# ¬ Le Référencement web

Il existe des centaines de millions de sites internet dans le monde, et pourtant votre projet se doit d'être vu! Les moteurs de recherche sont une source de trafic majeur de votre site internet avec en moyenne un tiers des visites des sites web.

Il est donc indispensable que votre site apparaisse en haut des classements, notamment sur la première page de Google, quand l'internaute va rechercher le produit ou le service que vous voulez lui vendre.

Comment rendre visible votre site internet?

- 1. En définissant une orientation stratégique de réferencement
- 2. En optimisant votre site pour qu'il s'adapte aux contraintes des moteurs de recherche
- 3. En établissant une politique d'accords de partenariat et d'échanges de liens avec d'autres sites

Un bon référencement repose sur ces **3 dimensions**, qui nécessitent une optimisation continue du contenu éditorial du site, de sa structure technique, et du "net-linking", anglicisation pour cette fameuse stratégie d'échange de liens.

# ¬ Chapter 1: Définir une orientation stratégique de réferencement

L'enjeu du référencement d'un site web n'est pas véritablement de "traquer" l'internaute ou le client, mais d'abord de répondre à une problématique d'identité.

- 1. Il s'agit d'abord que l'on trouve le site lorsqu'on le recherche, ce qui implique:
  - d'être accessible le plus souvent possible via une interrogation des principaux moteurs de recherche généralistes,
  - d'être accessible systèmatiquement via une interrogation "intelligente" d'un moteur de recherche spécialisé,
  - d'être accessible via une recherche "logique" dans les annuaires généralistes,
  - d'être accessible via une recherche "structurée" dans les annuaires spécialisés.
- 2. Ensuite, le site sera "identifié", selon:
  - des éléments relatifs à l'éditeur. à la finalité du site, au contenu documentaire et aux fonctionnalités et services accessibles sur le site.
  - des éléments relatifs aux typologies des internautes auxquels le site propose ses services,
  - des éléments relatifs au champ couvert par le site (thèmes abordés, niveau de précision du traitement des sujets, etc.),
  - l'adresse du site, et les modalités d'accès (l'itinéraire de circulation dans un annuaire, ou les mots-clés utilisés dans un moteur de recherche) et les circonstances de rencontre avec un lien ("lien réponse" à une interrogation d'un outil de recherche, "lien croisé" au format "publicitaire" sur un site partenaire, etc.) amèneront l'utilisateur à identifier le site en fonction de son schéma cognitif personnel de l'architecture du web,
  - et d'autres critères plus ou moins subjectifs: performance, attractivité, pertinence par rapport aux problématiques recherchées, etc.

C'est la pertinence de ce classement qui "qualifiera" le trafic du site. Le référencement consiste donc à oeuvrer pour permettre et optimiser ce classement.

# **▼ 1 Introduction**

Ces chiffres parlent d'eux-même:

- 1,2 milliard d'internautes dont 343 millions en Europe (source: InternetWorldStats, nov 2007)
- 250 millions de recherche sur Google chaque jour (source: Google, sept 2006)
- 33% des internautes pensent qu'une société qui apparaît dans les premières positions sur les moteurs de recherche est une société leader sur son marché (source: SIAO stat)

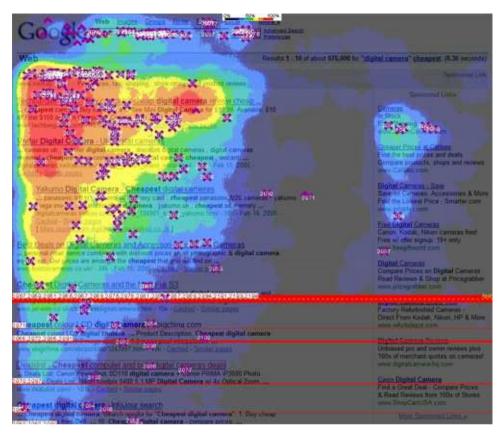
# **▼ 1.1 Le golden triangle**

Cette dernière statistique justifie que, pour être visible sur Google, il faut être placé en haut et à gauche.

Pour avoir la plus grande visibilité, il faut donc apparaître dans le **"golden triangle"**, sorte de zone chaude calculée par le survol des curseurs des internautes. Ce "triangle d'or" ne se limite pas aux résultats dits "naturels", mais bien à tout le SERP:

- les "top sponsored", ou liens sponsorisés "Google Adwords" qui apparaissent en haut à gauche,
- les "top organic results", les premiers résultats dits "naturels",
- et les résultats alternatifs de Google (liens shopping, news ou local suggestions, sitelinks, etc.).

La preuve en image:



Cette étude montre également que les résultats sponsorisés qui apparaissent en haut à gauche, les "top sponsored" sont beaucoup plus visibles, de l'ordre de 80 à 100% que les résultats sponsorisés qui apparaissent à droite, de l'ordre de 10 à 50%. Nous en reparlerons dans la partie dédiée au SEM.

# **▼ 1.2 Google**

Google est une société fondée le 7 septembre 1998 dans la Silicon Valley en Californie par Larry Page et Sergey Brin. Leur produit phare: le moteur de recherche Google. En 2008, Google annonce avoir identifié plus de mille milliard de pages web. Sa réussite lui vaut d'être copié par les plus grands: Yahoo, Live Search (Microsoft), etc. Que ce soit visuellement, par l'architecture de son SERP, ou par l'approche d'indexation, Google est régulièrement le fer de lance.

Note Aux Etats-Unis, "googler" est même devenu un verbe couramment utilisé. Je t'ai googlé signifie littéralement "j'ai tapé ton nom dans Google".

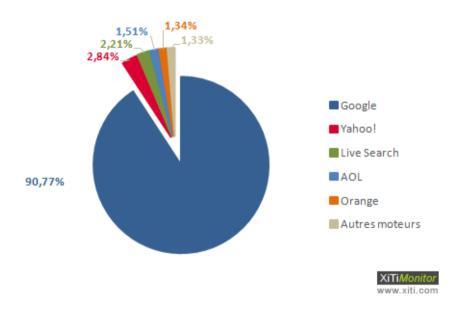
La preuve en image:



Et la notoriété de Google ne fait qu'augmenter partout dans le monde:

# TOP 5 des moteurs de recherche en parts de visites

Août 2008



### **▼ 1.2.1 L'ambition de Google**

L'ambition de Google est double:

- 1. Donner aux utilisateurs l'information la plus précise, et le plus rapidement possible.
- 2. Désindexer régulièrement les SPAM et les SCAM ("Fraude 4-1-9", type de "hoax") pour un SERP efficace.

Plus d'infos ici: http://www.google.com/corporate/tenthings.html

#### **▼ 1.2.2 Le projet Searchmash**

Pour donner aux utilisateurs l'information la plus précise, Google tend à devenir un "multi-moteurs": c'est le projet Google Searchmash http://www.searchmash.com/.



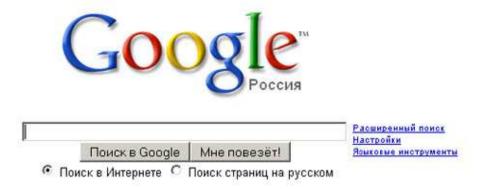
Searchmash permet à Google de tester une nouvelle interface et des fonctionnalités inédites. Car, pour l'heure, Google diffuse une large palette de moteurs, dédiés aux différents types de contenu:

### **Google Site Search**

http://google.com

Moteur de recherche dédié à la recherche de contenu sur les sites web (pages web, fichiers PDF, etc.). C'est le moteur le plus connu de la panoplie Google, et il et disponible dans 40 langues ; et parfois même des versions différentes pour chaque pays: google.ma pour le Maroc, google.fr pour la France, google.ru pour la Russie, etc.

La recherche avancée http://www.google.fr/advanced search permet de filtrer les recherches par type de contenu recherché, en recherchant des syntaxes exactes, etc.



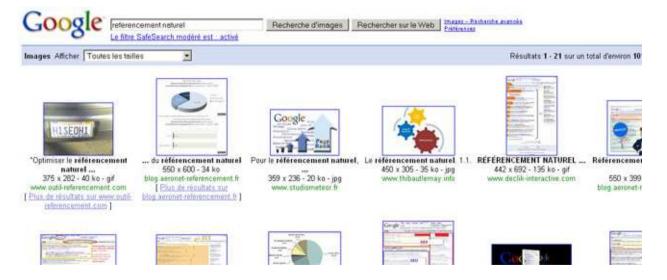
Рекламные программы - Решения для предприятий - Всё о Google - Google com in English

©2008 - Конфиденциальность

#### Google Image

http://images.google.fr/

Moteur de recherche dédié à la recherche d'images. La recherche avancée permet de filtrer par format d'image, par profondeur colorimétrique (couleurs, noir et blanc, etc.), etc.



# Google Vidéo

#### http://video.google.fr/

Moteur de recherche dédié à la recherche de vidéos. La recherche avancée permet de filtrer par domaine d'hébergement (youtube, dailymotion, etc.), par , langue, etc.



### **Google Blog**

http://blogsearch.google.fr/

Moteur de recherche indexant uniquement sur le contenu des blogs. La recherche avancée permet de filtrer, entre autre, par date de publication.



#### Knol

#### http://knol.google.com/

Projet d'encyclopédie collaborative en ligne, couvrant tous les sujets, semblable à Wikipedia, dont la politique est orientée "mise en lumière des auteurs". Selon Google, cette mise en avant aiderait significativement les utilisateurs à mieux se servir du contenu du Web.



#### Google "Recherche de codes"

http://www.google.fr/codesearch

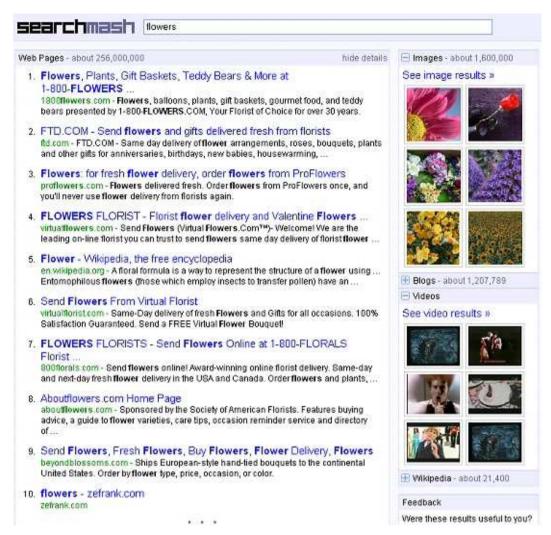
Moteur de recherche dédié à la recherche de codes informatiques: rechercher des expressions rationelles, des classes, des packages, par language, etc.



Il en existe encore de nombreux (Google Map, Google Book, Google Product Search ex-Froogle, etc.), prenez le loisir de tout découvrir ici: http://www.google.fr/intl/fr/options/

Les nouveaux outils Google sont souvent inclus dans le projet **Google Labs** <a href="http://labs.google.fr/">http://labs.google.fr/</a>, mais sont souvent, à leur sortie, limités au continent américain.. Sitons par exemple le projet **Google Ride Finder** <a href="http://labs.google.com/ridefinder">http://labs.google.com/ridefinder</a> permettant de trouver un taxi !

L'objectif de Google est donc de centraliser tous les résultats de ces moteurs dédiés sur un même SERP:



Google a développé, pour quelques-uns de ses projets, une liste de blogs, sur lequels il communique réqulièrement:

- des documentations utilisateur, lorsqu'il s'agit d'outils,
- des fonctionnalités diverses, et les évolutions de ces fonctionnalités,
- des base KM d'expérience utilisateurs, pour répondre aux moindres questionnements

Découvrez ces blogs: http://www.webrankinfo.com/actualites/200511-blogs-officiels-google.htm.

#### **▼ 1.2.2.1 L'évolution OneBox**

Le projet Searchmash se développe constamment, avec récemment, l'évolution OneBox de Google, qui permet de personnaliser les résultats en fonction des internautes!

Les auteurs du brevet "Presentation of Local Results" prennent l'exemple de la requête "Starbucks": un internaute "financier" qui requête cela dans Google espère sans doute obtenir le cours en bourse, tandis qu'un étudiant aimerait plutôt une carte lui indiquant les cafés Starbucks les plus proches de l'endroit où il se trouve, car il interroge depuis un smartphone. En plus de décider de faire apparaître des résultats spécifiques en haut de page, puisque l'encadré OneBox peut donner la météo, un cours de bourse, des horaires de cinéma, etc., Google décide parfois de faire "remonter" des résultats issus de recherches verticales (actualités, images, cartes, vidéos, blogs, scolaire, etc.) : c'est le famex Searchmash, soit la recherche universelle.



D'après ce brevet, Google décide d'afficher ces résultats non standards dans certains cas, en fonction de critères bien particuliers, incluant ceux-ci:

certains mots dans les requêtes déclenchent certains types de résultats

La requête en anglais "pizza places" est considérée comme une requête locale qui pourra faire apparaître une carte. Exemples de mots : restaurant, météo, film, itinéraire, etc.

Sur les mots clés génériques géolocalisés (hotel paris, taxi lyon, ..), Google semble également se servir de la notoriété des marques.

le type d'appareil utilisé pour faire la requête peut jouer un rôle (appareil mobile, ordinateur fixe, etc.)

La requête "pizza" effectuée sur un appareil mobile pourrait déclencher une carte listant des vendeurs de pizza dans le quartier de l'internaute faisant la requête.

