utiliser-le-singulier-et-pluriel-votre-mot-cle-dans-le-titre-webpage>

²⁵ Le nom de domaine est l'adresse que l'on affiche sur une carte de visite. Par exemple : ichec-entreprises.be

Il est important de préciser que les moteurs de recherche n'affichent dans leurs résultats que 3 ou 4 pages provenant du même nom de domaine. On appelle cela du *clustering* : Récupérer les meilleurs pages du site qui correspondent à la recherche.

En ce qui concerne les sous-domaines, ils ne sont plus considérés comme des sites différents. Ce qui veut dire que les adresses <http://music.monsite.com> et <http://forum.monsite.com> sont identiques aux yeux de Google. Une pratique qui consistait à « couper le site en sous-domaine » n'a, en conséquence, plus d'intérêt pour avoir plus de pages visibles dans une recherche.

Il faut savoir que le nom de domaine a de moins en moins de poids et que rien ne remplacera un contenu de qualité. Cela ne sert donc à rien de faire des achats en masse.

4.3.2 Nom des dossiers/pages

Le nom des dossiers et des pages, même si ceux-ci peuvent être virtuels, apportent aussi du poids dans la balance. En effet, *Google* (et l'humain aussi d'ailleurs) comprendra mieux <http://www.ikea.be/salon/meuble-tv> que <http://www.ikea.be/category/5784>. Il est important de préciser que les barres obliques²⁶ sont aussi considérées comme des séparateurs, contrairement aux tirets bas²⁷ qui semblent être encore mal gérés²⁸. Les caractères exotiques tels que « * », « ? », « + », les caractères accentués ou diacritiques sont à bannir car ils ne sont pas pris en compte.

4.4 Zone chaude 4 : Les balises « meta »

Les balises « meta » correspondent à des informations sur l'information qui sont placées dans l'en-tête (« head ») du fichier *HTML*. Elles sont invisibles dans le navigateur et peuvent indiquer la langue, l'encodage, l'auteur, etc... Nous allons parcourir les balises les plus importantes.

4.4.1 Keywords

Elle contient les mots clés importants d'une page. Malheureusement, suite aux abus de beaucoup de sites, cette balise n'est plus prise en compte. Elle est donc facultative.

4.4.2 Description

Sur sa page de résultat, *Google* affiche le contenu de la balise « description » sous le titre. S'il n'existe pas de description, qu'elle est non adéquate au contenu ou qu'elle est trop courte (150-200 caractères), le premier extrait pertinent de la page est utilisé. Par exemple, malgré le fait que la description du site <http://www.novomeo.fr/> soit « Novomeo, le logiciel homéopathique du futur », *Google* décide, malgré tout, d'afficher sur sa *SERP* le début du contenu éditorial de la page.

Voici quelques conseils à suivre :

- La description ne peut pas être identique au titre, ni même le contenir.
- Elle ne peut pas être une liste de mots-clés.
- Elle doit être un complément d'information. Par exemple, pour un site e-commerce, l'idéal est d'y mettre des informations complémentaires comme le prix, la taille, le constructeur, etc... On peut aussi y mettre le chapeau d'un article éditorial.
- Elle doit être différente sur chaque page. Si ce n'est pas le cas, il vaut mieux ne pas en mettre.

²⁶ « *Slash* » en anglais

²⁷ « *Underscore* » en anglais.

²⁸ <http://www.abondance.com/actualites/20091013-8825-les-underscores-dans-les-url-questionsreponses-avec-google-1.html>

4.4.3 Schema.org

« *Schema.org* » est le fruit d'une collaboration entre *Yahoo!*, *Microsoft* et *Google* pour promouvoir un web plus sémantique. Grâce à des balises spéciales insérées dans le code *HTML* de la page, les moteurs de recherche peuvent encore mieux comprendre le contenu et afficher des informations complémentaires pour des événements, des restaurants, des recettes, etc²⁹...

Pour un film par exemple, il est possible de préciser une centaine d'informations (facultatives) comme la note moyenne, les acteurs, la durée, la bande originale, le producteur, le réalisateur, etc... Sur *Google*, le résultat devient alors encore plus intéressant et attire l'œil.

X-Men (2000) - IMDb
www.imdb.com/title/tt0120903/ Traduire cette page
★★★★★ Note : 7,4/10 - 360.574 votes
21st century Marvel films, the important ones. ... X-Men -- Two mutants come to a private academy for mutants whose resident ... X-Men Origins: Wolverine.

Exemple du film "X-men" avec des étoiles représentant la moyenne des notes

De plus en plus de sites utilisent ces balises, elles sont d'ailleurs vivement recommandées si le site s'y prête.

4.5 Zone chaude 5 : L'attribut « alt »

L'attribut « alt » correspond au texte, lié à une image, qui est affiché si celle-ci n'arrive pas à se charger correctement. Etant donné que les *spiders* ne lisent que le texte brut, ils ne prennent pas en compte les images. Le seul moyen de les aider est d'indiquer la description de l'image dans cet attribut. Il est aussi utilisé pour insérer les mots sans accents, les pluriels ou les fautes d'orthographe qui n'ont pas pu être placé dans le texte éditorial.

Comme ce texte n'est pas réellement visible pour les internautes, son poids est assez faible dans l'algorithme de pertinence. Il est toutefois fort recommandé pour permettre à *Google* d'au moins identifier les images.

²⁹ Le site officiel contient la liste complète : <http://schema.org/docs/schemas.html>

5 Optimisation des pages du site : Critères « off page »

Il y a quelques années, *Google* a instauré dans ses critères de pertinence : la popularité, la réputation et l'indice de confiance. Ces critères n'ont pas de lien direct avec le contenu des pages, c'est pour cela qu'ils sont appelés « off page ».

Depuis ce moment-là, les liens provenant d'un autre site (appelés liens entrants ou *backlink*) ou d'une autre page du même site (appelés liens internes) sont dès lors devenus beaucoup plus importants! Leur intitulé s'est transformé en un mot-clé supplémentaire pour la page cible, d'où l'importance de ne pas utiliser des termes trop abstraits comme « cliquer ici ». *Google* va aussi mieux pouvoir analyser la pertinence d'une page et sa réputation en se basant sur une simple hypothèse : Si une page réputée fait un lien, celui-ci sera très probablement pertinent.

Cela a engendré un nouveau phénomène : Le « *Google bombing* ». Le principe est simple : Plusieurs sites pointent vers la même adresse avec le même terme (maintenant considéré comme mot-clé) en attendant que *Google* remonte petit à petit le résultat. Les premiers ciblés furent les hommes politiques et notamment *Nicolas Sarkozy*. En effet, plusieurs centaines de sites ont mis un lien vers la page *Facebook* de *Nicolas Sarkozy* en utilisant les mots « trou du cul » comme titre du lien.



En cherchant « trou du c.. », le premier lien est la page Facebook de Nicolas Sarkozy

Heureusement, *Google* a, depuis lors, réglé ses algorithmes et les bombes de ce genre se font de plus en plus rares.

5.1 PageRank

Le *PageRank* (*PR* ou indice de popularité) est un des paramètres les plus importants dans l'algorithme de pertinence des moteurs de recherche. Mais comment est-il calculé ?

Il y a quelques années, l'indice était calculé de manière quantitative : Plus une page avait de liens qui pointaient vers elle, plus son indice de popularité augmentait. A présent, les critères sont également d'ordre qualitatif. Ceci veut dire qu'il est possible qu'un page avec moins de liens entrants soit quand même plus visible car elle possède des liens de qualité.

Voici plusieurs aspects à prendre en compte en ce qui concerne le *PageRank* :

- *Google* prend en compte 2 paramètres : la popularité d'une page et celles pointant vers elle. Un lien est plus important s'il provient d'une page à forte popularité plutôt qu'un petit site personnel.
- Le nombre de liens sur une page a également son importance. Plus une page contient de liens, moins ils sont considérés comme important, tout simplement parce qu'ils ont plus de chance d'être noyés dans la masse. Cela permet aussi de contrer les pages, dites « ferme de liens »

(*links farms*) ou FFA (pour *Free For All*). Cette (mauvaise) pratique consiste à faire une page dont le contenu est constitué de liens exclusivement, en espérant augmenter la popularité de chacune d'entre elles. C'est tout à fait faux ! Nous y reviendrons au chapitre suivant 5.1 : Mode de calcul.