• le comportement des internautes : pour chaque requête, on peut analyser les types de résultats les plus cliqués.

Par exemple pour la requête "Marilyn Monroe", si la majorité des internautes cliquent sur un résultat de type "image" alors que ce n'est pas ce qui est présenté en premier dans le SERP, alors Google peut décider d'afficher dorénavant des résultats de type "image" en top result.

• la pertinence des documents dans chaque type de résultats

Si la pertinence des documents d'un certain type (actualités, images, etc.) n'est pas assez élevée pour une requête, ce type de document n'apparaîtra pas dans les résultats pour cette requête.

l'historique associé au profil de l'internaute

Si un internaute fait majoritairement des requêtes de type "cartographiques", Google favorisera ce type de résultats pour répondre aux requêtes de cet internaute.

Note Le fait que pour une même requête les résultats soient différents d'un internaute à l'autre peut perturber les référenceurs. Il faut donc essayer de cibler davantage le profil d'internautes que vous cherchez à transformer en trafic, puis adapter le plus possible vos pages pour les satisfaire. Tout ce qui peut inciter l'internaute à revenir sur votre site est bon pour la personnalisation:

- l'ajout à un soial bookmarking,
- l'abonnement à des flux RSS,
- la richesse du contenu (images, vidéos, cartes, etc.) et des supports acceptés (version mobile du site, etc.)

#### **▼ 1.2.3 CSE "Custom Search Engine"**

Google CSE Custom Search Engine <a href="http://www.google.com/coop/cse/">http://www.google.com/coop/cse/</a> est un **moteur personnalisable**, s'inspirant de nombreux prédecesseurs (Rollyo, Eurekster, Yahoo Search Builder, etc.), et qui permet facilement d'avoir la main sur:

- le mode de recherche,
- et le format du SERP



#### Plusieurs avantages:

- limiter ses recherches sur un nombre de sites bien précis.
- limiter ses recherches sur des flux RSS ou Atom pré-sélectionnés,
- intégrer un mode de recherche personnalisée sur le contenu de son site web.
- accessoirement, gagner de l'argent, puisque CSE est intimement lié à Google AdSense, le programme de régie puclicitaire du groupe, ce qui a tendance à élargir

#### **▼ 1.2.4 La Google Dance**

La Google Dance, renommée Google Dance Jagger sur certains sites dédiés au référencement, était le nom donné à la complexe mise à jour que Google opérait entre ses bases de données. Parfois, un site pouvait disparaître jusque 48h pendant cette mise à jour, surtout s'il était nouveau, ou venait de subir une profonde refonte, avant de réapparaître dans les premières lignes des SERP. La Google Dance était "anticipable" dans les années 2002-2004 grâce à un agenda précis.

Aujourd'hui, la Google Dance est tombée aux oubliettes. Les SERP sont mis à jour quasiment en temps réél. Seules les notations PageRank peuvent prendre jusqu'à quelques mois avant d'être mises à jour.

# **▽ 2 Etablir une stratégie de communication**

Pour réussir un référencement précis, il convient d'établir les réflexions préalables suivantes:

- 1. Déterminer les annuaires et moteurs cibles, même si le choix n°1 de Google paraît évident,
- 2. Définir un titre et un résumé du site.
- 3. Définir un corpus de mots-clés
- 4. Tester ce corpus,
- 5. Optimiser les mots-clés retenus dans le corpus,
- 6. Utiliser ces mots-clés lors des demandes de soumission dans les annuaires définis,
- 7. Dispatcher ces mots-clés au sein du site, selon une trame logique.

Ces premières étapes de réflexion nécessitent l'utilisation d'outils d'aide à la décision et d'analyse concurentielle.

# **▽ 2.1 les Outils d'audit et d'aide à la décision**

Un **SIAD** est un "Système Informatique d'Aide à la Décision"; parfois, ils sont nommés BI, pour "Business Intelligence". Plus clairement, il s'agit généralement d'un logiciel d'observation, qui permet, à partir de données de gestion ou de statistiques, de donner les moyens d'identifier des phénomènes particuliers, ou de suivre l'évolution d'une activité.

#### **▼ 2.1.1 Google Trends**



Lancé en mai 2006 dans la suite Google Labs, Google Trends est un outil gratuit qui permet de **connaître les tendances de recherche** autour des termes les plus requêtés. C'est un outil très stratégique dans le choix de ses mots-clés. Bien qu'actuellement disponible uniquement en anglais et en chinois, il peut être paramètré pour des requêtes en français.

Note Google Trends est disponible à l'adresse suivante: http://www.google.com/trends.

#### **▼ 2.1.1.1 Un outil stratégique**

L'outil délivre beaucoup d'informations utiles pour le référencement des sites web. Il permet de connaître les sujets qui ont la côte à une période donnée, et de connaître les expressions à la mode. Cet outil permet de découvrir l'intérêt porté à un thème donné, au travers de données et de graphiques. Il met donc le doigt sur les tendances du web.

Et en cas d'hésitation, il indique aussi quel est celui qu'il faudra privilégier entre plusieurs mots ou groupe de mots, par exemple:

- "rideau" ou "voilage" ?
- "pain" ou "baguette" ?
- "siège" ou "fauteuil" ?
- "achat voiture" ou "acquisition voiture" ?

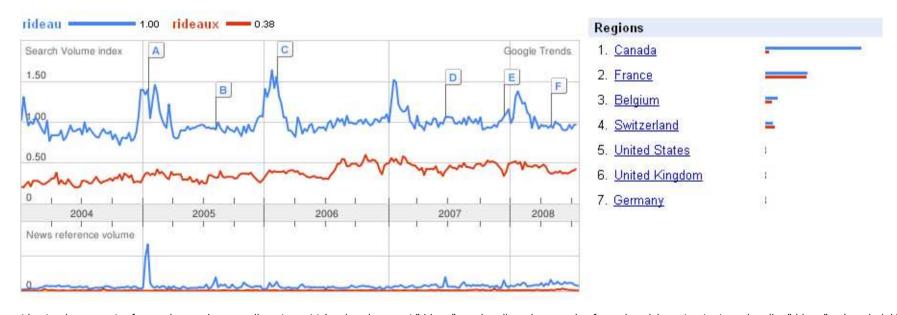
Une comparaison peut également être faite pour savoir si un mot-clé est plus utilisé au singulier ou au pluriel, avec ou sans accents, avec ou sans tiret, et si les internautes le saisissent régulièrement avec une faute d'orthographe, par exemple:

- "rideau" ou "rideaux" ?
- "chevals" ou "chevaux"
- "mot clé", "mot-clé", "mot cle" ou "mot clef" ?

De cette manière, le contenu d'un site se placera davantage en adéquation avec la demande des internautes et pourra profiter d'un trafic plus important en provenance des moteurs de recherche.

Les fréquences et les volumes de recherche peuvent être effectués simultanément sur plusieurs termes, et sont croisés à des informations de localisation, **capital pour un réferencement optimal**, à orientation marketing. On regrettera néanmoins que Google ne fournisse pas d'indication précise sur l'échelle de valeur.

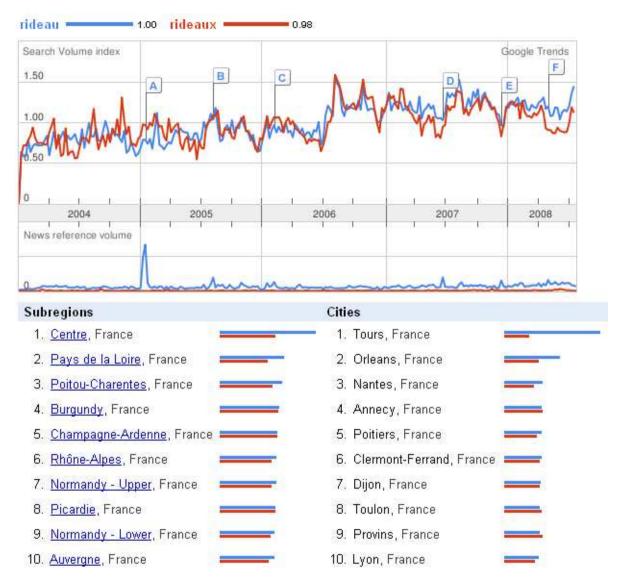
#### **▼ 2.1.1.2 Exemple n°1: aide à la décision des mots-clés**



On voit clairement les tendances entre francophones, les canadiens tapant très abondemment "rideau" au singulier, alors que les français saisissent autant au singulier "rideau" qu'au pluriel "rideaux". Le graphique du dessus nous montre toutefois que la moyenne générale est plutôt favorable à "rideau" au singulier.

Attention dans l'interprétation des chiffres, il semblerait que Google se base sur les requêtes au sens large. Cela signifie par exemple qu'en comparant "rideau" et "rideaux", on compare en réalité l'ensemble des requêtes utilisant le mot "rideaux" à l'ensemble des requêtes utilisant le mot "rideaux".

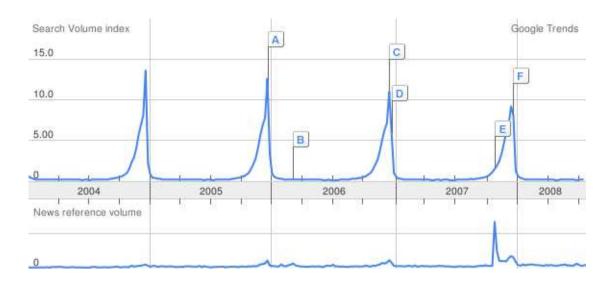
Et on peut aller encore plus loin dans la "géolocalisation", en cliquant sur "France", on obtient des détails, par région et même par villes:



Google Trends permet d'avoir une vision très internationale ou très locale. Pour optimiser son réferencement, il s'agit donc, avant tout, de bien déterminer sa cible!

#### **▼ 2.1.1.3 Exemple n°2: faire un choix stratégique**

Pour se préparer aux grands pics d'activités, il est préférable de connaître quand commence la saisonnalité d'un thème ou d'un produit, pour mieux anticiper la demande en provenance des moteurs. Ainsi, pour le mot clé "noël", les requêtes ont commencé à s'amplifier à partir du mois d'octobre, chaque année. Un site marchand doit donc préparer son contenu dès la rentrée, à la fin septembre, pour pouvoir mieux se placer sur les moteurs et répondre dès le départ à cette demande!



#### **▼ 2.1.1.4 Les "Hot trends"**

Pour vérifier les tendances, 2 modes d'utilisation:

- Si le mot clé est déjà déterminé: il suffit d'entrer ce dernier dans la zone de recherche.
- Sinon, il est possible de consulter directement les mots-clés du moment dans la liste "Hot Trends", fournie dès l'accueil de l'outil. Cette liste est remise à jour toutes les 24 heures. Pour la constituer, Google a choisi de ne montrer que les mots-clés qui connaissent une poussée soudaine en popularité, plutôt que l'ensemble des recherches les plus populaires où figureraient souvent les mêmes termes génériques.

Cette liste concerne pour l'instant uniquement le continent américain.

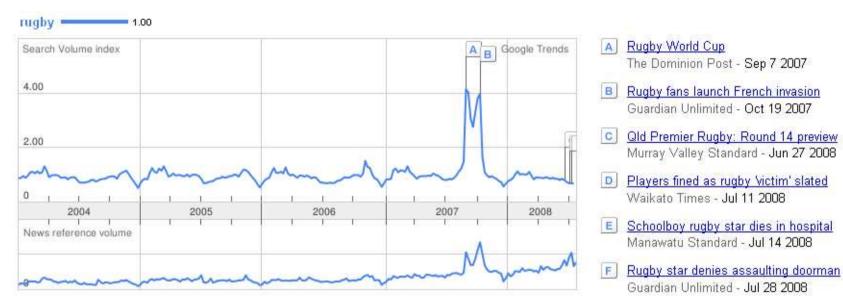
Today's Hot Trends (USA)				
1. <u>diane borchardt</u>	6. <u>acrylamide</u>			
2. <u>chiguire</u>	7. <u>seafair</u>			
3. beach to beacon	8. boomer grigsby			
4. dinotopia	9. <u>flugtag portland</u>			
5. <u>bobby sippio</u>	10. <u>spurning</u>			
	More Hot Trends »			

La liste complète est disponible ici: <a href="http://www.google.com/trends/hottrends">http://www.google.com/trends/hottrends</a>.

#### **2.1.1.5** L'intéraction des contenus

Autre fonctionnalité intéressante, la présentation d'articles d'actualités **directement associés aux pics de popularité** du mot-clé. Qui permet d'ailleurs de se rendre compte de l'interaction entre le moteur de recherche Google, et son site d'actualités Google News.

Par exemple, pour la coupe du monde de rugby en 2007, on remarque un pic similaire entre les recherches dans le moteur et le nombre d'articles de presse couvrant les matches. Et en plus des dernières actualités de Google News, l'internaute peut accéder à toutes les archives conservées par le moteur. Issues la plupart du temps de journaux américains comme le New York Times, ces informations permettent de remonter très loin et même de retrouver des informations datant du 18ème siècle!



#### **₹ 2.1.1.6 La recherche avancée**

Parmi les autres possibilités de recherche, Google Trends propose les fonctionnalités suivantes:

l'utilisation de guillemets permet d'effectuer une recherche sur une expression précise,

ex: "rideau de douche"

pour exclure des termes, utiliser le tiret "-",

ex: python -serpent

pour cadrer sa recherche sur le language de programmation python en excluant le serpent du même nom ;

• pour voir combien de requêtes contiennent au moins un des deux termes, il suffit de les séparer par la touche "pipe",

ex: pavillon|villa

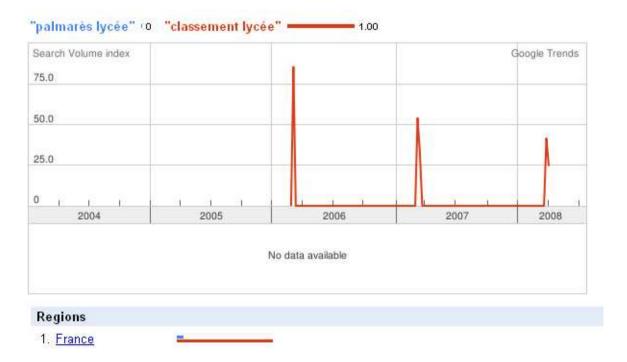
#### **₹ 2.1.1.7** Les limites de l'outil

Encore au stade de développement, Google Trends montre parfois certaines limites sur les requêtes en langue française. Par exemple, le nombre de requêtes dans Google sur l'expression "palmarès lycée" n'est pas suffisant pour que la comparaison puisse aboutir. Alors qu'en faisant un comparatif entre les expressions "palmarès lycée" et "classement lycée", on voit qu'il y aurait eu des résultats sur "palmarès lycée"!

Your terms - "palmarès lycée", "palmares lycee" - do not have enough search volume to show graphs.

# Suggestions:

- Make sure all words are spelled correctly.
- · Try different keywords.
- Try more general keywords.
- · Try fewer keywords.
- · Try viewing data for all years and all regions.



# **2.1.1.8** Un analyseur de trafic

Google Trends a accueilli récemment une nouvelle fonctionnalité: le Google Trends for Websites, qui permet de comparer la popularité de différents sites web et blogs.

Se basant sur le nombre de visiteurs uniques, le volume de recherches, les données fournies par Google Analytics et d'autres paramètres non divulguées, Google Trends for Websites vous permet de visionner en un clin d'oeil le nombre de visiteurs uniques via un graphique.

Ce service vous informe également des principales régions d'ou proviennent les visiteurs des sites comparés, ainsi que les autres sites visités par ces mêmes visiteurs et enfin les mots clés/requêtes ayant généré le plus de trafic sur ces sites. Il s'agit donc d'un véritable outil dédié à l'analyse concurrentielle.

Google Trends for Websites étant encore en phase de test, Google annonce clairement que ces résultats sont à prendre avec un certain recul. Certainement pour quelques unes des raisons suivantes:

- les sites web qui enregistrent un trafic trop faible, sont difficilement identifiables;
- tous les sites, surtout les nouveaux, ne sont pas encore dans la base Google;
- certains sites ne respectent pas la charte de qualité de Google: http://www.google.com/support/webmaster/bin/answer.py/?answer=35769.

#### **₹ 2.1.1.9 Alexa**

Sur la base de cette dernière fonctionnalité d'analyse du trafic des sites, Google va plus loin que l'outil Alexa, fourni par Amazon:



Note Alexa est disponible à l'adresse suivante: http://www.alexa.com.

Alexa réalise un classement pour tous les sites du web mondial, appelé "Traffic rank". Le web comptait 166 millions de sites internet en avril 2008 (source Netcraft). Alexa fournit des données chiffrées pour les sites classés dans les 25 premiers millions. Le classement d'Alexa est réalisé en fonction du nombre de visiteurs uniques et du nombre de pages vues par chaque visiteur. Le calcul de cette moyenne sur les trois derniers mois déterminent le "Traffic rank".

Toutes les extensions d'une adresse internet sont prises en compte et sont comptabilisées pour le même site. Ex: mail.example.com, example.com et example.fr seront considérés comme 1 seul et même site, à condition que les extensions soient gérés par des redirections permanentes, et qu'il ne s'agisse pas de sites différents.

Un peu à l'image de Google Trends for Websites. Alexa fournie égalment des statistiques détaillées par zone géographique, des "Hot Trends", etc.

Enfin, Alexa propose de télécharger différentes barres d'extension, pour obtenir des informations en temps réél des sites visités:

Sparky, la barre de statut officielle,

téléchargeable ici: http://sparky.mozdev.org/

SearchStatus, montre le classement d'un site sur Google et son classement trafic d'Alexa,

téléchargeable ici: http://www.quirk.biz/searchstatus/

About This Site, un plug-in Firefox qui montre les metadonnées du classement trafic d'Alexa.

téléchargeable ici: http://www.ginatrapani.org/workshop/firefox/aboutsite/

Un autre plugin nous permet d'accéder au Google PageRank, au classement Alexa, et à celui de Compete simultanément: **SearchStatus** <a href="https://addons.mozilla.org/fr/firefox/addon/321?application=firefox&version=1.0&os=Windows&numpg=10&id=32">https://addons.mozilla.org/fr/firefox/addon/321?application=firefox&version=1.0&os=Windows&numpg=10&id=32</a>

#### **▼ 2.1.1.10 Compete**



Lancé en 2000, Compete est un autre outil gratuit dans la lignée de Google Trends for Websites, qui permet d'analyser le trafic de différents sites web et blogs. Un espace pro payant permet d'accéder à différentes fonctionnalités, par exemple de remonter plus loin dans l'historique du trafic, par défaut limité à la dernière année.

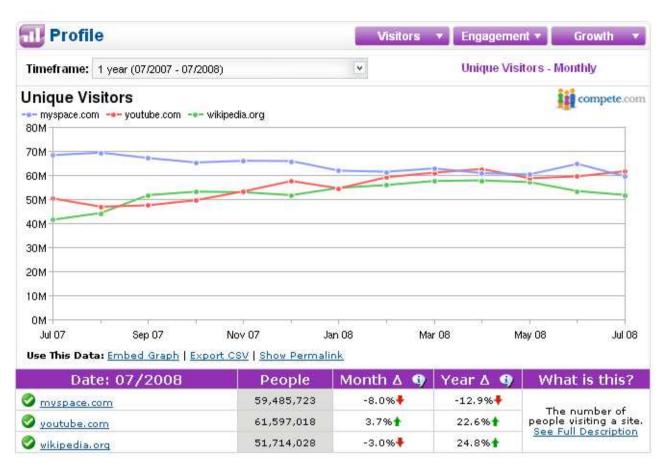
Cet outil se limite toutefois au trafic en provenance des Etats-Unis.

Note Compete est disponible exclusivement en anglais à l'adresse suivante: http://www.compete.com/

Plus précis que Google Trends for Websites. Compete indique une échelle de valeur et permet d'analyser et de comparer des sites selon différents paramètres:

- Comparaison du nombre de visiteurs uniques par mois,
- Comparaison du nombres de visites par mois, un visiteur pouvant revenir sur le site dans le mois en cours,
- Comparaison du classement dans le top million des sites US,
- Comparaison du temps moyen passé sur le site, par jour ou par mois, en %,
- Comparaison du nombre de pages, par visite, et par mois

L'espace pro permet d'obtenir d'autres statistiques encore plus abouties: comparaison du nombres de pages vues par mois, etc.



Sous le graphe, le lien "Embed Graph" permet de l'afficher et de le mettre à jour en temps réél sur un site distant.

Compete permet de connaître également les mots-clés qui ont permis aux internautes de trouver le site sur un moteur de recherche. Compete permet de connaître le nombre de mots-clés ayant créé un trafic entrant, et surtout d'obtenir cette liste de mots. Le mode gratuit ne permettant d'en connaître que les 5 premiers, le mode pro permettant d'obtenir toute la liste!

Report Overview	
Domain / Category	kayak,com
Time frame	2008-05-28 to 2008-08-26
Available results	5,066 terms
Displaying	1 - 5   Access: <u>all 5,066 terms</u>

Volume Rank 🗓	Keyword 🗓	% of Site's Search Traffic j	Average Time Index 🗓	Total Time Index i)
1	<u>kayak</u>	19.77%	60.17	100.00
2	kayak.com	9.13%	69.77	53.56
3	www.kayak.com	1.70%	65.81	9,40
4	kayak.com flights	1.04%	100.00	8.78
5	kayak travel	1.03%	57,91	5.00

On peut également obtenir les mots-clés les plus utilisés, par thèmes et/ou activités,

Report Overview	
Domain / Category	Shopping >> Home and Garden
Time frame	2008-05-28 to 2008-08-26
Available results	100,000 terms
Displaying	1 - 5   Access: <u>all 100,000 terms</u>

Volume Rank 1)	Keyword 🤄	% of Category's Search Traffic i	Average Time Index 1/	Total Time Index i
1	home depot	5.44%	57,90	81.16
2	lowes	4.80%	80.89	100.00
3	ikea	2,23%	100.00	57,38
4	bed bath and beyond	2.04%	78.64	41.25
5	homedepot	0.73%	57.29	10.72

connaître les sites qui ont généré le plus de trafic en provenance de ces mots-clés,

Report Overview	v	
Search term	home depot (exact match)	
Time frame	2008-05-28 to 2008-08-26	
Available results	625 domains	
Displaying	1 - 5   Access: all 625 domains	

Volume Rank 1)	Destination Website (j)	% Total Keyword Volume _i)	% of Site's Search Traffic ()	Average Monthly Search Referrals
1	homedepot.com	90.75%	40.96%	2,381,127
2	lowes.com	0.81%	0.28%	2,726,925
3	anrdoezrs.net	0.52%	5.73%	97,728
4	citibank.com	0.40%	0.50%	1,020,038
5	yahoo.com	0.35%	0.00%	130,390,593

et comparer les avantages des 2 sites sur des mots-clés stratégiques:

Report Overview		
Domains	facebook.com / orkut.com	
Time frame	2008-05-28 to 2008-08-26	
Available results	100,000 terms referring to facebook.com or orkut.com	
Displaying	1 - 5   Access: all 100,000 terms	

Volume Rank 🗓	The second secon	Keyword Volume Advantage (j)	Average Time Advantage 🗓
1	facebook	>1,000%	564%
2	facebook.com	>1,000%	>1,000%
3	www.facebook.com	>1,000%	>1,000%
4	facebook login	>1,000%	>1,000%
5	face book	>1,000%	>1,000%

# **▼ 2.1.1.11 SEOmeter**



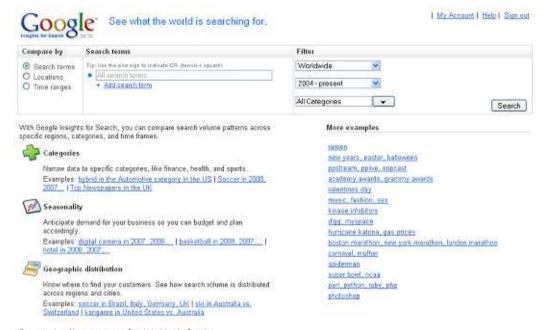
Lancé en 2008, SEOMeter est un autre outil gratuit, qui permet d'analyser le CC "Crawl Cycle" du Googlebot, c'est à dire de savoir quand est-ce que Google a parcouru les différents sites web et blogs, et de connaître l'espace temps entre ces crawls.

Note SEOMeter est disponible exclusivement en anglais à l'adresse suivante: <a href="http://www.seometer.com/">http://www.seometer.com/</a>



Il est possible de comparer jusqu'à 3 sites simultanément. Et d'afficher un widget sur son site, permettant de suivre à distance le CC.

# **▼ 2.1.1.12 Google Insights for Search**



See examples of how you can use Google Insights for Search.

Note Google Insights for Search est disponible à l'adresse suivante: http://www.google.com/insights/search/.

Pour aller encore plus loins dans l'analyse des comportements, Google a mis en place en 2008 le **Google Insights for Search**. Encore plus précis que son cousin Google Trends, il fournit un contenu plus étoffé, malgré tout à l'objectif semblable, et **mis à jour tous les jours**.

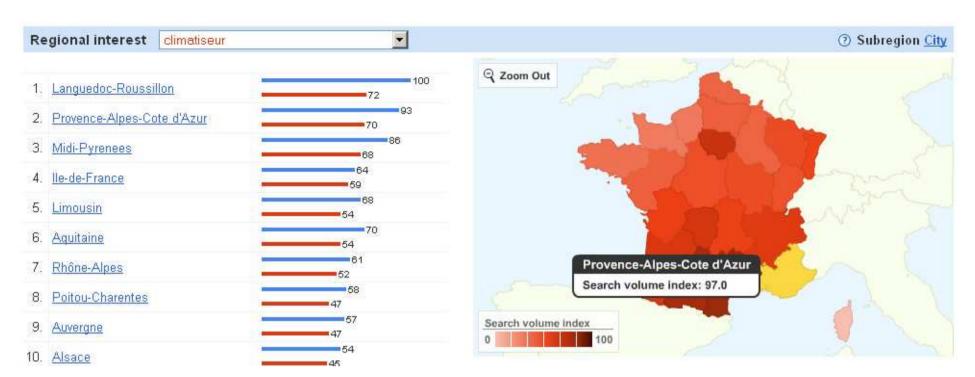
Toutefois, les valeurs demeurent "normées", c'est-à-dire que les chiffres réels de volumes de recherche sont modifiés afin de faciliter les comparaisons. Les valeurs indiquées sont calibrées entre 0 et 100. 100 correspond au record de trafic journalier d'une requête, les autres valeurs sont donc un indice représentant le volume de recherche journalier.