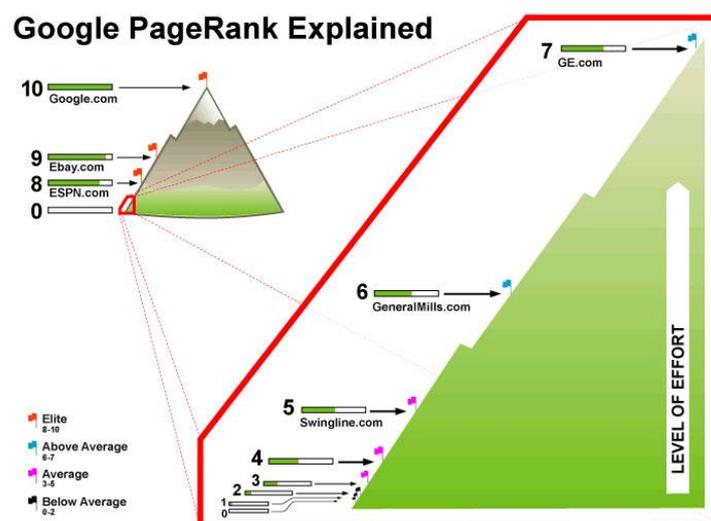
- Un lien provenant d'un site extérieur a bien sûr plus de poids qu'un lien interne. Ce qui est logique vu que ces liens sont plus durs à obtenir.
- L'indice de popularité est attribué à **une page** et non à un site entier. Il y a fort à parier que la page d'accueil sera celle qui a la plus grande popularité. *Google* considère « monsite.com » et « www.monsite.com » comme étant différent, il faut donc bien communiquer une adresse unique sous peine de voir sa popularité éparpillée.
- Pour le moment, le *PageRank* d'une page ne se calcule que sur base des liens qui pointent vers celle-ci.

Il est indispensable de pratiquer une veille sur l'indice de popularité des pages du site. Régulièrement, il faut :

- Vérifier l'existence de nouveaux sites potentiellement intéressants.
- Vérifier que les partenaires n'ont pas supprimé un lien.
- Fidéliser les meilleurs partenaires.
- Assurer qu'il n'y a aucune erreur 404, auquel cas, faire une redirection.

5.1.1 Mode de calcul

L'indice du *PageRank* a une valeur entre 0 et 10. Il peut être observé via la *Google Toolbar* ou via des sites comme <http://www.whatsmypr.net/>. Le chiffre affiché par ces outils est obsolète de plusieurs mois et ne sert donc qu'à titre indicatif. Malheureusement, c'est tout ce que *Google* veut bien communiquer.



Le niveau d'effort augmente à chaque palier

Une page partage son *PR* à **tous** ses liens. On appelle cela le « jus de lien » (ou *linkjuicer*). La page peut être vue comme une bouteille dans laquelle on perce un trou pour chaque lien. Si celle-ci a un *PR* de 6 avec 10 liens, chaque lien aura 0,6³⁰ de popularité en plus. Nous l'aurons compris, au moins il y a de

³⁰ Pour être exact, il faut préciser que ce chiffre est multiplié par 0,85. Du jus de lien est donc « perdu » au passage.

lien, au plus le jus de lien sera bénéfique à chacun. C'est pour cette raison que les fermes de liens sont inefficaces car elles contiennent en général des centaines, voire des milliers de liens.

Ce calcul est un processus très long car il prend en compte des matrices de plusieurs millions (voir milliards) d'enregistrements. De plus, *Google* doit répéter plusieurs fois sa formule pour atteindre nombre qui reste stable.

Il est important de préciser que le *PageRank* à lui tout seul ne suffit pas. En effet, cet indice n'est qu'un facteur parmi 200 dans l'algorithme. Il est possible d'avoir un bon score tout en étant introuvable. Le contenu reste la priorité numéro 1.

5.2 Echange de lien

Etant donné que les liens sont pris en compte, on pourrait se demander si un simple échange (*netlinking*) de lien est intéressant. Malheureusement, ce n'est pas le cas, même si le partenaire a un meilleur *PageRank*³¹.

Cependant, s'il y a *netlinking*, certaines choses sont à prendre en compte :

- Le site partenaire doit avoir un indice de popularité de minimum 4 avec beaucoup de pages référencées. L'outil <https://ahrefs.com/> est très utile pour connaître ce genre d'informations. Il donne, par exemple, le nombre de *backlink* du site, ce qui est un bon indice sur sa popularité.
- Le site partenaire doit absolument avoir un thème en commun. Autrement, *Google* pourra l'ignorer.
- Le libellé pointant vers le site a aussi de l'importance, puisqu'il est considéré comme mot-clé.
- Il faut éviter que le lien soit en bas de page car il peut être ignoré³².
- Ajouter un site dans un annuaire peut rapporter quelques points s'il utilise un lien direct. Il faut également préférer une rubrique avec le moins de sites, afin de limiter la perte du jus de lien.

Chaque jour, les plus gros sites reçoivent une centaine de demandes d'échange de lien. Elles sont, en général, déplacées dans la corbeille sans même être lues. Il existe d'autres types de partenariat³³ plus intéressant pour les deux partis. Un site peu connu pourrait, par exemple, fournir du contenu de qualité à un site célèbre en échange d'un lien : C'est du gagnant-gagnant!

5.3 Paid linking

Un autre moyen de se procurer des liens est « tout simplement » de les payer, mais est-ce une bonne idée ?

Des sites comme <https://www.matomyseo.com/> sont des sites qui mettent en relation vendeurs et acheteurs de liens. *Google* est contre cette pratique et propose même un formulaire qui sert à la dénoncer. Certains sites comme [Overstock.com](http://www.abondance.com/actualites/20091218-8846-la-ponderation-des-liens-entrants-ou-backlinks-questionsreponses-avec-google-3.html)³⁴ ont déjà été sévèrement pénalisés. Le jeu n'en vaut clairement pas la chandelle.

³¹ Démonstration par : <http://mygreatname.com/google-pagerank/pagerank-loss-by-reciprocal-link-01-1.html>

³² <http://www.abondance.com/actualites/20091218-8846-la-ponderation-des-liens-entrants-ou-backlinks-questionsreponses-avec-google-3.html>

³³ <http://searchenginewatch.com/sew/news/2064719/analyzing-most-common-link-building-strategies-best-sew-2010-8221>

³⁴ <http://www.abondance.com/actualites/20110225-10719-overstock-com-nouveau-site-puni-pour-spam-par-google.html>

5.4 Link baiting

Nous l'avons vu, l'échange de lien ne fonctionne plus, payer des liens est une mauvaise idée, il reste alors une astuce, qui elle, est de plus en plus utilisée : Faire du contenu de qualité et attractif qui appelle au partage. Le *Link baiting* est en quelque sorte un « attrape-liens »³⁵. Voici une liste qui montre que ce n'est peut-être pas si compliqué de créer du contenu qui est susceptible d'être partagé :

- Faire un concours qui met en concurrence des sites ou des personnes. Ex : Quel site à la meilleure communauté ? Votez pour votre site préféré !
- Une interview exclusive.
- Proposer un article qui peut potentiellement faire un buzz parce qu'il va à contre-courant de la « pensée actuelle ». Par exemple, lors du drame de « Charlie Hebdo », « Je suis Charlie » était visible partout sur les réseaux sociaux. Un blogueur a eu l'idée d'appeler son article « Je ne suis pas Charlie »³⁶.
- Traduire un article connu (avec l'autorisation de son auteur).
- Les articles ayant un titre avec chiffre fonctionnent assez bien. Par exemple : « Les 5 choses à savoir avant votre interview », « 10 astuces pour... », etc...
- Publier une étude, un travail de fin d'étude ou une infographie intéressante.
- Proposer une liste de ressources, comme « les meilleurs blogs sur l'e-commerce », et tenir cette liste à jour.
- Faire un article le 1^{er} Avril, avec un peu de chance il sera relayé.

Le but est créer un contenu original de qualité. Une fois le contenu mis en ligne, il est parfois nécessaire de contacter des « leaders d'opinions » afin que ceux-ci le partagent. Les réseaux sociaux sont aussi un bon endroit pour faire le buzz.

Une autre technique, appelée *link ninja*, consiste à aller poster ses liens sur des forums ou commentaires d'autres articles. Cela peut fonctionner à condition de ne pas spammer d'autres sites au risque d'être désactivé par ceux-ci.

5.5 Les liens sortants

Comme vu précédemment, le jus de lien dilue équitablement le *PageRank* pour le partager à tous les liens sortants de la page. Pour éviter de trop le gaspiller, il faut faire attention à ne pas insérer plusieurs fois le même lien comme on peut le voir sur certains sites d'information où il y a 3 liens pour le même article : le titre, l'image et le texte « lire la suite ».

Pour aider les webmasters, *Google* a instauré un nouvel attribut qui indique aux *spiders* de ne pas suivre un lien : « rel=nofollow ». Malheureusement, depuis peu, le partage du jus de liens se fait sur les liens « classiques » et les « nofollow ». Ceci veut dire que pour chaque lien à ne pas suivre, du jus se perd et c'est préjudiciable pour tous les liens classiques.

Lorsqu'il existe plusieurs liens similaires sur une page, il **faut** que le premier soit un lien « classique » et que les suivants soient en « nofollow ». En effet, une fois que *Google* trouve un lien à ne pas suivre, il ne prend pas en compte ceux qui ont la même destination dans la suite du document.

Pour *Google*, les liens les plus importants se trouvent dans la partie éditoriale du site. Les liens peu pertinents doivent être limités autant que possible.

³⁵ <http://adscriptum.blogspot.be/2006/02/linkbait-linkbaiting-une-tentative-de.html>

³⁶ <http://blog.marcelssel.com/2015/01/07/je-ne-suis-pas-charlie/>

Les liens « hors contexte », comme par exemple, la charte, les conditions générales, le mot du PDG sont à limiter. On peut penser qu'ils pourraient également prendre du jus de liens alors qu'il n'apporte aucune popularité. L'idéal serait de les placer dans une page contenant le plan du site par exemple.