Quelques nouvelles fonctionnalités intéressantes:

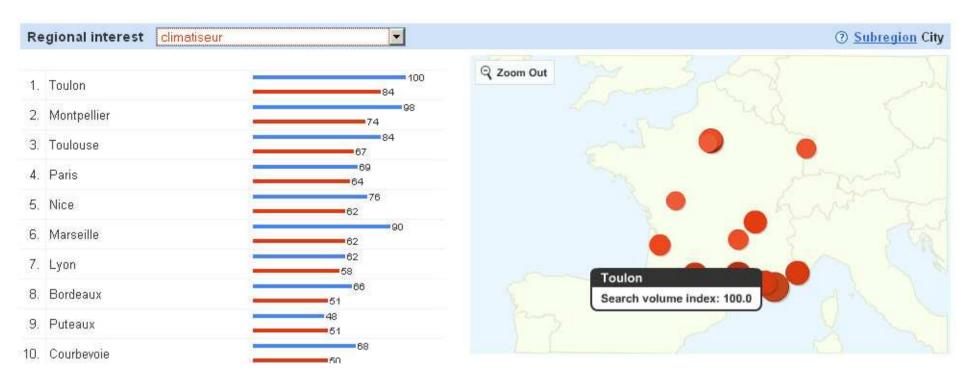
• Une correlation du volume de recherche, avec une carte du monde interactive, par pays,



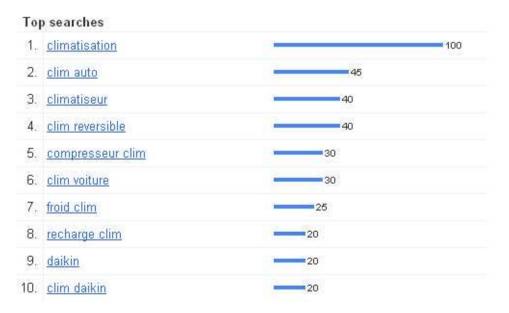
par région,



par ville,



• un Top 10 des requêtes contenant le(s) terme(s) "insightés",



• un Top 10 des expressions-clés contenant les termes audités, et qui ont connu des pics de tendances: les "Rising searches",

Ris	ing searches	
1.	winning eleven 10	Breakout
2.	sensible soccer 2006	Breakout
3.	pes6	Breakout
4.	pes 6	Breakout
5.	fifa 2007	Breakout
6.	soccer aid	+930%
7.	2006 world soccer	+580%
8.	soccer world cup	+280%
9.	worldcup soccer	+250%
10.	world cup 2006	+250%

Ces résultats sont calculés entre la période d'analyse sélectionnée et celle qui la précède. Par exemple, si la période sélectionnée est l'année 2007, la comparaison est faite par rapport à l'année 2006; si la période est Juin 2007, la comparaison est faite par rapport à Mai 2007. Par ailleurs, un terme fait son apparition ici: "Breakout", se dit des mots-clés connexes ayant une très forte augmentation (soit supérieure à 5000%).

• On peut également comparer les données sur plusieurs années. Et même distinguer des termes par "catégorie":

# Search Volume: France vs. United States vs. Japan for Movies, Last 30 days



Il ne reste plus qu'un atout manquant à Google Insights pour être devant Google Trends: l'analyseur de trafic.

#### **▼ 2.1.1.13 Conclusion**

Google Trends est donc un outil qui s'avère particulièrement utile pour auditer des sites de contenu. Il permet notamment aux équipes éditoriales de **sélectionner les bons mots et expressions clés** et d'en connaître **les alternatives de positionnement**. L'outil reste cependant assez limité lorsque le nombre de requêtes est trop faible.

### **▼ 2.1.2 Google Keyword Sandbox**

Nous avons vu que Google Trends vous permettait de répondre à la question suivante: "Dois-je réferencer mon site sur "achat de voiture" ou "acquisition de voiture" ?". Google n'a pas fini de nous aider dans la création d'un contenu de qualité, puisqu'il nous propose cette fois-ci le **Google Keyword Sandbox**, parfois nommé Google Keyword Tool.

Note Google Keyword Tool est disponible à l'adresse suivante: https://adwords.google.com/select/KeywordToolExternal.

Ce Générateur de mots clés recense de façon automatique une liste de requêtes courantes effectuées sur le moteur de recherche de Google, en rapport avec le mot clé que vous avez saisi. Il permet

d'avoir des estimations chiffrées des volumes de recherche des internautes du mois précédent, et en moyenne. Il fournit également une jauge indicative du niveau de concurrence des liens sponsorisés Adwords sur le(s) terme(s) en question. Et, au vu des parts de marché de Google, il a l'avantage de fournir des données récentes et largement représentatives.

Concurrence entre annonc	eurs Volume de recherche approximatif	Volume de recherche moyen
Mots clés en rapport avec le(s) term	ne(s) entré(s) - trié par pertinence ②	
thé vert	<u> </u>	170
bienfaits du thé vert	Données insuffisantes	480
thé vert en	Données insuffisantes	480
de thé vert	Données insuffisantes	1 900
thé vert de chine	Données insuffisantes	Données insuffisantes

Utilisez-le pour obtenir de nouvelles suggestions de mots clés et tester vox choix.

On regrettera néanmoins que Google n'ait pas (encore ?) développé de fonction d'exportation dans un format de transport plus pratique que les .CSV (on ne peut même pas exporter le contenu via PDF Creator, puisque les données s'affichent dans une iframe), et surtout que les fonctions d'export actuelles se limitent à la liste des mots-clés.

Yahoo Search Marketing, ex-Overture, a diffusé pendant longtemps un outil similaire mais il a été coupé récemment.

D'autres outils sont néanmoins listés ici: <a href="http://www.keywordbox.com/article-138.html">http://www.keywordbox.com/article-138.html</a> et vous trouverez un ouvrage intéressant sur le sujet ici: le SEO Book Keyword Research Tool, disponible à l'adresse: <a href="http://tools.seobook.com/keyword-tools/seobook/">http://tools.seobook.com/keyword-tools/seobook/</a>

### **▼ 2.1.2.1 Les termes protégés**

Dans l'introduction de son outil, Google précise que "Avant de sélectionner un mot clé, vous devriez vérifier qu'il ne s'agit pas d'un terme protégé (marque, nom commercial, dénomination sociale) en consultant un registre des marques (ex: www.icimarques.com) et des sociétés (ex: www.euridile.com). Vérifiez bien, car on est souvent confronté à des surprises de taille!

#### ▼ 2.1.2.2 La prise en compte du singulier et du pluriel

Les moteurs repèrent la racine commune du mot, le "stemming". Si bien qu'en utilisant le singulier sur une page, on peut sortir dans les résultats pour une requête au pluriel, et réciproquement.

Yahoo s'est vu attribué un brevet, le "Word pluralization handling in query for web search", le 7 août 2008. Ce système construit une sorte de dictionnaire des mots concernés (ceux qui prennent une forme différente au singulier et au pluriel, et qui valent la peine d'être étudiés). Les mots retenus pour cette analyse sont déterminés en cherchant dans la requête le dernier mot qui ne soit pas un stop-word (ou mot vide, c'est-à-dire un mot non porteur de sens, qui n'apporte rien ou presque à la requête, comme par exemple "le" dans la requête "le Pape Benoit XVI"). D'après les auteurs du brevet, environ la moitié des requêtes des internautes comportent au moins un mot qui peut ainsi être transformé entre le singulier et le pluriel. Le brevet prévoit donc que le moteur puisse estimer si cela vaut le coup d'afficher des résultats qui auraient été sélectionnés si on avait utilisé le pluriel, et fait alors un savant mélange des 2 groupes de résultats. Est-ce que le résultat en sera bénéfique ? Trop tôt pour le dire.

D'un point de vue réferencement, que faut-il en conclure pour le choix des mots-clés ? Les internautes font majoritairement leurs requêtes au singulier ; autant donc privilégier la forme utilisée par ceux qui vous cherchent Teste-le au préalable grâce aux outils cités ici.

Nous verrons plus loin le balisage précis à suivre au sein de vos pages web. Mais préférez les endroits les plus stratégiques de votre site (titre de la page, titre des rubriques, backlinks) pour utiliser la

forme prépondérante et utilisez la forme secondaire à d'autres endroits de la page.

#### **▼ 2.1.2.3** La co-occurrence et le ratio E/F

La **co-occurence** est un indice exprimé en ppt "Points Per Thousand" qui **mesure le relation entre les termes**: plus cet indice est élevé, plus les termes semblent proches sémantiquement parlant, quel que soit l'ordre des mots. Cela se traduit généralement par une forte concurrence dans Google.

Sur le même principe, mais cette fois avec le respect de l'ordre des mots, plus le ratio E/F est fort, plus il est fréquent de trouver sur le web les mots dans cet ordre.

Vous pouvez utiliser cet outil pour tester les co-occurences: http://www.webrankinfo.com/outils/co-occurrence.php

Termes	Indice de co-occurrence	Distance	Ratio EF
test bonjour resultat	1,1 ppt	6,778	0,0 ppt
test bonjour	6,5 ppt	E 027	0,5 ppt
bonjour test		5,037	0,3 ppt
bonjour resultat	30,2 ppt	3,499	0,0 ppt
resultat bonjour			0,1 ppt
test resultat	0404094620240	4 420	0,9 ppt
resultat test	11,9 ppt	4,430	14,3 ppt

Google permet de compléter ce test de co-occurence, en connaissant le nombre de résultat et le niveau de concurrence:

- "allinanchor: bonjour test", ne va afficher que les résultats dont les backlinks vers les snippets affichés contiennent les mots-clés recherchés,
- "allintitle: bonjour test", ne va afficher que les résultats dont le titre contient les mots-clés recherchés,
- "allintitle: bonjour test", ne va afficher que les résultats dont le descriptif contient les mots-clés recherchés,

#### **▼ 2.1.2.4 La recherche par contenu de site web**

Une autre très bonne fonctionnalité est disponible dès la page d'accueil de GKT: un outil d'analyse concurentielle qui permet de générer une liste de mots-clés, en se basant sur un site web!

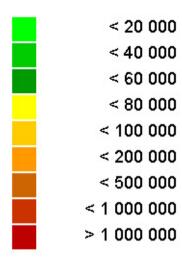
Le reporting va ainsi se construire en recherchant des mots-cls avec le contenu de la page choisie:

Affichage des mots clés regroupés en fonction des termes suivants : decathlon convertible car seat (6), ski equipment (18), car seat (5), ski boots (10), sport equipment (10), mountain climbing (6), ski rental (5), ski (37), sport (25), equipment (8), decathlon (6), gear (5), Mots clés divers (11)

Mots clés	Concurrence entre annonceurs ②	Volume de recherche approximatif : août ②	Volume de recherche moyen approximatif ②	Type de ciblage : 👩
				Large
Mots clés concernant decathlon conv	vertible car seat - trié par per	tinence ②	-4040	
decathlon convertible car seat		Données insuffisantes	6 600	<u>Ajouter</u> 🍣
britax decathlon convertible car seat		Données insuffisantes	5 400	Ajouter 💝
britax decathlon convertible car seat hampton		Données insuffisantes	Données insuffisantes	<u>Ajouter</u> 🍣
britax decathlon convertible car seat onyx		Données insuffisantes	1 300	<u>Ajouter</u> ×
britax decathlon convertible car seat tiffany		Données insuffisantes	1 000	<u>Ajouter</u> 💝
decathlon convertible car seat tiffany by britax		Données insuffisantes	Données insuffisantes	<u>Ajouter</u> 💝
				Tout ajouter - 6 »
Mots clés concernant ski equipment	- trié par pertinence 🕐			
skiing equipment		Données insuffisantes	368 000	<u>Ajouter</u> 💝
crested butte ski equipment rental		Données insuffisantes	Données insuffisantes	<u>Ajouter</u>
used ski equipment		Données insuffisantes	2 900	Ajouter 💝
ski equipment rental		Données insuffisantes	2 900	<u> Ajouter</u> 💝
ski equipment rentals		Données insuffisantes	1 000	Ajouter 💝

# **▼ 2.1.2.5 Degré de difficulté d'indexation**

Si vous devenez réferenceur professionnel, vous entendrez régulièrement vos clients vous demander de les positionner sur des termes très génériques ("sport", "tourisme", etc.). Malheureusement, les places sont souvent déjà prises, et très chères! Pour vous aidez dans vos raisonnements, voici une **grille estimative** de la difficulté de s'indexer sur la 1ère page du SERP de Google:



Cette grille n'est pas une vérité absolue, car elle ne tient compte que des paramètres de quantité, mais elle peut être déterminante dans la majorité des cas.

# **▼ 2.1.2.6 Google Suggest**



Lancé en 2004, dans la suite Google Labs, Google Suggest est un outil gratuit qui vous présente des **suggestions de mots-clés**, à mesure que vous saisissez du texte dans le champ de recherche. Aujourd'hui, ces suggestions sont activées dès l'interface du moteur de recherche Google.

Note Google Suggest est disponible à l'adresse suivante: http://www.google.com/webhp?complete=1.