5.6 TrustRank

Le *TrustRank* est une variable très difficile à connaître car gardée secrète par *Google*. Il est utilisé par le moteur de recherche pour savoir si un site est « de confiance ». Il semblerait qu'il se base sur certains critères (non-exhaustifs) comme :

- Les données du propriétaire du nom de domaine.
- L'ancienneté du nom de domaine.
- La disponibilité du site en mode sécurisé (https).
- La taille du site.
- Le nombre de pages du site.
- Une note donnée par un humain en interne chez Google.
- Le nombre de « +1 » comptabilisé par Google+.

Il y a fort à parier que dans un futur plus ou moins proche, *Google* prenne également en compte les médias associés à un site via des plateformes comme « Youtube » et « Flickr ».

5.7 Temps de chargement

Depuis 2009³⁷, *Google* prend en compte le temps de chargement des pages pour son indexation. Il faut savoir³⁸ que, pour *Google*, perdre une demi-seconde, c'est perdre 20% du trafic. En plus d'avoir un site avec du bon contenu, il faut aussi qu'il soit rapide.

Il existe énormément d'outils pour mesurer les performances. Ce sujet pourrait faire l'objet d'un travail à part entière et n'est donc que brièvement mentionné dans ce chapitre. Pour plus d'informations à ce sujet l'ouvrage « Web Performance Daybook Volume 2 »³⁹ est à consulter.

Il faut retenir que les performances sont un critère dans l'algorithme de *Google*, mais il semblerait que son influence soit encore assez minime. Il n'existe encore aucun exemple concret de site qui s'est vu pénalisé. Cependant, un site rapide sera toujours mieux perçu par les visiteurs. Ce n'est donc jamais perdu.

³⁷ <http://googlewebmastercentral.blogspot.be/2010/04/using-site-speed-in-web-search-ranking.html>

³⁸ <http://performance.survol.fr/2008/06/a-quoi-ca-sert/>

³⁹ <http://www.amazon.fr/Web-Performance-Daybook-Stoyan-Stefanov-ebook/dp/B008CQA8BA/>

6 Référencement multimédia, multi support

La recherche multi-support prend de plus en plus de place chez *Google*. Avec l' « eye tracking », il a été démontré que notre manière de visualiser les résultats a également changé. Du fait que *Google* modifie sa page de résultat, l'ère du triangle d'or⁴⁰ arrive petit à petit à sa fin.

La recherche des médias est la deuxième préoccupation des internautes utilisant *Google*. C'est pourquoi il peut être intéressant de bien les optimiser.

6.1 Image

6.1.1 Le nom de l'image

Le nom du fichier doit avoir un nom descriptif. Il ne faut pas l'appeler « image001 », mais plutôt « bougie-noel » par exemple. Comme pour les adresses, il est préférable d'utiliser un tiret pour séparer les mots-clés pour éviter que *Google* considère « bougienoel » comme un seul mot. Les caractères accentués sont aussi à éviter.

6.1.2 L'attribut « alt »

Cet attribut a déjà été présenté au chapitre 4.5. Nous n'y reviendrons pas. Rappelons toutefois qu'il ne doit pas contenir plus de 10 mots, ni être bombardé de mots-clés sous peine d'être considéré comme *spam*.

6.1.3 Le texte autour de l'image

Le texte qui entoure l'image a également son importance. L'idéal est de placer une légende sous l'image pour rajouter encore plus de contexte.

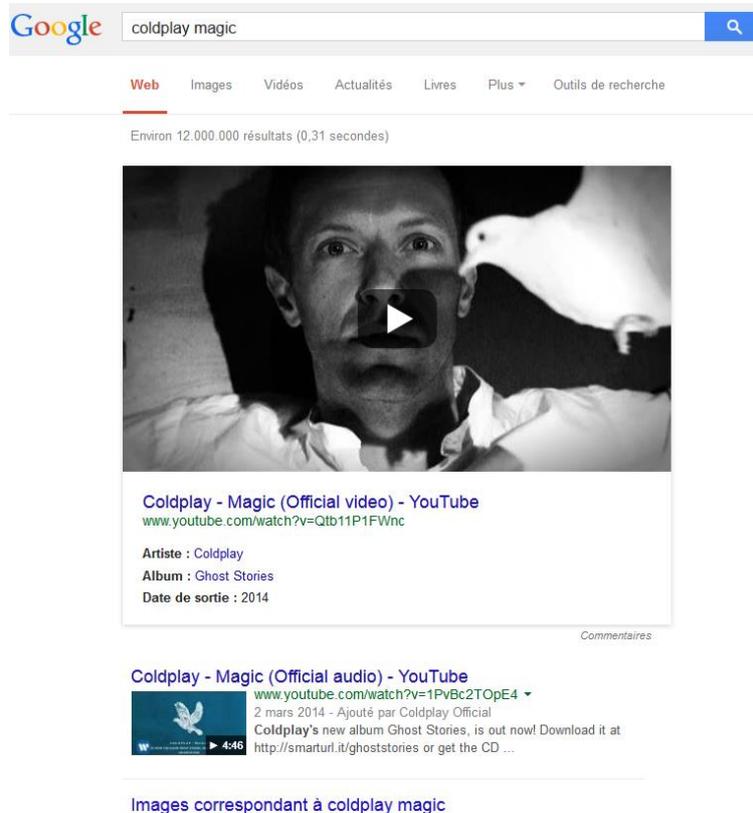
6.1.4 Le texte du lien

Si l'image n'est pas affichée directement, mais accessible via un lien, il faut que celui-ci soit le plus précis possible comme par exemple : « Une bougie rouge et blanche pour les fêtes de Noël ». Si le texte est en gras, le texte prendra d'avantage de poids par rapport aux critères de pertinence.

⁴⁰ <http://searchengineland.com/new-google-eye-tracking-study-shows-downfall-golden-triangle-205274>

6.2 Vidéos

Les vidéos ont énormément d'importance. Suite à une étude⁴¹, il a été démontré qu'une vidéo a beaucoup plus de chance de se retrouver sur la première page des résultats qu'une page au contenu textuel.



Cette recherche renvoi deux liens vers la même vidéo, des images, et enfin des liens "classiques"

Si un site contient des vidéos, il est primordial de bien les optimiser. Les bonnes pratiques sont les mêmes que pour les images, avec cependant quelques nuances.

6.2.1 Meta données

Les métas données sont des données « invisibles » stockées dans le fichier même. Il peut, par exemple, contenir une description, un titre, les coordonnées GPS, l'appareil utilisé, etc... Ce sont ces 2 premières informations qui sont indispensables pour un bon référencement.

6.2.2 Caractéristiques techniques

Une vidéo ne doit idéalement pas dépasser 10 minutes, 100 Mo et garder un *framerate* de 30 images par seconde.

Google arrive de plus en plus à « comprendre » le contenu d'une vidéo : Il peut lire le texte inséré et même retranscrire ce qui est dit dans la vidéo. Il vaut donc mieux également soigner le son.

6.2.3 Les tags

Les tags qui accompagnent la vidéo sont pris en compte. Les vidéos déposées sur *Youtube* profitent de ce genre d'informations.

⁴¹ http://blogs.forrester.com/interactive_marketing/2009/01/the-easiest-way.html

6.2.4 Le texte du lien

Comme pour une image, si la vidéo se trouve sur une autre page, le lien doit être le plus descriptif possible. Aussi il est conseillé de faire une page par vidéo pour optimiser chaque page. En effet, une page avec 10 vidéos différentes sera finalement très mal référencée n'ayant pas un titre et une description claire.

6.2.5 Le texte autour de la vidéo

Une vidéo a besoin d'un maximum de contexte. Il est conseillé de préciser le plus d'informations possible, comme une retranscription par exemple.

Les fichiers « sous-titre », utilisé par *Youtube* sont d'ailleurs excellent pour le référencement.

6.2.6 Indexage

Mettre la vidéo sur *Youtube* et la partager via les réseaux sociaux est idéal car au plus une vidéo est vue, partagée et aimée, au plus elle sera visible dans le moteur de recherche. Indiquer l'adresse du site dans la description est également une bonne pratique afin de ramener quelques visiteurs en plus.

Google propose aux gestionnaires de site de créer un *Sitemap* pour vidéos⁴². Un *sitemap* permet de s'assurer que toutes les vidéos seront bien lues par les *spiders*. Ce fichier XML contient les caractéristiques de chaque vidéo. On y retrouve pour chacune :

- un titre
- une description
- une date de création
- une date de publication
- des tags
- une catégorie
- une note sur 5
- La durée

6.3 Fichier word, powerpoint et pdf

Les fichiers au format pdf (.pdf), powerpoint (.ppt) et word (.doc) sont aussi référencés. *Google* utilise leur contenu, leur titre et leur nom de fichier pour les indexer.

La description affichée dans les résultats correspond au premier contenu contextuel identifiable. Cela veut dire qu'il est préférable de ne pas commencer le fichier par un tableau, par exemple, car son formatage peut être mal interprété par *Google*.

Pour ce qui est de l'indexation, comme pour les images, il faut utiliser un nom de fichier et un lien clair (« Cliquez ici » à éviter). La taille du fichier ne doit pas dépasser 1 Mo auquel cas il ne sera peut-être pas totalement indexé.

Pour le reste *Google* considère ces fichiers comme une « page classique ». Ils ont leur propre *PageRank* et leurs liens sont pris en considération pour le partage du jus de lien. Les images qui s'y trouvent ne sont, par contre, pas encore prises en compte.

⁴² <https://support.google.com/webmasters/answer/80472?hl=fr>

6.4 Le référencement local

Google apporte de plus en plus d'importance aux résultats locaux. Ils sont caractérisés par la présence d'une carte (*Google Maps*) lors de l'affichage des résultats.

The screenshot shows a Google search for "restaurant bruxelles". The search bar at the top contains the text "restaurant bruxelles" and a magnifying glass icon. Below the search bar, there are navigation tabs for "Web", "Maps", "Images", "Vidéos", "Actualités", "Plus", and "Outils de recherche". The search results indicate "Environ 50 600 000 résultats (0,27 secondes)".

The main content area is divided into two columns. The left column lists several restaurant entries, each with a name, website URL, and Google review rating:

- Belga Queen Brussels**: www.belgaqueen.be, 3.5 stars, 88 avis de Google
- Vincent Restaurant**: www.restaurantvincent.com, 3.5 stars, 44 avis de Google
- Villa Lorraine**: www.villalorraine.be, 3.8 stars, 12 avis de Google
- Le Café de Bruxelles**: plus.google.com, 2 avis de Google
- Le Jaloa**: www.jaloa.com, 4.0 stars, 7 avis de Google
- Comme Chez Soi**: www.commechezsoi.be, 4.3 stars, 54 avis de Google

The right column features a map titled "Plan de 'restaurant bruxelles'" showing the location of the results in Brussels. Below the map, there are several restaurant listings with their addresses and phone numbers:

- A** Rue du Fossé aux Loups 32, Bruxelles, 02 217 21 87
- B** Rue des Dominicains 8-10, Brussel, 02 511 26 07
- C** Avenue du Vivier d'Oie 75, Bruxelles, 02 374 31 63
- D** Grand Place 12A, Bruxelles, 02 503 33 25
- E** Place Sainte-Catherine 5, Bruxelles, 02 513 92 62
- F** Place Rouppé 23, Bruxelles, 02 512 29 21

Additional listings on the right include:

- Brasserie Restos Bruxelles**: www.cavesdelabbaye.be/, 071 51 98 28
- Restaurant L'Altro Mondo**: www.liotsacre.be/altromondo.htm
- Restaurant Le Marmiton**: www.lemarmiton.be/
- Restaurant Bar COSPAIA**: www.cospaia.be/

Les résultats "non locaux" sont placés au second plan

Il est clair que tout gérant d'activité locale, comme un restaurant ou un hôtel, doit se placer dans ces résultats, sous peine de très rapidement se retrouver en seconde page⁴³. D'autant plus que ce n'est pas si compliqué : Il suffit de s'inscrire sur « Google + Local »⁴⁴. Lors de l'inscription, il faut être le plus complet possible pour que le site soit mieux classé. Encore une fois, il ne faut pas tomber dans l'excès et sur-peupler de mots-clés les champs titre, description et autres.

Google apprécie qu'on lui donne un maximum de photos et de vidéos. L'avis des internautes aura également un impact, surtout s'il est positif. Dans le cas contraire, il vaut mieux répondre cordialement aux notes négatives pour montrer à Google notre bonne volonté.

Depuis peu, il est possible de payer pour publier des annonces sur *Google Adwords Express*⁴⁵. Cela correspond à *Google Adwords*, mais pour les résultats géo-localisés.

⁴³ <http://www.abondance.com/actualites/20110502-10857-google-adresses-premiere-source-de-traffic-pour-les-sites-locaux.html>

⁴⁴ Inscription gratuite sur <https://plus.google.com/u/0/dashboard?hl=fr>

⁴⁵ <http://www.google.fr/adwords/express/>

7 Technologies à éviter

Ce chapitre présente les technologies à ne pas utiliser. Des solutions existent pour chacune d'entre elles. Cependant, elles ne seront pas mentionnées dans ce chapitre car elles sont trop techniques et pourraient faire l'objet d'un travail à part entière.

7.1 Les frames

Les *frames*, ou fenêtres en français, permettent de diviser une page en plusieurs fenêtres plus ou moins indépendantes, de sorte qu'une fenêtre puisse changer sans recharger toute la page. Cela se faisait beaucoup dans les années 2000 car la bande passante y était très réduite. Actuellement, il est fortement déconseillé de les utiliser. Cette technologie n'est d'ailleurs pas compatible avec tous les navigateurs et n'est plus supporté par *HTML5* (la dernière version d'*HTML*).

En général, les frames étaient, le plus souvent, utilisées pour diviser une page en deux parties: Une pour le menu, et une pour le contenu. Malheureusement *Google* analyse ces fenêtres comme étant 2 pages différentes. Ce qui veut dire qu'il indexe le menu d'une part, et le contenu de l'autre. Par conséquent, lorsqu'un internaute arrive sur une page via *Google*, il ne peut pas naviguer sur le site étant donné qu'il n'a pas son menu.

7.2 Le flash

Le flash est une technologie qui permet, entre autres, de faire des animations complexes. Pour fonctionner, elle nécessite l'installation d'un programme tiers⁴⁶. Ces animations, qui représentent parfois **tout le site**, sont enregistrées dans des fichiers « .flv » et sont indexés par *Google* mais sont quasiment indéchiffrables.



FLASH SERVICES PORTFOLIO CONTACT ALLIANCES CARE...
www.garciamijdi.com/main_stage1.swf
SERVICES. PORTFOLIO · CONTACT · ALLIANCES · CAREERS. ARCHITECTURE.
INTERIORS. PO. RESIDENTIAL. RETAIL. ENVIRONMENTAL. GRAPHICS.

Exemple d'un site en flash (très mal référencé)

Enfin, il ne faut pas oublier que les « Iphone », « Ipad » d'Apple, et maintenant les « Android » ne sont pas en mesure de lire ce format. Ceci ne fait qu'empirer la situation du *Flash* et *Google* pourrait commencer à « descendre » dans ses résultats les sites utilisant cette technologie.

7.3 Le javascript

Le javascript est une technologie qui a le vent en poupe en ce moment. Elle permet de faire de plus en plus de choses sur une page : animations, alertes, modales, modification de la page en temps réel, etc. *Google* n'interprète pas encore le javascript, il faut dès lors éviter d'écrire les liens dans ce langage de programmation.

7.3.1 Ajax

Ajax⁴⁷ est une manière, **en javascript**, de récupérer du contenu dynamique sans recharger la page. C'est ce qu'utilise Facebook pour pouvoir faire du scroll infini. Cela permet aux pages de se charger plus rapidement car elles n'affichent, dans un premier temps, que le minimum vital.

Malheureusement, comme *Google* n'interprète pas le javascript, les informations chargées de manière asynchrones ne seront pas lues par le *spider*. Ceci veut dire que le contenu additionnel ne sera pas indexé.

⁴⁶ Le bien connu « Flash player » de Adobe : <http://get.adobe.com/fr/flashplayer/>

⁴⁷ *Asynchronous JavaScript and XML*

Tant que *Google* ne prévoit pas d'interpréter le javascript, *Ajax* ne rimera pas avec optimisation *SEO*. Cependant, les choses risquent peut-être de s'accélérer car *Google* est entrain d'écrire une librairie⁴⁸ en javascript...

7.4 Formulaire

Google ne sait pas remplir de formulaire. Si des pages ne sont accessibles que via une barre de recherche de recherche ou une liste déroulante, *Google* ne pourra jamais les indexer.

Une solution possible serait de prévoir un lien vers un plan du site qui regroupe toutes les pages qui ne sont pas accessibles pour *Google*.

7.5 URL paramétrée

Les sites sont, pour la plupart, des sites dynamiques. Un site dynamique, contrairement à un site statique, est un site qui peut être géré par un administrateur : Il peut ajouter des pages, modifier du contenu, ajouter des articles si c'est un e-commerce, etc...

Dans ce cas-ci, les pages n'existent pas physiquement, et il se peut que l'adresse ressemble plutôt à ceci : <http://www.monsite.com/article.php?category=5&article=6&page=58>. *Google* n'aime pas ce qu'on appelle des paramètres dans l'adresse, c'est-à-dire, tout ce qui se trouve après le « ? » et qui est séparé par des « & ». Dans ce cas-ci, les paramètres de cette adresse sont faciles à comprendre: La page « 58 » est appelée avec l'article « 6 » dans la catégorie « 5 ».

A partir de 4 paramètres différents, *Google* semble ne plus indexer la page. Au-delà de 4, il est nécessaire de faire appel à la réécriture d'adresse⁴⁹ pour qu'elle soit plus conventionnelle. Dans notre exemple, une meilleure adresse serait : <http://www.monsite.com/article/category-5/article-6/page-58>. Ici, *Google* n'y verra que du feu et croira qu'il a affaire à des dossiers « classiques ».