Ces suggestions sont déterminées en fonction du nombre de résultats connus par le moteur Google, mais surtout par le **nombre de requêtes effectuées par les internautes**. Elles risquent donc d'être encore davantage favorisées, puisque ces suggestions sont faites pour être sélectionnées!

deve	
developpez	2,370,000 résultats
developpement photo	2,600,000 résultats
developpement photos	883,000 résultats
devendra banhart	1,780,000 résultats
développement durable	9,140,000 résultats
devenir riche	482,000 résultats
developper photo	4,060,000 résultats
devenir mannequin	375,000 résultats
devenir avocat	255,000 résultats
devenir gendarme	286,000 résultats
ne reconstruit de la company d	fermer

# Une évolution intéressante dans la suggestion: Xippee

Xippee est un plug-in pour différents browser, et offre une aide très intelligente pour accélérer vos recherches dans Google notamment. Découvrez une vidéo d'utilisation, sur le site officiel: <a href="http://www.xippee.com">http://www.xippee.com</a>.



# Yahoo! Inquisitor

Yahoo! avait déjà développé "Search Assist", un équivalent du Google Suggest. Et depuis Mai 2008, après le rachat du plugin Inquisitor, il a diffusé un plugin pour Safari, puis Firefox et Internex Explorer. Au programme toujours, des requêtes possibles commençant par les lettres demandées, au fur et à mesure de la saisie, mais également des propositions de sites web, parfois en fonction des sites dernièrement visités par l'utilisateur! A télécharger ici: http://www.inquisitorx.com/firefox/index en.php



#### **₹ 2.1.3 Website Grader**



Website Grader est un outil qui permet d'auditer synthétiquement le SEO de votre site, ou de votre blog.

Note Website Grader est disponible à l'adresse suivante: http://www.websitegrader.com/.

Website Grader dresse un reporting, comprennant plusieurs statistiques et détails instructifs, même si pour la plupart on peut les calculer très facilement autrement:

• le PageRank, que l'on peut simplement obtenir par la Googlebar <a href="http://toolbar.google.com/">http://toolbar.google.com/</a>



- les métadonnées.
- le nombre de liens,
- le nombre d'images,
- le domaine,
- l'aspect social marketing

Cet outil est toutefois à utiliser avec grande prudence, car "audit automatisé" signifie que beaucoup de paramètres "naturels" et humains ne sont pas pris en compte:

- pas d'analyse sémantique, à proprement parler
- pas d'analyse en fonction des objectifs fixés
- etc.

Mais cela répond à la problématique du webmaster ou du bloggueur qui souhaite faire un point rapide et synthétique sur son positionnement SEO, et également avoir quelques conseils rapides. En quelques sortes, cet outil permet essentiellement de se rendre compte si son site ou son blog suit les **directives générales** en terme de réferencement naturel.

#### **₹ 2.1.3.1 Les recommandations**

Toutefois, les recommandations et données fournies par le reporting de Website Grader donne des pistes, selon différents aspects:

# La pertinence du site web

- il vérifie la saisie et la taille des balises meta title, description, et keywords,
- il vérifie et les liste les headings avec leur niveau de hiérarchie,
- il vérifie la présence d'images, et si elles sont correctement acompagnées de texte ALT,
- il vérifie quelques pages choisies de façon aléatoire et teste leurs metaéléments.

Enfin, il détermine une note, discutable, qu'il attribue au "niveau de lisibilité" du site.

#### La notoriété du site web

- il vérifie si le nom de domaine a une validité d'au moins 1 an,
- il calcule le PageRank de Google
- il détermine le nom de pages indexées dans Google

Note Vous pouvez tester cela facilement pour n'importe quelle URL. Rendez-vous sur Google, testez la requête suivante: "site:wikipedia.com". Seules les pages du nom de domaine "wikipedia.com" s'afficheront dans le SERP. Avec à droite de l'écran, le nombre de pages indexées dans Google.

Résultats 1 - 10 sur un total d'environ 222 sur wikipedia.com (0,14 secondes)

• il détermine le nom de pages indexées dans DMOZ

Note DMOZ <a href="http://dmoz.com">http://dmoz.com</a>, ou "Open Directory Project" est un annuaire communautaire entièrement gratuit, qui fait l'interaction entre des contenus thématiquement proches, en faisant abstraction de la langue. Sachez que l'annuaire de Google <a href="http://www.google.com/dirhp">http://www.google.com/dirhp</a> complète ses données grâce à DMOZ, en triant ses données par PageRank et non par ordre alphabétique. Il est fort intéressant de s'y enregistrer!



- il détermine la date du dernier crawl de Google sur le site
- il détermine le nombre de "inbound links", les liens externes pointant sur le site.

#### La notoriété de votre site, sur la blogosphère

- il vérifie si le site possède un blog
- il détermine le ranking du blog sur le moteur Technorati. C'est d'ailleurs dommage que d'autres outils ne soient pas pris en compte.

Note Les Blogs ont leurs propres moteurs de recherches et leurs propres annuaires:

- □ Google Blog http://blogsearch.google.fr/
- □ Technorati http://technorati.com/
- ☐ Wikio http://www.wikio.fr/
- etc.
- il vérifie la présence d'un flux RSS sur votre blog. Et détermine le nombre d'inscrits, qui le suivent depuis l'agrégateur BlogLines. A nouveau on peut regretter qu'un agrégateur comme Feeburner (appartenant à Google) ne soit pas cité ici.

#### La notoriété de votre site, sur les services sociaux web 2.0

• il se limite simplement à déterminer le nombre de bookmarks Del.icio.us pointant chez vous, et le nombre d'articles à votre sujet stockés sur Digg.com.

Note Partager vos bookmarks sur des "Sharewares", tels que <u>Del.icio.us</u>, <u>Ma.gnolia</u>, <u>Mister Wong</u>, <u>Yoolink</u>, etc., créer une fiche sur des "Groupwares", tels que <u>Facebook</u>, Twitter, Linkdln, Ziki, etc. optimisera effectivement la notoriété de votre site. Et tout cela gratuitement!

#### Votre capacité à qualifier l'internaute en lead...

• ...simplement en vérifiant si votre site diffuse un formulaire de contact. Là encore, simplement avoir un formulaire ne va pas tout faire, encore faut-il déterminer s'il est facile d'accès, et compatible sur tous les browser par exemple !

#### La notoriété de votre site, sur des critères compétitifs

• il vise quant à lui à faire un focus sur l'historique des données et à dresser un petit tableau comparatif, suite aux sites "concurrents" saisis en amont, lors de la validation de l'outil.

De nombreux outils sur internet vont vous permettre d'effectuer d'autres comparatifs entre vous et vos concurrents, par exemple:

# Dresser un tableau comparatif "PageRank + Age du Domaine + Backlink Yahoo"

http://www.referencement-page1.fr/classement-google-pr/

POSITION SUR GOOGLE	‡ URL.	PAGERANK	‡ AGE	\$YAHOO BACKLINKS
01	http://www.climshop.com/	4	2004	52826
02	http://fr.wikipedia.org/wiki/Climatisation	5	2004	166
03	http://www.toshibaclim.com/	3	2003	574
04	http://www.atlantic-climatisation.com/	4	2002	610
05	http://www.ubaldi.com/electromenager/climatisation/climatisation.php	4	2007	43
06	http://www.daikin.fr/	4	2002	1197
07	http://www.travaux.com/dossier/climatisation.html	3	2003	44
08	http://www.edf.fr/accueil/conseils-pratiques/mes-equipements/chauffage-et-climatisation-141138.html	3	2003	4

Lors de la soumission d'un audit sous Website Grader, l'outil détermine une note globale, suite aux points évoqués. A nouveau, on peut se poser la pertinence de cet indice, puisqu'il se limite essentiellement à des critères d'ordres techniques.

#### **▼ 2.1.3.2 Les audits archivés**

Website Grader stocke la plupart des actions effectuées par les internautes. Ce qui permet d'obtenir des audits intéressants, sur des thématiques diverses. Le contenu de ces thématiques n'intéressera pas forcément les moins curieux, mais ce qui est réellement intéressant das notre cas actuel, c'est le type d'informations qu'il nous communique. Car c'est le type d'informations qui peuvent être intéressantes d'obtenir lorsque l'on rédige un audit, pour un client par exemple!

Un audit comparatif sur les sites web des candidats à la maison Blanche - Octobre 2007

http://www.websitegrader.com/Internet Marketing Report on Presidential Campaign Websites/tabid/53122/Default.aspx

Un comparatif du Top 20 des blogs, selon Technorati - Octobre 2007

http://www.websitegrader.com/SEO Reports/tabid/49252/Default.aspx

Vous pourrez trouver d'autres informations intéressantes en explorant Google grâce à la requête "site:websitegrader.com".

#### **▼ 2.1.3.3 Press Release Grader**



La société HubSpot, à l'origine de l'outil Website Grader, diffuse également Press Release Grader, un outil permettant d'auditer le contenu d'un communiqué de presse.

Note Press Release Grader est disponible à l'adresse suivante: http://pressrelease.grader.com/.

Tout comme son cousin Website Grader, Press Release Grader dresse un reporting, comprenant plusieurs statistiques et détails instructifs, sur la structure d'un contenu:

- Statistiques générales
  - □ Nombre de mots composant le communiqué,
  - Nombre de phrases,
  - Nombre de liens,
  - ☐ Le niveau intellectuel sensé avoir pour pouvor lire ce communiqué.

Et va proposer une liste d'améliorations sur chacun de ces points:

- Réaliser un contenu entre 250 et 750 mots,
- Spécifier les coordonnées de la société téléphone, email, etc.),
- Définir un contenu qui parle bien de la société (l'outil étant seulement disponible en anglais, il recherche un contenu de type "About the company etc."),
- Saisir l'URL de la société,
- Une analyse basique des liens,

Et fabriquer un nuage de mots-clés permettant de visualiser la densité des termes employés, très utile pour améliorer son contenu. En somme, cet outil dédié aux communiqués de presse comporte quelques éléments intéressants, même pour tester la saisie d'un billet de blog!

Si seule la génération du tag-cloud vous intéresse, de nombreux autres outils existent sur internet, en voici un parmi tant d'autres: http://www.referencement-page1.fr/tag-cloud/tag,cloud.php

#### **▼ 2.1.3.4 Compléter l'audit avec le Spider-Simulator**



Note Spider Simulator est disponible à l'adresse http://www.spider-simulator.com/fr

Le Spider-Simulator est un outil lancé par le site <u>Abondance</u> qui permet de voir le site **avec l'oeil d'un bot**. Et donc de se rendre compte de la construction de votre site, ou celui de vos concurrents, en vérifiant précisement:

- le contenu textuel de la page vu par les moteurs,
- les liens internes et externes suivis par les robots,
- des informations sur les images: nombre d'images et textes alternatifs utilisés,
- etc.

#### Contenu textue

 Accuell 9| Magazine 9| Recherche 9| Forum 9| 9 Chat 9| 9 Connecte; vous 9| 9 Inscription 9| 9 Témoignages Les dédicaces De CHÔLIVAL pour TOUT LE MONDE 🐠 SALAM, c'est le 27 du Ramadan, Leila L9adr Mobarak à tout les Marocaines (s) du monde et avec plus de lecture d CORAN. SiMed Agadir De NOUR995 pour CASABLANCA \*\* Ramadan mabrouk... je cherche bent nass de casa pour fonder une famille serieuse... De ADILO03 pour TOUT LE MONDE 66 Bonjour II ramadan karim II J'aimeraix bien faire votre connaissance I venez voir mon profil et laisser moi vos messages! Bisous.. De BASTIO75 pour TOUT LE MONDE 🍑 bonjour, je cherche des connaissances à casablanca. travaillant à l'etranger je passe ue petite semaine par mois à casablanca. De LADY 19 pour TOUT LE MONDE 🍑 Salut tout le monde ! "l'imagination est plus imporatrite que le savoir" ramadan kareem De BRUNETTE pour CASABLANCA 🚧 Icherche un mec s'appellant achraf, amine ou soufiane, habitant ds les environs de maarif ou ziraoui, almant la house les films.... De DAERON pour TOUT LE MONDE 😻 Salut la terreil Un bon ramadan a tous et surtout un grand salut a mes fréres et seeurs marocains :)! De ZINE\_111 pour TOUT LE MONDE 🐿 bonsoir tout le monde ramadan karim De XELAGOO pour MARRAKECH 😻 gros bisous a nivine de l'hotel tikida garden tro tro belle si quelqun la connais dite lui mera De KREZGUINI pour AGADIR 🐠 bisoir tous le monde mebrok remdan o 3vachrkom mebrouka De AMIRA78 pour TOUT LE MONDE 👐 je souhaite un bon ramadan a tout les marocains dans le monde entier mabrouk 3vasher v ramadan karim. De HINDOO pour TOUT LE MONDE 💖 sitt c'hind de casa je te dit bon ramadan pritt les marocainns e marocainnes De AMAZROY pour AL-HOCEIMA 👐 salut tout le monde ramadan moubarak sa'ild De PIZZA30 pour CASABLANCA 👽 salam je veux dire tout d abord mabrouk ramadan pour toutes et tous et ma familles sans oublier arnel de casa mero a toi merci Ecrire une dédicace \* Demiers connectés redevane kamalest spyder chanab07 samo\_55 colibri adil13 machar limame irem coyo69 tachfine Ils sont sur le chat (23 membres) sofia1230 meri nehal pure\_poison hind001a souka acceslimite yassino01 tachfine nouhammou24 highteck minoch aza2im simohammed badboy mous1 amine188 maylist chase dersimlyim nabil11 riade pizza20 Demiers inspits apyder adil13 hatim007 nasrallah1 khalidgt acceslimite vaffffae meltor daho6625 jonny235 ramo81 kalagane Top membres

rasha

abla

imane booba

manua mimin07

ritabrune

nadat

cocolina

sisita

nineu

mamo missiolo

koutika

Jajar666 Anniversaires saidsport10 abdallah82 tipo11 tipo01 kamel100 almohavir karim0101 cyrilpouvade ennati sanae med29 prince 123 imanet euro men toutou1234 boy vic essaf hammouda9999 querty kascadeur aminos de nadorman brain99 alisiamo ala elkhattabi22 sslimou massou770 bambie abbadmed 80 youeminem01 scope, amour sblues bilove nene23 ayourino fouadnet xmen3 sachiddddd ricos epicure sedd ddarkus kaka10100 Bienvenue sur vaniss le nouveau portail de rencontre totalement gratuit pour tout ceux qui désirent trouver l'amour, l'âme soeur, créer de nouvelles amitiés, ou simplement participer au chat et forums de discussions! Hommes et femmes, dans toutes les villes du Maroc (Cazablanca, Marrakech, Rabat, Agadir, Fès, Tanger) trouveront la perzonne au profil tant rêvé et qui ni lattend que l'illoccasion de nouer de nombreux contacts. Ni lattendez plus inscrivez-vous et faitas la rencontre qui changera votre vie. Accueil | Qui sommes nous ? | Partenaires | Publicité | Plan du site | Contact Agadir - Casablanca - Dakhla - Fès -Essacuira - Ifranc - Kénitra - Leayounc - Marrakech - Meknès - Mohammédia - Quarzazate - Rabat - Safi - Salé - Settat - Tanger - Taza - Tétouan -MaghrebSpace - Fleurs Maroc - Musique gratuite - Rencontre gratuite Maroc et Afrique - Prénoms - Prénom - Tarot gratuit - chat gratuit - rencontre au sénégal - Site de rencontre - Tchat - stream series - stream series - streaming series - Music Maroc - Tchat - ZiCLip.com - couple - Sites de rencontres -QuébecSpot Tous les droits réservés • www.waniss.com 2007

Le contenu textuel est primordial, pour se rendre compte de l'ordre de lecture du contenu de sa page. Plus les mots seront positionnés en début de liste, plus le site sera optimisé sur ces termes.

#### **▼ 2.1.4 Les Web Analytics**

L'objectif du réferencement est de développer en l'audience de son site web, et traduire cela par l'augmentation de ses visites journalières. Il est donc indispensable de disposer d'indicateurs permanents permettant de:

- mesurer l'évolution du trafic de son site web (on parle de "mesure d'audience" ou d'"audimétrie",
- qualifier son audience afin de fournir un contenu toujours plus proche de l'attente de son public

On distingue donc logiquement deux types d'études:

- les études de mesure de la fréquentation "site centric". axée sur le site
- les études de qualification du trafic "user centric", axée sur l'utilisateur

Les Web Analytics sont des outils, gratuits ou payants, proposés sous 4 formes distinctes:

### Les systèmes de type "Logs"

Il s'agit dans ce cas de l'exploitation des fichiers de logs du serveur web par un outil spécifique. Chaque activité du serveur est enregistrée, puis envoyée, par planification automatisée, à l'outil de statistiques. Il s'agit donc de choisir un outil capable d'analyser ces journaux d'activité et de réaliser un tableau de bord contenant les principaux indicateurs de fréquentation du site web.

ex: WebTrends http://www.webtrends.com, Urchin, par Google http://www.urchin.com

### Les "sniffer" ou "analyseur de trames"

Un "analyseur réseau" est un dispositif permettant d'"écouter " le trafic d'un réseau, c'est-à-dire de capturer les informations qui y circulent: les paquets réseau ou "packet sniffing". Il sert plus généralement aux administrateurs. A l'instar d'une solution de type logs, on y perd les notions d'évènements et d'interactivité.

ex: Ethereal http://www.ethereal.com

# Les systèmes "Ad hoc"

Il s'agit d'outils stockés sur le site, et capables de stocker en temps réel des données. Pour les sites web possédant un trafic important, ce type de dispositif peut provoquer une charge du processeur importante et un fort taux d'utilisation de l'espace disque, notamment si les données collectées sont stockées dans un système de gestion de base de données.

ex: PhpMyAdmin http://www.phpmyadmin.net

#### Les services de" mesure d'audience"

Ce système consiste à insérer dans chaque page un "marqueur " ou "tag" en javascript permettant au service de mesure du trafic de collecter les données sur un serveur. Ce type de service a l'avantage d'économiser des ressources matérielles, car l'ensemble des traitements est réalisé sur le serveur distant. De plus, la société proposant le service se charge de faire évoluer les indicateurs et le panel des tableaux de bord proposés afin d'être constamment en adéquation avec l'évolution des technologies d'accès à internet et des navigateurs web. En contrepartie les statistiques ainsi recueillies ne seront pas nécessairement exhaustives dans la mesure où:

- certains utilisateurs arrêtent le chargement de la page avant que le code du marqueur ne soit téléchargé,
- des serveurs proxy intermédiaires sont susceptibles de gêner le chargement du tag,
- les infrastructures de sécurité, notamment les dispositifs pare-feu, peuvent bloquer la remontée d'informations,
- aucune information n'est fournie concernant les visites des robots.
- aucune information n'est fournie sur les accès aux autres documents (images, vidéos, fichiers divers).

ex: Google Analytics http://www.google.com/analytics, Yahoo IndexTools http://www.indextools.com/, Xiti http://xiti.com/, Weborama http://weborama.com etc.

Certains de ces systèmes ont une orientation "user centric" plus "commerciale" et KPI "Key Performance Indicator", en analysant les paramètres marketin clés de performance (ROI "Return On Investissement", CPL "Cost per Lead", etc.). Parmi ces outils, les plus reconnus sont Omniture <a href="http://www.omniture.com">http://www.omniture.com</a> et Médiamétrie E-stats <a href="http://www.estat.com">http://www.estat.com</a>

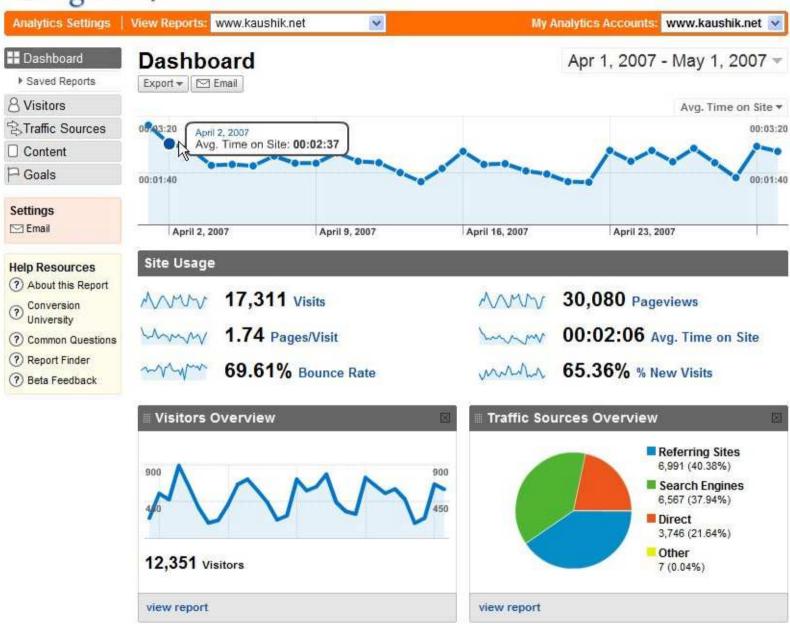
# Les systèmes "Hybrides"

Certains outils de web Analytics utilisent une méthode hybride qui consiste à insérer par Javascript (via un serveur séparé) les informations correspondants au tags dans les logs. De nombreux administrateurs de système utilisent un outil d'analyse des logs conjointement avec un système de marqueurs.

Pour pouvoir comparer les différentes solutions du marché, rendez-vous ici: http://www.webanalyticsworld.net/2007/03/ultimate-web-analytics-comparison.html

# **2.1.4.1 Google Analytics**



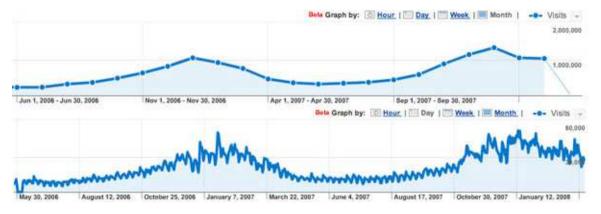


Note Google Analytics est disponible à l'adresse suivante: <a href="http://www.google.com/analytics">http://www.google.com/analytics</a>. Pour l'activer sur votre site, vous aurez juste un petit script JS à déposer en bas de chaque page de votre site. Il vaut mieux le déposer en bas de page, sinon l'appel à un JS distant risque de ralentir l'affichage de votre site, et d'impacter négativement sur votre référencement naturel.

Créé en 2005, Google Analytics est un logiciel d'analyse de statistiques de trafic très complet multi-sites et multi-domaines. D'innombrables fonctionnalités jalonnent cet outil. Nous n'allons pas toutes les énumérer, mais bien nous concentrer sur l'intérêt d'une telle plateforme pour auditer un site web. Car il est important d'analyser le **comportement des visiteurs** sur son site, pour **développer un trafic de qualité**:

- quelles sont les pages préférées ?,
- quand viennent les visiteurs ?,

**Note** Google Analytics s'est vu récemment affiné, d'une fonctionnalité permettant d'analyser toutes les statistiques, par **échelle temporelle** (heure, jour, mois, etc.). Essentiellement intéressant pour tester des pics d'audience à des moments-clés, comme les emailing, ou les opérations type advergames éenementielles.



- quel est le niveau de fidélité des visiteurs? Les visiteurs reviennent-ils sur le site? A quelle fréquence? Quand sont-ils venus pour la dernière fois?,
- les liens externes qui rapportent le plus de visiteurs.

Note Vous pouvez tester cela facilement pour n'importe quelle URL. Rendez-vous sur Google, testez la requête suivante: "link:wikipedia.com". Seules les pages pointant vers le nom de domaine "wikipedia.com" s'afficheront dans le SERP. Avec à droite de l'écran, le nombre de liens indexés dans Google.

Résultats 1 - 10 sur un total d'environ 16 400 liés à wikipedia.com. (0,34 secondes)

- les mots-clés les plus fréquents qui créent du trafic sur votre site,
- quel est le pays de provenance de mes visiteurs ? Ai-je besoin de traduire mon site dans une autre langue ?
- etc.

Une analyse approfondie de ces statistiques peut être très utile pour améliorer son travail de référencement, et connaître le profil type de ses visiteurs. Il existe de nombreuses solutions similaires: Xiti, Direct-stats, PhpMyVisites, c'est une question de goûts et de besoin.

Note Malgré ce qu'on peut lire ici et là, au sein de la communauté web, utiliser Google Analytics n'offre aucun point de bonus dans le cadre de son référencement. Que l'outil utilisé ne soit pas de la lignée Google, n'impactera rien en ce sens. Seule l'utilisation du contenu des informations fournies vous aidera à avancer.

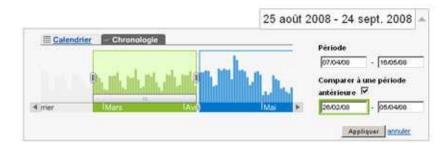
On peut toutefois imaginer que Google Analytics aidera Google à optimiser encore davantage son algorithme d'indexation. Et qu'il déterminera d'autant mieux les **fermes de liens**, et les opérations frauduleuses de référencement.

Néanmoins, Google Analytics apporte des statistiques très avancées, pour certaines de type datamining:

• le taux de rebond ou bounce rate, des internautes qui sont venus avec tel mot-clé, et ont quitté le site immédiatement ; on comprend ainsi qu'il n'est pas très utile de continuer à se référencer sur ce(s) terme(s)...! Google Analytics permet à chaque fois de trier les statistiques, par provenance géographiques des internautes, et de comparer des périodes temporelles. Retenez que plus le taux de rebond est élevé, mois bonne est votre stratégie de référencement!

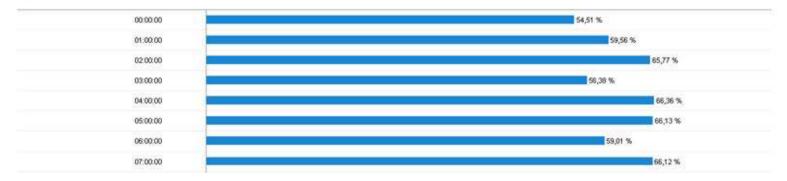
Toutefois, sachez que, **au bout de 30 minutes d'inactivité**, si le visiteur accomplit une nouvelle activité, il sera de nouveau comptabilisé comme étant un nouveau visiteur. Cette statistique n'offre donc pas une fiabilité de 100%, mais en est tout de même très proche.