Cette méthode demande une modification côté serveur⁵⁰ et un peu de rigueur pour fonctionner. La majorité des hébergeurs ont ce module installé. Il suffit de l'activer et de modifier les liens déjà présents sur le site pour qu'ils utilisent le nouveau format.

Les agences web sont de plus en plus conscientes de ce problème, il est d'ailleurs assez rare de voir de nouveaux sites naître avec des adresses paramétrées.

7.6 Cookie & session

Les cookies sont des petits fichiers qu'un site peut déposer sur un ordinateur. Ils sont utilisés pour reconnaître un visiteur lorsqu'il revient sur le site. N'importe quel visiteur peut refuser l'utilisation des cookies, c'est d'ailleurs ce que fait *Google*. Il est indispensable de ne pas bloquer l'accès aux visiteurs qui ne les acceptent pas auquel cas le site ne sera pas référencé.

Une session est une sorte de connexion temporaire entre le site et le visiteur. Elle dure généralement le temps qu'un utilisateur parcourt le site. La plupart des sites l'utilisent pour garder un utilisateur connecté ou pour gérer le panier d'achat en cours. L'identifiant de session peut soit être enregistré dans un cookie, soit être passé dans l'url, via un paramètre⁵¹. Dans les deux cas, *Google* ne prendra pas ces pages en compte. Il faut donc que la session soit utilisée le plus tard possible pour que le *spider* puisse parcourir un maximum de pages avant d'être bloqué.

⁴⁸ <https://angularjs.org/>

⁴⁹ Également appelé *url rewriting*

⁵⁰ Activer le « mod_rewrite » dans « apache » et changer le « .htaccess » ou le « virtual host »

⁵¹ Adresse de type : <http://www.monsite.com/page?sessionId=4d898ed57fdf5a>

8 Conseils à suivre

8.1 Redirection

Dans le cas d'une refonte de site, il est possible que quelques pages aient changé d'adresse. Dans ce cas, il est préférable de prévenir *Google* qu'elles ont « déménagé » afin que les visiteurs n'atterrissent pas sur les fameuses erreurs « 404 : Page introuvable ». La seule chose à faire est d'effectuer une redirection de type « 301 ». Ce code signifie que la page a été déplacée de manière permanente. *Google* se chargera de lui donner son ancien *PageRank*⁵².

8.2 HTTPS

Depuis le 6 Aout 2014, *Google* a indiqué, via son blog⁵³, qu'il améliorera le référencement des sites sécurisés en *HTTPS*. Ce critère ne « pèse » pas encore très lourd et reste bien moins important que le contenu. *Google* semble vouloir donner du temps aux agences pour passer leur site en *HTTPS* avant de lui donner plus de poids. Autant dire qu'il est préférable d'y penser dès aujourd'hui.

8.3 Duplicate content

En Février 2011, *Google* a instauré un nouvel algorithme qui a fait l'effet d'une bombe. Plus connu sous le nom de « Panda », cet algorithme détecte les contenus dupliqués afin de ne pas proposer les deux dans sa page de résultats. Le « challenge », pour les honnêtes rédacteurs, est de prouver que leur contenu est l'original⁵⁴ sous peine d'être déplacé dans le second index de *Google*.

Dans le cas d'un partenariat, comme le fait l'*AFP*⁵⁵ par exemple, il existe plusieurs moyens de démontrer un contenu comme étant la source:

- Il faut que le partenaire ajoute un **lien exact** vers l'article original, en bas de page par exemple.
- Le partenaire doit ajouter, dans son en-tête, un code avec l'adresse de l'article original:

```
<link rel="canonical" href="http://site-original.com/article-original" />
```

Cette balise est gérée par les principaux moteurs de recherche. L'avantage de ce système est que les pages contenant le contenu dupliqué transmettent leur jus de liens à la page originale. C'est une étape à ne pas égliger lors de la rédaction du contrat.

Si les articles d'un site sont copiés à l'insu du propriétaire et qu'aucune de ces pratiques n'a pu être mises en place, *Google* choisira l'article original lui-même sur base de celui qui a le plus de liens entrants ou qui a le meilleur *TrustRank*. Un bon moyen de contrer le plagiat est d'insérer des liens internes vers d'autres pages du site en espérant que ceux-ci soient également copiés. Dans ce cas-là, *Google* remarquera que le site plagié semble être l'original car il contient plus de liens internes.

⁵² <https://support.google.com/webmasters/answer/93633?hl=fr>

⁵³ <http://googlewebmastercentral.blogspot.fr/2014/08/https-as-ranking-signal.html>

⁵⁴ Aussi connu sous le nom de « canonique »

⁵⁵ Agence France Presse : <http://www.afp.com/>

Il faut également faire attention à ce qu'une page ne soit disponible qu'à une seule adresse pour éviter ce qu'on appelle le *DUST*⁵⁶. Cela peut sembler logique, mais voici un exemple d'adresses qui fournissent toutes la page d'accueil d'un site:

- <http://www.monsite.com>
- <http://www.monsite.com/>
- <http://www.monsite.com/home>
- <http://www.monsite.com/home/>
- <http://www.monsite.com/home?pagination=0>
- <http://monsite.com>
- <http://monsite.com/>
- <http://monsite.com/home>
- <http://monsite.com/home/>
- <http://monsite.com/home?pagination=0>
- <http://monsite.be/>
- <http://mon-site.com>

Google considère toutes ces pages comme différentes, même si la même page est renvoyée. A moins de ne communiquer toujours la même adresse, le *PageRank* sera dilué, ce qui est fort préjudiciable. La manière la plus simple de gérer cela est d'utiliser la redirection permanente (Code 301) et surtout de rester cohérent sur le format des liens dans l'entièreté du site.

Le *Duplicate content* s'applique également sur deux pages du même site. Si *Google* remarque que deux pages se ressemblent trop, il placera l'une des deux dans le second index. Il faut donc faire attention à ne pas, par exemple, donner le même titre et la même description à toutes les pages, sous peine d'être considérées comme identiques. Il faut les différencier au maximum et éviter une succession de blocs complètement identiques.

Si toutefois, avoir le même contenu est voulu, mais que le problème du *Duplicate Content* doit être évité, il faut changer au maximum la structure de la page pour essayer de « duper » *Google* :

- Inverser les balises dans l'en-tête *HTML* ou en ajouter/supprimer
- Changer l'ordre des paragraphes lorsque c'est possible
- Changer l'attribut des images
- Changer la structure de l'*html*
- Changer le titre de la page

Ces points ne sont que des exemples et aucune méthode ne fonctionne à 100%⁵⁷.

8.4 Multi langue

Si le site est accessible en plusieurs langues, il existe plusieurs manières d'écrire des adresses :

La première est d'avoir un nom de domaine par langue. Le référencement sera dès lors optimal. Il y a cependant quelques inconvénients. D'abord, il faut payer et renouveler « tous » les noms de domaine chaque année. De plus, les pays ayant plusieurs langues comme la Belgique poseraient problème. En effet, il serait inopportun de rediriger les francophones vers le site français « .fr » et les néerlandophones vers le site néerlandais « .nl ».

⁵⁶ *Duplicate URL, Same Text*

⁵⁷ <http://googlewebmastercentral.blogspot.be/2008/09/demystifying-duplicate-content-penalty.html>

Il est possible d'utiliser les sous-domaines pour spécifier la langue : « fr.monsite.com », « nl.monsite.com », etc... Il ne faudra plus acheter plusieurs noms de domaine. Par contre les sous-domaines sont moins bien compris par les internautes car l'usage du « www » est encore fort ancré.

La dernière possibilité est de faire un dossier par langue : « www.monsite.com/fr », « www.monsite.com/nl », etc... C'est, de manière générale, la solution la plus utilisée.

Dans tous les cas, il est fortement déconseillé de faire une page d'accueil où il faut choisir une langue pour deux raisons :

- Celle-ci aura un très faible *PageRank* dû au peu d'information disponible sur la page.
- Les internautes n'aiment pas cliquer et certains peuvent quitter le site dès la première page à cause d'un clic.

Il existe des moyens pour connaître la langue préférée du visiteur et ainsi le rediriger vers la bonne langue. Il doit également pouvoir changer de langue facilement. En général, les liens « changement de langue » se trouvent en haut ou en bas du site.

Pour aider *Google* à reconnaître toutes les langues du site, il est conseillé d'insérer dans l'en-tête *HTML* des balises « *link* »⁵⁸ pour préciser toutes les langues disponibles.

⁵⁸ <https://support.google.com/webmasters/answer/189077?hl=fr>

9 Accélérer le référencement d'un site

9.1 Soumission manuelle d'un site

Lorsqu'un nouveau site est créé, il est possible de le soumettre à *Google* via l'adresse suivante : <http://www.google.fr/addurl/?continue=/addurl>. Malheureusement, les spiders ne passent que quelques jours (ou semaines) plus tard pour analyser le site. La voie manuelle n'est donc pas idéale pour référencer un site rapidement.

Pour accélérer le processus, il vaut mieux ajouter un lien sur une page existante à fort potentiel. Encore faut-il en avoir une disponible qui :

- est déjà indexée dans *Google*
- est régulièrement mise à jour,
- possède un *PageRank* de minimum 4.

Dans ce cas, *Google* passera sur le nouveau site par l'intermédiaire de cette page au bout de 24 à 48h.

9.2 Optimiser le temps d'indexation

Une autre solution pour voir son site indexé rapidement sur *Google* est de soumettre l'adresse 2-3 mois avant le lancement du site. Il faut alors créer un mini site (une page suffit) qui indique que le site est en construction.

Ceci a comme avantage d'anticiper le fait que *Google* prend son temps avant d'indexer les nouveaux sites et, en plus, d'y être déjà bien installé lorsque le site ouvrira ses portes. Etant donné que les spiders repassent régulièrement sur les pages indexées, ils reviendront rapidement récupérer le contenu de toutes les pages mises en ligne.

Le mini site a pour but d'attirer les futurs consommateurs. C'est le moment idéal pour attiser leur curiosité en proposant un compteur par exemple. Une stratégie « push » peut également s'avérer payante, c'est pourquoi il est souvent intéressant de proposer aux internautes de s'inscrire à une newsletter qui les préviendra lorsque le site sera mis en ligne.

Il faut également proposer du contenu de qualité et essayer de mettre à jour le site régulièrement afin d'habituer *Google* à repasser le plus fréquemment possible. Ainsi, le jour de la mise en ligne, le site définitif sera parcouru au bout de quelques heures seulement.

Une fois le site lancé, il ne faut pas hésiter à s'inscrire dans les annuaires et en discuter sur les forums et blogs ayant les mêmes centres d'intérêt. Tout cela sans pour autant en abuser ou se faire passer pour un spammeur.

9.3 Plan du site

Le plan du site est considéré comme le meilleur moyen de donner les liens des pages principales aux spiders. Il est assez fréquent d'y voir des liens vers des pages intermédiaires qui sont optimisées pour les moteurs de recherche, mais qui ne sont pas disponibles via une navigation classique.

9.4 Sitemap.xml

Pour indiquer à *Google* les pages disponibles sur le site, il est possible de créer une sorte de fichier récapitulatif. Ce fichier, au format *XML*, est appelé « sitemap.xml ». *Google* ne s'engage pas à indexer toutes les pages précisées, mais il a le mérite d'au moins une fois les visiter. En plus de cela, il continuera à parcourir le site de manière « classique ».

Pour remplir ce fichier, il existe des générateurs automatiques⁵⁹ si cela n'est pas déjà fait dans la partie d'administration⁶⁰ du site. En y regardant de plus près, il permet de préciser, pour chaque page, la fréquence des mises-à-jour, la date de dernière modification et la priorité (entre 0 et 1. 1 étant la priorité la plus haute comme une page d'accueil). Si une page venait à être ajoutée, modifiée ou supprimée, le fichier doit être mis à jour au plus vite.

Pour que *Google* prenne en compte le fichier, il faut indiquer son adresse via le « Webmaster tools »⁶¹.

⁵⁹ <https://www.xml-sitemaps.com/> ou <http://gsitecrawler.com/> pour les professionnels.

⁶⁰ Via un *CMS* (pour « content management system ») par exemple.

⁶¹ <https://www.google.com/webmasters/tools/?hl=fr>

10 Mauvaises pratiques

Google a publié⁶² une série de consignes pour les agences web. On y parle de bonnes pratiques mais également des techniques à éviter sous peine de pénalité.

10.1 Techniques à éviter

Parmi les techniques à éviter, nous pouvons en énumérer quelques-unes :

- Le *cloaking* : Afficher du contenu différent aux internautes et aux spiders suite à une détection.
- Cacher du texte en insérant du texte blanc sur fond blanc par exemple.
- Pages satellites⁶³ : Pages sur-optimisées pour les moteurs de recherche suivies d'une redirection automatique vers une autre page.
- *Keyword stuffing* : Répétition non naturelle de mots clés.
- Mettre tout un texte « en gras » ou indiquer trop d'informations dans le titre.
- Redirection frauduleuse : Rediriger les internautes vers une page différente que les moteurs de recherche.

10.2 Détection

Google a mis en place des algorithmes qui sont capables de détecter toutes mauvaises pratiques. Vouloir « être plus malin » que la machine est visiblement possible, mais ce n'est pas pour autant une bonne idée. En effet, si Google ne reconnaît pas (encore) toutes les fraudes, les concurrents peuvent, eux, les signaler via un formulaire⁶⁴ prévu à cet effet.

Le "spam" fait référence à des pages qui essaient de tromper Google pour obtenir un meilleur classement. Avant de signaler du spam, vérifiez si la page ne présente pas un autre problème :

Liens payants Cette page sert à vendre ou acheter des liens.	Droits d'auteur et autres questions juridiques Selon la législation en vigueur, cette page doit être supprimée.
Contenu inapproprié Cette page est inappropriée.	Personnel/privé Cette page divulgue des informations confidentielles.
Logiciels malveillants Cette page est infectée.	Phishing Cette page tente d'obtenir des informations sensibles.
Autres produits Google Cette page porte atteinte à des produits Google autres que la recherche Google, par exemple AdSense, Google Maps, etc.	Extraits enrichis Cette page ne respecte pas les consignes relatives aux extraits enrichis de Google.
Il y a un autre problème Cette page présente des problèmes sans rapport avec le spam.	

Cette page contient réellement du spam.

Le Spam Report

10.3 Types de pénalité

Il existe plusieurs types de pénalités infligées. Elles n'ont jamais été communiquées officiellement, mais elles existent bel et bien. Voici un aperçu des sanctions déjà infligées :

- La *sandbox*⁶⁵: Ce syndrome s'applique aux nouveaux sites référencés, mais pas encore bien indexés. Si Google remarque un comportement anormal en ce qui concerne les liens entrants du site (beaucoup de liens avec le même libellé), il va le mettre « en quarantaine » afin de vérifier si le site fait ses preuves. Au bout de quelques semaines, lorsque le site aura atteint une bonne popularité, la sanction sera annulée et les pages du site auront un meilleur positionnement dans les résultats.

⁶² <https://support.google.com/webmasters/answer/35769>

⁶³ Aussi appelées *doorway pages*

⁶⁴ Le *Spam Report* : <https://www.google.com/webmasters/tools/spamreport?hl=fr>

⁶⁵ <http://www.webrankinfo.com/dossiers/strategies-de-liens/sandbox>

- Le déclassement : Le site se voit reculer de plusieurs places pour une requête spécifique. Cela arrive lorsque *Google* perçoit une répétition abusive du terme dans les pages du site ou dans les libellés des liens entrants. On parle alors de déclassement de 6 à 60 positions, ce qui est énorme car, au mieux, le site se retrouve sous la ligne de flottaison, au pire, il passe en deuxième ou troisième page.
- La baisse de *PageRank* publique : On l'a vu, le *PageRank* publique n'affecte pas le positionnement du site. Cette sanction sonne plutôt comme un avertissement.
- La liste noire : Le pire cas enregistré, le cauchemar de tout *SEO*. *Google* supprime purement et simplement le site de ses index. C'est un cas assez rare, mais qui est déjà arrivé. On peut imaginer qu'à ce stade le fautif doit certainement savoir pourquoi son site a été réprimandé.

10.4 Sortir d'une pénalité

Google a publié une marche à suivre⁶⁶ pour les sites ayant assaini leur *SEO*. Il suffit de se connecter à l'outil « webmaster tools » et soumettre un formulaire avec l'adresse du site et des explications quant aux mesures qui ont été prises sur ce qui aurait pu être vu comme frauduleux.

L'outil « webmaster tools » contient également un système de messagerie afin de communiquer en direct avec *Google*. Lorsqu'un site est sanctionné, c'est ce canal de communication⁶⁷ qui est le plus souvent utilisé.

⁶⁶ <https://support.google.com/webmasters/answer/35843>

⁶⁷ <http://googlewebmastercentral.blogspot.be/2007/07/message-center-let-us-communicate-with.html>

11 Instructions aux *spiders*

11.1 Robots.txt

Le fichier « robots.txt », qui est à la racine du site, contient des indications pour les *spiders* avant qu'ils ne commencent à parcourir le site. Si ce fichier n'existe pas, le site sera visité de manière standard.

Voici un exemple de fichier qui contient à peu près toutes les instructions possibles.

```
Sitemap : http://www.monsite.com/sitemap.xml
User-agent: *
Disallow: /*.jpg$
Disallow: /*.png$
Disallow: /*.gif$
Disallow: /admin/
Allow: /
```

Les points suivants expliqueront les lignes *Sitemap*, *user-agent* et *Disallow/Allow* du fichier exemple ci-dessus.

11.1.1 Sitemap

Le *Sitemap* a déjà été vu précédemment. Cette ligne ne fait qu'indiquer l'adresse où se trouve le fichier.

11.1.2 User-agent

Un « User-agent » représente un moteur de recherche. Voici une liste des noms les plus connus :

Nom du moteur	User-Agent
Alta Vista	Scooter
Excite	ArchitextSpider
Google	Googlebot
HotBot	Slurp
InfoSeek	InfoSeek Sidewinder
Lycos	T-Rex
Voilà	Echo

Cette ligne précise pour quel robot s'appliquent les commandes qui vont suivre. Il est dès lors possible de ne donner d'instructions qu'à certains moteurs. L'étoile (*) signifie « tout le monde » en langage informatique.