# Taux de rebond, tous visiteurs confondus





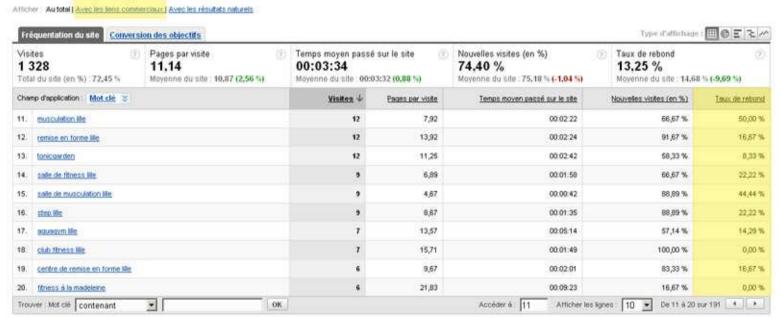
# 59,07 % Taux de rebond



#### la source de trafic:

- □ "Accès directs": les internautes connaissaient l'URL et ne l'ont pas cherché sur un moteur ou un annuaire. Cette statistique va donc déterminer le **niveau de réputation** d'un site web, et déterminer l'impact de sa **comunication globale**, on-line et off-line. Peut-être s'agit-il également de liens stockés dans les favoris ?
- □ "Sites référents": les internautes ont trouvé un lien sur internet, qu'il s'agisse d'un lien sponsorisé, d'une bannière de publicité, d'un lien sur un site partenaire, sur un annuaire, ou encore d'un lien dans un emailing commercial; cette statistique détermine l'impact de la stratégie global on-line, et la stratégie de net-linking. C'est là toute la force de Google Analytics, qui arrive encore à disséquer l'information, en croisant cette statistique sur de nombreux critères: le temps moyen passé sur le site depuis chaque backlink, le taux de rebond depuis chaque backlink (ce qui permet de s'apercevoir que mettre un lien chez ce référent n'est pas avantageux, et peut être néfaste pour son ROI, si ce backlink, ou cette bannière de publicité a un coût), etc.
- □ "Moteur de recherche": Google Analytics nous liste tous les moteurs de recherche utilisés pour trouver le site web, et nous donne la possibilité de cliquer sur ces référents pour connaître les **mots-clés utilisés** pour trouver le site ; à nouveau ces statistiques sont croisées avec le bounce rate.





Google Analytics permet également de différencier les origines des sources, s'agit-il d'un organic result, ou d'un sponsored link? Ce qui permet notamment de déterminer l'impact de sa campagne AdWords de liens sponsorisés par exemple. Une campagne AdWords bien réalisée, doit ramener plus de trafic que du référencement naturel, sinon l'investissement n'est pas intéressant! C'est là qu'une fonctionnalité de comparaison entre une période donnée et une seconde, permet de déterminer davantage si l'évolution dans sa stratégie SEM a été correctement déterminée. Nous reparlerons en détail des fonctionnalités statistiques du SEM dans la partie dédiée à ce sujet.

En parallèle, pour ce qui est du réferencement naturel, cette fonction de comparaison permet de vérifier si l'évolution du contenu de son site web, et sa stratégie générale de net-linking s'avère bonne et génère de plus en plus de trafic.

# • l'appréciation du contenu, en décortiquant:

- □ les pages les plus consultées,
- □ les landing pages, ou "pages d'atterissage",
- □ les pages par lesquelles les internautes quittent le site,

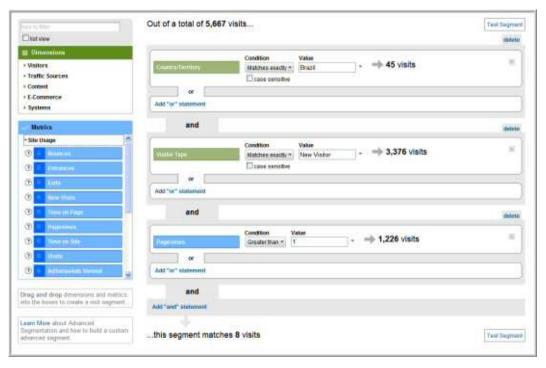
- □ le parcours complet depuis l'entrée jusqu'à la sortie du site,
- □ les pages qui sont à l'origine de visites d'1 seule page (="le rebond"):
  - si les pages détectées contiennent beaucoup de liens externes, il est probable que cela fasse considérablement chuter le nombre de pages visitées!
  - ♦ dans le cas d'un blog, où la plupart des billets seront postés sur la page d'accueil, le taux de rebond risque d'être très fort,
  - il se dit qu'une stratégie correcte pour un site corporate de PME requiert que ce taux de visite tourne aux alentours de 20%

Ces statistiques sont à nouveau croisées avec les taux de rebond, le temps passé, etc.

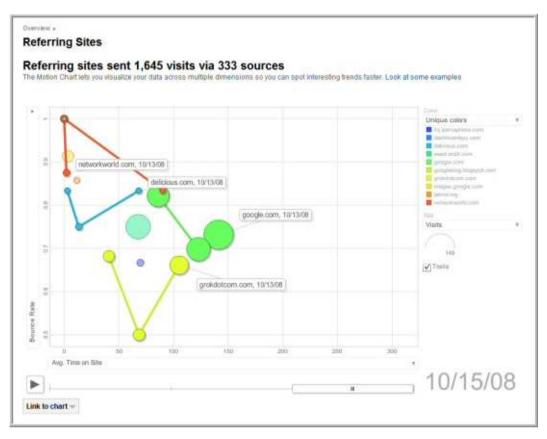
#### La version v3

Sortie en Octobre 2008, outre une évolution de son interface générale, et une amélioration de l'ergonomie, Google propose de nombreuses évolutions statistiques très intéressantes:

• Une meilleure segmentation des visiteurs: vous pourrez désormais créer des segments de visiteurs en vous appuyant sur plusieurs paramètres. Vous ne serez donc plus limités à 2 dimensions comme avec la version actuelle. Avec cette évolution, vous pourrez filtrer vos rapports préférés par type de public. Cette option correspond aux ASI segments & DataWareHouse de SiteCatalyst chez Omniture.



• Des graphiques animés ou motion charts, vous permettent de voir l'évolution de vos données dans le temps, et pas juste de façon statique. Vous pourrez ainsi faire de nouvelles découvertes sur vos visiteurs, et corréler les données avecdes outils de type KPI:



Un exemple dynamique ici: http://spreadsheets.google.com/pub?key=pCQbetd-CptE1ZQeQk8LoNw

# **Exclure son propre trafic de Google Analytics**

Pour ne pas fausser les statistiques générées par Google, il vaut mieux filtrer les entrées sur le site.

Par défaut, on pensera simplement à exclure son adresse IP, mais ce n'est pas une solution 100% efficace, si:

- vous utilisez un firewall,
- vous n'avez pas une IP fixe.

Autre solution, Google Analytics met à votre disposition une variable que vous pouvez utiliser comme bon vous semble. Cette variable user-defined est définie en appelant la fonction \_setVar('valeur'); dans votre page HTML, au sein de votre code Google Analytics. Par exemple:

```
<script type="text/javascript">
var gaJsHost = (("https:" == document.location.protocol) ? "https://ssl." : "http://www.");
```

```
document.write(unescape("%3Cscript src='" + gaJsHost + "google-analytics.com/ga.js' type='text/javascript'%3E%3C/script%3E"));
</script>
</script type="text/javascript">
    var pageTracker = _gat._getTracker("UA-123456789-1");
    pageTracker._setVar('OX2:Julien');
    /* autres paramètres */
    pageTracker._trackPageview();
</script>
```

Google va alors lancer un mécanisme de création de cookie utmv qui contiendra le nom "Julien". Faites attention à n'écrire ce code que sur une page externe du site, que vous seul connaissez!

Note "OX2" est un règle définie par Google, qui vous permet d'exclure un ou des visiteurs, par exemple:

```
OX2:(Julien|David|Marie)
```

pour exclure 3 personnes,

```
OX2:.*
```

pour exclure tout le monde!

Puis, il reste à exclure le trafic dans l'interface de l'outil: "Gestionnaire de filtres > Ajouter un filtre":



# **▽** 2.1.4.2 Yahoo! Web Analytics

Actuellement en BETA fermée à l'heure d'écriture de ce cours, Yahoo! se lance sur les traces de Google: http://web.analytics.yahoo.com/

A la différence de Google, Yahoo! devrait proposer des statistiques en temps réel, Google ayant des statistiques décalées entre l'instant "t" et les stats affichées.



# **2.1.4.3 Woopra**

Woopra est une double application web & desktop, de statistiques de trafic, qui fonctionne également par tag javascript: http://www.woopra.com/



La version desktop est très orientée "interface riche", et risque de consommer bcp de ressources systèmes.



Et une option est disponible pour interagir avec les visiteurs en temps réel (via chat)!

Cet outil est effectivement pensé en priorité pour les **blogs**, et diffuse d'ailleurs un plugin pour WordPress: Woopra WordPress Plugin from the WordPress Plugin Directory



Mais il faudra encore attendre un peu ... l'outil est encore en BETA fermée à l'heure de l'écriture de ce cours!

# Whos.amung.us

Encore un outil dans la lignée de Woopra, qui permet également de connaître le trafic en temps réel, via une petite carte: http://amung.us/

Et, pour les blogs, d'ajouter un widget qui affiche directement sur le site l'évolution du trafic.



# Clicky

Encore un outil très utilisé dans la blogosphère, qui propose des statistiques en temps réel, une compatibilité avec les iPhone, de trier les visiteurs par sociétés (il est capable de dire si les visiteurs sont des gens de chez Microsoft par exemple), de connapître l'impact de Tag Cloud, etc. la liste est impressionnante et disponible ici: <a href="http://getclicky.com/">http://getclicky.com/</a>



# **▽** 2.1.4.4 YouTube Insight



YouTube est un site web d'hébergement de vidéos sur lequel les utilisateurs peuvent envoyer, visualiser et se partager des séquences vidéos. Racheté en 2006 par Youtube, et prochainement bénéficiant d'un **nouveau concept publicitaire**, en incluant des liens sponsorisés directement dans le chargement des séquences, Google permet désormais d'étoffer le contenu de son site, sans pour autant le ralentir.

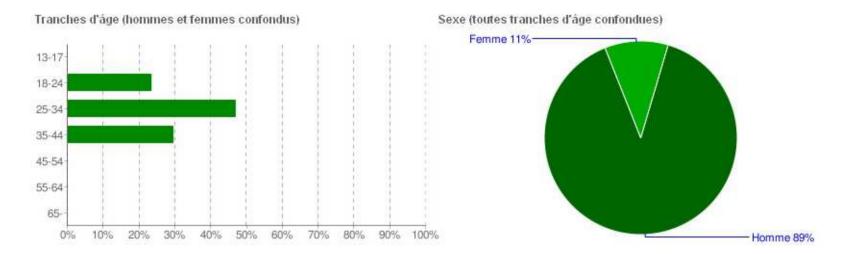
Note Vous pouvez accéder à YouTub Insight ici: "My account > Videos, Favorites, Playlists > Manage my Videos > About this Video".

Déposer une vidéo sur Youtube, ou d'autres espaces comme Daylimotion, Kewego, Wat, Widéo, etc. permet donc très facilement de créer un lien entrant sur son site.

YouTube s'est vu récemment compléter d'une zone de statistiques, **YouTube Insight**, qui permet désormais de connaître les tendances des internautes qui regardent les vidéos: le nombre de fois où la vidéo a été regardée, par date et provenance géographique des internautes, etc. Même s'il semble évident que l'adaptation des contenus à un public donné n'est bien sûr pas une obligation pour les diffuseurs amateurs de YouTube, il dénote une sorte de professionnalisation de l'activité.

- Qui regarde les vidéos ?
  - □ YouTube arrive à déterminer l'âge et la civilité des internautes. Certainement s'agit-il des internautes déjà connectés sur YouTube. Ce qui dénature un peu la statistique. Néanmoins, cela permettra à Google d'orienter les stratégies des liens sponsorisés, lorsque la publicité fera apparition dans les vidéos, et d'orienter ainsi les possibilités des investisseurs en SEM.

Et c'est un premier pas, pour s'assurer que la cible n'est pas faussée.



- La popularité des vidéos
  - □ YouTube Insight permet également de connaître, selon une chaine du temps, la cote de popularité comprise entre 0 et 100, et le nombre total de consultations des vidéos par zone géographique (pays ou continent).



Votre chaîne est très populaire dans ces pays/États.

France	5,24	
Réunion	0,37	
Luxembourg	0,16	
Belgique Algérie Maroc Bulgarie	0,1	
<u>Algérie</u>	0,1	
Maroc	0,05	
<u>Bulgarie</u>	0,05	

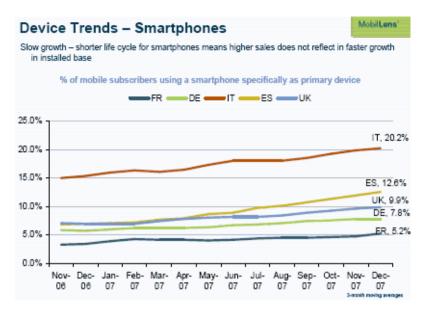
Cette popularité ne s'arrête pas au visionnage, mas on peut également connaître succintement le **nombre de notations** et la **note moyenne** de chaque vidéo. Et s'assurer de son efficacité et l'interaction de son contenu avec le reste du site où elle est diffusée.

Attendons-nous à ce que ce système de notation soit amplifié par la suite, au vu du récent rachat par Google d'Omnisio. Soyez donc précis dans la construction de vos vidéos!

# **2.1.4.5** L'avenir des Web Analytics

Le marché de la mesure d'audience se développera certainement dans le monde du **mobile** ; les besoins en terme de mesure d'audience sont encore limités dans ce domaine (nombre de visites sur le site mobile, taux de conversion depuis une campagne de SMS, etc.), car les annonceurs sont encore largement dans une problématique de communication générale, pas encore du véritable retour sur investissement.