11.1.3 Disallow/Allow

Autorise ou non la navigation sur certaines pages ou dossiers. Dans l'exemple donné, toutes les pages peuvent être parcourues à l'exception du dossier « admin » (ce qui est logique) et des images.

11.2 Balise « meta »

Les balises « meta » ont déjà été évoquée au chapitre 4.4 : Zone chaude 4 : Les balises « meta ». En plus de donner des informations sur la page, il est possible d'y insérer des instructions supplémentaires aux *spiders* via la balise « meta name='robots' ».

12 Récapitulatif et derniers conseils

On dit souvent qu'il faut optimiser son site dès sa conception. Les 3 « C » sont là pour rappeler certaines bonnes pratiques :

- Conception : les spiders doivent pouvoir accéder aux pages facilement.
- Code : Il est simple d'extraire le contenu éditorial du design.
- Contenu : Il doit être clair.

12.1 Conception

Une bonne conception facilitera le travail des spiders. Il faut donc, dès le départ, fournir un fichier « robots.txt » valide et un *Sitemap* à jour. Voici un rappel d'autres pratiques indispensables :

- Les liens ne sont pas en javascript.
- N'importe est accessible en 3 clics à partir de la page d'accueil.
- Un plan du site est disponible.
- Les erreurs 404 (et autres) sont bien gérées.
- Le *flash* est les *frames* ne sont pas utilisés.
- Une technique de « url rewriting » est mise en place pour les adresses avec paramètres.
- Les redirections se font à l'aide du code 301.
- La page d'accueil n'a pas de redirection.
- Il n'y a pas de redirection en cascade.
- Les *cookies* et *sessions* ne sont pas indispensables pour parcourir le site.
- Les pages ne comportent qu'une seule langue.
- Afin qu'elle soit optimisée, une page n'a qu'un thème précis: une vidéo, un article, etc...
- Les rubriques sont créées en fonction de ce qui est le mieux pour l'internaute, pas en fonction de la technique.

12.2 Code

Le code *HTML* doit lui aussi être optimisé car de zones chaudes bien remplies apporteront une meilleure visibilité. Il faut se rapprocher au mieux des recommandations suivantes:

- Le titre contient entre 7 et 10 mots descriptifs et est représentatif du contenu de la page.
- Chaque page a un titre unique.
- La structure du titre est : Titre éditorial – Rubrique – Nom du site.
- La balise « description » contient 200 à 300 caractères et est unique.
- La balise « meta : content-language » est précisée sur chaque page.
- Des balises « meta : robots » sont prévues en cas de nécessité.
- L'encodage des pages est la même sur l'entièreté du site.
- Les fichiers javascript et feuilles de style sont dans des fichiers externes
- Le site est navigable sans javascript.
- Les menus sont lisibles par les *spiders*.
- le texte le plus important est situé le plus haut possible, qu'importe la mise en page.
- Les images contiennent un attribut « alt » représentatif.
- Il n'existe aucun contenu caché

12.3 Contenu

Comme souvent indiqué, le contenu éditorial est le nerf de la guerre. Il est important d'éduquer les rédacteurs afin qu'ils rédigent du contenu aussi « sexy » que possible aux yeux des moteurs de recherche et, bien sûr, des internautes. Certains parlent de « *text appeal* »⁶⁸. Voici une liste assez complète de points à prendre en compte :

- La balise « h1 » contient le titre de la page
- Il n'existe qu'une seule balise « h1 »
- Le chapô/premier paragraphe (100 premiers mots) contient des mots déjà présents dans le titre et en propose d'autres.
- Les mots importants sont mis en gras, via la balise « strong », et/ou en lien
- Les autres niveaux de titres sont gérés via les balises « h2 », « h3 », « h4 », etc...
- On retrouve, dans le contenu éditorial, les mots importants dérivés au singulier/pluriel, en synonyme ou en version non-accentuée.
- Le contenu éditorial contient au minimum 200 mots.
- Les mots clés importants sont analysés, au préalable, sur un outil tel que « Google Trends ».
- Le contenu éditorial n'est pas sur-optimisé pour les moteurs de recherche.
- Des liens internes et externes sont proposés dans le contenu éditorial.
- Les liens ont un libellé qui correspond à la page ciblée (« cliquer ici » est à proscrire).
- Les liens ne sont pas en *javascript*.

⁶⁸ <http://s.billard.free.fr/referencement/index.php?2006/11/30/319-optimisation-du-contenu-travaillez-votre-text-appeal>

13 Assurer un suivi

Une fois tout cela mis en place, il faudra mesurer le retour sur investissement et l'efficacité du travail fourni.

Lorsqu'un contrat entre le client et le référenceur est mis en place, il n'est pas rare de discuter sur la position que doit avoir un site dans les résultats de Google. Ceci n'est intéressant que si cela génère du trafic. En effet, il n'est pas vraiment difficile d'être premier sur des mots clés que personne ne saisit.

De plus, Google personnalise 20 %⁶⁹ de ses résultats en fonction de la localisation, de l'actualité, de la langue, des dernières recherches, etc... Etant donné que le positionnement peut changer pour chaque utilisateur, il est difficile de « promettre » une position à un client. Pour surveiller son positionnement, il existe des outils comme « *Advanced Web Ranking* »⁷⁰.

Au final, ce qui importe le plus est de générer du trafic de qualité, c'est-à-dire qu'il doit :

- Augmenter les ventes en ligne.
- Améliorer la notoriété. Pour une campagne de publicité, par exemple, le nombre de visiteurs journalier est utilisé comme référence.
- Engendrer de l' « opt-in » via l'abonnement à une *newsletter* par exemple.

Il est fortement recommandé de se munir d'un outil de statistique tel que « *Google analytics* » afin de mieux connaître ses visiteurs : Combien de temps sont-ils restés ? Par quels canaux sont-ils arrivés, quels sont leurs navigateurs ? Quels sont les mots-clés utilisés ?... Il est très intéressant d'analyser régulièrement quels sont les mots générateurs de trafic. Si un mot n'est jamais visible dans les statistiques, c'est que le site est mal positionné ou tout simplement que ce mot n'est jamais saisi par les internautes.

Enfin, le taux de rebond est aussi important à analyser : Si les visiteurs ne visitent qu'une seule page, le taux de rebond sera élevé. Plus il est faible, plus ils naviguent sur le site et plus ils ont de chances d'y revenir.

⁶⁹ <http://www.abondance.com/actualites/20100309-10271-20-des-resultats-de-google-sont-personnalisés-a-l'internaute.html>

⁷⁰ <http://www.advancedwebranking.com/>

14 Conclusion

Nous l'avons vu, le *SEO* demande énormément de temps et sera encore plus chronophage sur un site existant. Dans un premier temps, il peut être opportun de faire appel à une sous-traitance, surtout pour les *TPE* et les *PME*. *Google* a, d'ailleurs, écrit un article⁷¹ indiquant une série de questions à se poser à ce sujet.

L'idéal veut que « tout le monde soit un bon *SEO* » : Les développeurs, les rédacteurs, les marketeers, etc... Tous doivent partager la même vision d'optimisation. Le but étant de trouver un juste milieu afin de ne pas « sur-optimiser » les pages et ne surtout pas se laisser embarquer dans des pratiques douteuses sous peine de se faire pénaliser par *Google*.

Bien sûr, lorsque deux sites abordent le même sujet, ont quasiment le même contenu et sont tous deux optimisés, la seule manière qu'à *Google* pour les départager est de compter le nombre de liens entrants. C'est pourquoi, il est indispensable de bien gérer sa « visibilité online » en proposant des partenariats, en partageant son contenu sur les réseaux sociaux, etc... *Google* a une préférence assez prononcée pour les « pages qui vivent », il est donc difficile, voire impossible, de bien positionner un site statique.

Etant donné que 30%⁷² des visites proviennent d'un moteur de recherche, il faut également faire attention à ne pas trop en être dépendant et assurer un suivi quasi journalier. En effet, *Google* communique beaucoup via son blog à propos des futures modifications de son algorithme. Les critères évoluent⁷³ au fil du temps : De quelques dizaines, nous en sommes arrivés à plus de 200⁷⁴. Du jour au lendemain, il est possible de descendre (ou monter) dans les résultats. Il vaut donc mieux diversifier au maximum ses sources de trafic car, même pour le *SEO*, « il ne faut pas mettre tous ses œufs dans le même panier ».

Enfin, comment ne pas terminer ce travail sans évoquer la « devise » de tous bons *SEO*. Une phrase à placarder sur les murs qui résume bien l'état d'esprit à cultiver :

« Content is king, link is his Queen, and optimized content is emperor. »

⁷¹ <https://support.google.com/webmasters/answer/35291>

⁷² <http://www.referenceur.be/etude-31-visites-site-web-proviennent-recherche-organique-13196.html>

⁷³ <http://www.searchmetrics.com/wp-content/uploads/seo-ranking-factors-2014.png>

⁷⁴ <http://backlinko.com/google-ranking-factors>

15 Bibliographie

- (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20121012-11966-taux-de-clics-80-pour-les-3-premiers-resultats-naturels.html>
- Message Center: Let us communicate with you about your site.* (s.d.). Récupéré sur Google Webmaster Central Blog: <http://googlewebmastercentral.blogspot.be/2007/07/message-center-let-us-communicate-with.html>
- 20% des résultats de Google sont personnalisés à l'internaute.* (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20100309-10271-20-des-resultats-de-google-sont-personnalisés-a-l'internaute.html>
- A New Click Through Rate Study For Google Organic Results.* (s.d.). Récupéré sur Marketing land: <http://marketingland.com/new-click-rate-study-google-organic-results-102149>
- Analyzing the 9 Most Common Link Building Strategies.* (s.d.). Récupéré sur Search engine watch: <http://searchenginewatch.com/sew/news/2064719/analyzing-most-common-link-building-strategies-best-sew-2010-8221>
- Andrieu, O. (2013). *Réussir son référencement web*. Eyrolles.
- AngularJS and SEO.* (s.d.). Récupéré sur Year of moo: <http://www.yearofmoo.com/2012/11/angularjs-and-seo.html>
- At a loss for words?* . (s.d.). Récupéré sur Google official Blog: <http://googleblog.blogspot.be/2008/08/at-loss-for-words.html>
- Au coeur de la recherche.* (s.d.). Récupéré sur Google: <http://www.google.com/insidesearch/features/search/knowledge.html>
- Avez-vous besoin d'un référenceur (SEO) ?* (s.d.). Récupéré sur Google support: <https://support.google.com/webmasters/answer/35291>
- Comment éviter les modifications de titre par Google.* (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20120524-11493-comment-eviter-les-modifications-de-titre-par-google.html>
- Consignes aux webmasters.* (s.d.). Récupéré sur Google support: <https://support.google.com/webmasters/answer/35769>
- Création d'un sitemap pour vidéos.* (s.d.). Récupéré sur Google support: <https://support.google.com/webmasters/answer/80472?hl=fr>
- Demandes de réexamen.* (s.d.). Récupéré sur Google support: <https://support.google.com/webmasters/answer/35843>
- Demystifying the "duplicate content penalty"* . (s.d.). Récupéré sur Google Webmaster Central Blog: <http://googlewebmastercentral.blogspot.be/2008/09/demystifying-duplicate-content-penalty.html>
- En ligne SEO Marketing: Utiliser le singulier et pluriel votre mot clé dans le titre Webpage.* (s.d.). Récupéré sur Guyot16: <https://guyot16.wordpress.com/2012/12/23/en-ligne-seo-marketing-utiliser-le-singulier-et-pluriel-votre-mot-cle-dans-le-titre-webpage>

Etude : 31% des visites d'un site web proviennent de la recherche organique. (s.d.). Récupéré sur Referenceur: <http://www.referenceur.be/etude-31-visites-site-web-proviennent-recherche-organique-13196.html>

Eye tracking. (s.d.). Récupéré sur googleblog.blogspot.de: <http://googleblog.blogspot.de/2009/02/eye-tracking-studies-more-than-meets.html>

Eye tracking. (s.d.). Récupéré sur Pr Web: <http://www.prweb.com/releases/2005/3/prweb213516.htm>

Google Adresses première source de trafic pour les sites locaux. (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20110502-10857-google-adresses-premiere-source-de-traffic-pour-les-sites-locaux.html>

Google et l'effet "sandbox". (s.d.). Récupéré sur WebRankInfo: <http://www.webrankinfo.com/dossiers/strategies-de-liens/sandbox>

Google PageRank Decrease Example - Reciprocal Link. (s.d.). Récupéré sur MyGreatName: <http://mygreatname.com/google-pagerank/pagerank-loss-by-reciprocal-link-01-1.html>

Google Suggest : nouveau jugement, cette fois favorable. (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20090727-9982-google-suggest-nouveau-jugement-cette-fois-favorable.html>

Google's 200 Ranking Factors: The Complete List. (s.d.). Récupéré sur Backlinko: <http://backlinko.com/google-ranking-factors>

HTTPS as a ranking signal . (s.d.). Récupéré sur Google Webmaster Central Blog: <http://googlewebmastercentral.blogspot.fr/2014/08/https-as-ranking-signal.html>

La pondération des liens entrants - ou backlinks - (Questions/réponses avec Google #3). (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20091218-8846-la-ponderation-des-liens-entrants-ou-backlinks-questionsreponses-avec-google-3.html>

Le référencement d'un site à page unique. (s.d.). Récupéré sur Responsive mind: <http://www.responsive-mind.fr/seo-site-page-unique/>

Le référencement, les accents, le pluriel et l'orthographe. (s.d.). Récupéré sur Axenet: <http://blog.axenet.fr/le-referencement-les-accents-le-pluriel-et-lorthographe>

Les underscores dans les URL (Questions/réponses avec Google #1). (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20091013-8825-les-underscores-dans-les-url-questionsreponses-avec-google-1.html>

Liens sitelink. (s.d.). Récupéré sur Google: <https://support.google.com/webmasters/answer/47334?hl=fr>

Link Building Case Study: How I Increased My Search Traffic by 110% in 14 Days. (s.d.). Récupéré sur Backlinko: <http://backlinko.com/skyscraper-technique>

Linkbait & linkbaiting : une tentative de traduction. (s.d.). Récupéré sur Adscriptum: <http://adscriptum.blogspot.be/2006/02/linkbait-linkbaiting-une-tentative-de.html>

Longue traine. (s.d.). Récupéré sur Olivier Corneloup: <http://www.olivier-corneloup.com/ressources/dico-referencement/definition-longue-traine>

Minty Fresh Indexing. (s.d.). Récupéré sur Matt Cutts: <https://www.mattcutts.com/blog/minty-fresh-indexing>

Modifier l'URL d'une page à l'aide d'une redirection 301. (s.d.). Récupéré sur Google support: <https://support.google.com/webmasters/answer/93633?hl=fr>

Mythes SEO décortiqués: longueur de la balise title. (s.d.). Récupéré sur Alekseo: <http://alekseo.com/longueur-balise-title/>

Netlinking. (s.d.). Récupéré sur Wikipedia: <http://fr.wikipedia.org/wiki/Netlinking>

New Google Eye Tracking Study Shows The Downfall Of The Golden Triangle. (s.d.). Récupéré sur Search engine land: <http://searchengineland.com/new-google-eye-tracking-study-shows-downfall-golden-triangle-205274>

Optimisation du contenu : travaillez votre text appeal ! (s.d.). Récupéré sur S.Billard: <http://s.billard.free.fr/referencement/index.php?2006/11/30/319-optimisation-du-contenu-travaillez-votre-text-appeal>

Overstock.com, nouveau site puni pour spam par Google. (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20110225-10719-overstock-com-nouveau-site-puni-pour-spam-par-google.html>

Performance web. (s.d.). Récupéré sur Performance web: À quoi ça sert?: <http://performance.survol.fr/2008/06/a-quoi-ca-sert/>

Plus de 4 mots clés en moyenne dans les requêtes moteurs. (s.d.). Récupéré sur Abondance: <http://www.abondance.com/actualites/20120126-11246-plus-de-4-mots-cles-en-moyenne-dans-les-requetes-moteurs.html>

Ranking factors 2014. (s.d.). Récupéré sur Search metrics: <http://www.searchmetrics.com/wp-content/uploads/seo-ranking-factors-2014.png>

Ranking Factors Checklist. (s.d.). Récupéré sur Backlinko: http://backlinko.com/wp-content/uploads/2014/05/Ranking_Factors_Checklist_Backlinko.pdf

Saisie semi-automatique. (s.d.). Récupéré sur Google support: <https://support.google.com/websearch/answer/106230>

Schema.org. (s.d.). Récupéré sur Schema.org: <http://schema.org/docs/schemas.html>

Search Queries Are Getting Longer: Hitwise Report. (s.d.). Récupéré sur Search engine land: <http://searchengineland.com/search-queries-getting-longer-16676>

SEO : La Balise TITLE dans les SERPS de Google. (s.d.). Récupéré sur Linkeo: <http://labs.linkeo.com/seo-la-balise-title-dans-les-serps-de-google/>

Seo Strategy. (s.d.). Récupéré sur Backlinko: <http://backlinko.com/seo-strategy>

The Easiest Way to a First-Page Ranking on Google. (s.d.). Récupéré sur Forrester: http://blogs.forrester.com/interactive_marketing/2009/01/the-easiest-way.html

The long tail. (s.d.). Récupéré sur The long tail: http://longtail.typepad.com/the_long_tail/2005/01/definitions_fin.html

Using site speed in web search ranking. (s.d.). Récupéré sur Google Webmaster Central Blog:
<http://googlewebmastercentral.blogspot.be/2010/04/using-site-speed-in-web-search-ranking.html>

Utiliser l'attribut "hreflang" pour afficher l'URL correspondant à une langue ou à une région. (s.d.).
Récupéré sur Google support:
<https://support.google.com/webmasters/answer/189077?hl=fr>