En cela, la mesure d'audience mobile reste donc dominée par la méthode "user-centric", comme le montre la solution M:metrics http://www.mmetrics.com,



ou encore <a href="http://www.flurry.com/">http://www.flurry.com/</a>, qui demeure gnéraliste mais est dédié aux iPhones:



Il y a encore des places à prendre dans ce domaine!

# **▽ 2.1.5 Wayback Machine**



Note La Wayback Macine est accessible à l'adresse http://web.archive.org

La Wayback Machine a été créée en 1996 par les ingénieurs d'Alexa. Il s'agit d'une sorte de machine à remonter le temps, puisqu'elle indexe à intervalle régulier les pages de **tous les sites internet**depuis leur création. Ce service est appelé "index à trois dimensions". Les "clichés" sont disponibles de 6 à 12 mois après leurs captures. La fréquence des instantanés est variable, toutes les mises
à jour de sites Web ne sont pas enregistrées, et des intervalles de plusieurs semaines peuvent être remarqués.



En 2006, la Wayback Machine contenait près de **2 pétaoctets de données** (rendez-vous ici pour les unités de mesure: <a href="http://fr.wikipedia.org/wiki/Octet">http://fr.wikipedia.org/wiki/Octet</a>). Le volume augmente à un rythme de **20 téraoctets par mois**. Un plugin Firefox existe: **Wayback** https://addons.mozilla.org/fr/firefox/addon/318?application=firefox&id=318&vid=1075 pour accéder plus facilement aux archives.

Il semblerait que le moteur de Google base son algorithme d'indexation en fonction de plusieurs paramètres, dont **l'âge du nom de domaine**. Il se dit qu'un âge supérieur à 3ans est un bonus pour le Googlebot. Plus vous avez de données stockées dans la Wayback Machine, meilleure sera donc votre indexation. Vous pouvez vérifier l'âge du nom de domaine grâce à des outils de type WHOIS, comme: http://www.raynette.fr/services/whois/.

Vous pouvez également utiliser les archives du Wayback Machine, pour les passer sur le Spider Simulator, et ainsi vérifier l'évolution dans l'optmisation d'un site de vos concurrents, par exemple. Ou vous inspirez de leurs anciens contenus pour écrire les vôtres !

#### **2.1.6 Plug-ins Firefox**

De nombreux browsers existent sur le marché: Internet Explorer, Google Chrome, Safari, Firefox, etc. Ce dernier, open-source, a développé une large communauté dédiées à la création de plugins. Dont des plugins dédiés au SEO:

#### **SEO Links**

téléchargeable ici: http://www.webmasterbrain.com/seo-tools/firefox-extensions/seo-links/



Grâce à ce plugin, il suffit de survoler un lien sur une page web pour obtenir des informations sur la page concernée. En passant le curseur au-dessus d'un lien pointant vers une page "PAGE", ayant comme texte de lien "TEXTE", vous obtiendez les informations suivantes:

- nombre de backlinks de "PAGE" sur Google, Live Search et Yahoo
- positionnement de la page "PAGE" pour la requête "TEXTE" sur Google, Live Search et Yahoo

Cette extension peut vous permettre d'analyser le référencement des sites que vous visitez, et vous aider à trouver des sites avec lesquels échanger des liens. Attention toutefois, si vous laissez cette extension activée en permanence, l'outil va envoyer de nombreuses requêtes aux serveurs de Google et en conséquence vous risquez d'être bloqué: Google affichera un message d'erreur quand vous irez sur son moteur. Ne vous inquiétez pas, il suffit d'attendre quelques heures pour être débloqué.

#### **SEO for Firefox**

téléchargeable ici: http://www.websiteburo.com/goldorank/



Permet d'obtenir de nombreuses informations précises: le PageRank, si le site est référencé dans l'annuaire de Yahoo.com, le nombre de liens issus de .edu pointant vers l'ensemble du domaine (utilise la commande **linkdomain** sur Yahoo), etc.

### **GCache**

téléchargeable ici: https://addons.mozilla.org/fr/firefox/addon/301?application=firefox&version=1.0&os=Windows&numpg=10&id=301

Permet simplement d'afficher la version en cache de la page chez Google.

# **User Agent Switcher**

téléchargeable ici: http://extensions.geckozone.org/UserAgentSwitcher/

Ce plug-in permet de modifier la chaine d'identification du navigateur, appelée **User Agent**. Pour surfer en vous faisant passer pour un robot, il vous suffit de configurer les bons user agent. Et tester ainsi les sites qui pratiquent la technique frauduleuse de **cloaking**.

# Web Developper

téléchargeable ici: http://joliclic.free.fr/mozilla/webdeveloper/



WebDevelopper se présente sous la forme d'une barre d'outils, fournissant un nombre incroyable de fonctionnalités utiles pour tester l'affichage d'une page web. Le must pour les référenceurs, avec entre autres: désactiver les styles CSS, désactiver les cookies, désactiver le JavaScript, afficher les textes alternatifs ALT à la place des images, lister les liens présents sur une page, etc.

# GoldoRank



Goldorank est une extension permettant de connaître de façon rapide le positionnement d'une page ou d'une URL sur les principaux moteurs et annuaires.

# **Link Checker**

téléchargeable ici: http://extensions.geckozone.org/LinkChecker

Vérifie la validité de tous les liens d'une page, et vous liste les liens morts.

# Google PageRank Status

téléchargeable ici: http://pizzaseo.com/firefox-pagerank

Ce plugin affiche le PageRank de chaque page visitée. Exactement comme sur la GoogleBar, mais en plus léger.

# **Live HTTP Header**

Ce plug-in affiche de façon très détaillée les entêtes HTTP transmises pendant votre surf, et permet par exemple de vérifier si une redirection est du type 301permanente ou 302 temporaire.

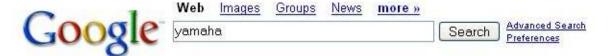
#### **SEO Quake**

téléchargeable ici: http://www.seoguake.com/



SEO Quake permet d'obtenir des informations intéressantes dans l'analyse des positionnements de la conccurence.

Un peu moins avancé, Google Preview reprend cette idée d'optimisation du SERP, et ajoute un preview du site: <a href="http://extensions.geckozone.org/GooglePreview">http://extensions.geckozone.org/GooglePreview</a>



# Web

# Welcome To The World Of Yamaha!

Yamaha guitars and accessories. www.yamaha.com/ - 9k - Cached - Similar pages SeoQuake: GPR; 7 GI; 35,900 € Rank; 10717

# America's Most Diverse - Yamaha Motor Corporation, USA, Home

Motorcycles, ATVs, Outboard Motors, Snowmobiles, Sport Boats, WaveRunners, Race Kart Engines & Generators distributed by **Yamaha** Motor Corporation USA.

www.yamaha-motor.com/ - 14k - Cached - Similar pages

Seo Quake: GPR: 7 GI: 90,900 @ Rank: 11320

# Yamaha UK Home Page

Yamaha Motor UK · Yamaha Electronic UK · XG · Digital Music World · CD Recording ... www.yamaha.co.uk/ - 10k - Cached - Similar pages
SeoQuake: GPR: 6 GI: 7,750 Rank: 377173

# Yamaha Canada Music

Yamaha Music Canada. About Yamaha | Dealer Locator | CD Rewriters | Yamaha Motors | Worldwide Links ... 2005 Yamaha Canada Music Ltd/Ltée. All Rights Reserved.

www.yamaha.ca/ - 16k - Cached - Similar pages SeoQuake: GPR: 6 GI: 4,550 @ Rank: 283814

Seoquake in SERP Image 4 of 5

# **KGen**

téléchargeable ici: http://kgen.elitwork.com/

8	sQ	<b>4</b> ×
8 PR	: 4	
8 I:	90	
32 L:	5 <u>,755</u>	
W LD	: 20,584	8
W LD	2: 20,58	4
# I:	<u>0</u>	
<b>2</b> Ra	nk: <u>5110</u>	0
III Ag	e: <u>Apr 10</u>	, 2007
- I:	752	
🔯 wh	ois .	
sourc	<u>e</u>	
Robo:	yes	
Sitem	ap: <u>no</u>	
G tre	nds: <u>3.0+</u> 1	<u>K</u>
lh De	<u>nsity</u>	
(b) Lin	ks: 17 2	9

Vertical mode of Seotoolbar Image 5 of 5



KGen est un extracteur sémantique de mots-clés qui permet d'analyser les mots contenus dans une page et de déterminer ceux qui, pour les moteurs de recherche, auront le plus d'importance. et il permet d'optimiser votre référencement et de déterminer la stratégie de vos concurrents.

# **▼ 2.2 Le PPC "Pay per click"**

Le sigle SEM "Search Engine Marketing" regroupe les activités de référencement, de positionnement publicitaire, de développement de la popularité et plus généralement toutes les formes de marketing liées à la recherche sur Internet. Le SEM comprend:

- le SEO "Search Engine Optimization", qui regroupe toutes les techniques pour optimiser le réferencement dit "naturel" de son site web. Et c'est essentiellement l'objectif de ce cours.
- le **PPC** "Pay Per Click" (CPC "Coût par clic" en français), qui est de la publicité payante, figurant dans les résultats des moteurs de recherche et sur certains sites web. Ce sont des "**liens** sponsorisés":



Les moteurs ont chacun leurs propres programmes de type PPC: Google propose les Google AdWords, Yahoo propose le réseau Yahoo Search Marketing, anciennement nommé réseau Overture,

MSN/Live Search les MSN Keywords, Ask Jeeves les Ask Jeeves Sponsored Listings, etc.

La différence majeure entre le PPC et le SEO, est la suivante:

le SEO se voue	à être une opération régulière et durable:
□ durable	
□ lent	
□ coûts de de	épart élevés
□ coûts à lon	g terme faibles
□ image de n	narque forte
annonceurs ne p haut du classem	ord un acte d'achat. Chaque clic sur les liens sponsorisés sont ensuite énumérés précisément dans un outil statistiques de suivi. Il y a maîtrise de l'affichage et du budget, les paient que lorsque l'on clique sur leurs liens commerciaux. Le PPC a une dimension plus "évenementielle" (fêtes, soldes, etc.), et assure un positionnement de l'annonce dans la ent. Ce positionnement est déterminée à la fois par son taux de clic (CPC "Cost Per Click", ou "Coût par clic", ou encore CTR "Click-Through Ratio") et la somme que expte de payer à chaque clic ("bid", ou enchère).
□ éphémère	
□ rapide	
□ coûts de de	épart faible
□ coûts à lon	g terme élevés
□ image de n	narque faible
	Note Les snippets des liens sponsorisés sont conçus pour être visuellement attractifs. Utilisez-y des arguments commerciaux forts: mots clés, superlatifs, comparatifs (il
	faut demander l'accord auprès de Google dans ce cas), affichez-y des prix, etc. Pensez aussi à effectuer des tests d'affichage: mots clés dans l'URL ou dans le titre ?
	Et surtout, pensez à ce que votre landing page soit en accord avec le snippet: ne menez pas vers la page d'accueil si l'annonce affiche un zoom produit! Préparez
	même des pages dédiées s'il le faut, pleines de "call to action" afin de transformer le plus possible !

• le SEO implique de réaliser des optimisations sur, ou autour de votre site, pour en augmenter sa pertinence ou sa notoriété. L'évolution du SEO se calcule à l'aide Web Analytics par exemple. Et

#### **▼ 2.2.1 Les variations de mots-clés en campagne PPC**

Quand vous planifiez une campagne de référencement PPC ou que vous cherchez des mots-clés pour cibler un contenu, il est fréquent d'oublier un grand potentiel de mots-clés pourtant simples à utiliser. Surtout que ces éléments alternatifs sont faciles à dénicher et, étant généralement peu compétitifs, le budget requis sera plus bas que prévu.

Nous avons déjà vu comment déterminer un choix précis grâce à Google Trends pour le référencement naturel. Cette fois-ci, nous allons plutôt explorer des méthodes décisionelles et générer une liste précise dédiée au PPC.

### **▼ 2.2.1.1 Les erreurs de frappe**

C'est la variation la plus connue, et le plus fréquente. Un exemple avec "restaurant, restaurant, rest

A ce sujet, l'exemple le plus flagrant, une anecdote diffusée par Google, relatif à la chanteuse Britney Spears. Une étude menée pendant 3 mois, près de 40.000 personnes ont fait des erreurs sur son

prénom en requêtant "Britanny Spears", et 36.000 ont fait une erreur sur son prénom "Brittney Spears" avec deux "tt".

Pour vous aider dans la décision de ces termes:

- Microsoft AdCenter Labs: http://adlab.msn.com/Keyword-Mutation-Detection/Default.aspx
- Searchspell TYPO: <a href="http://www.searchspell.com/typo/">http://www.searchspell.com/typo/</a>
- Trellian Keyword Spelling Mistake (payant): http://www.keyworddiscovery.com/tour-spell.html
- Générateur de fautes de frappe http://bvwg.actulab.net/fautes-de-frappe.php
- Keyword Typo Generator, orienté claviers QWERTY <a href="http://www.seochat.com/seo-tools/keyword-typo-generator/">http://www.seochat.com/seo-tools/keyword-typo-generator/</a>

# **2.2.1.2** Les acronymes

Les acronymes sont des mots constitués par les premières lettres d'une phrase à mots multiples. SEO est l'acronyme de "Search Engine Optimisation" et PPC est celui du "Pay Per Clic". Quel est donc l'expression la plus requêtée ? "SEO" ou "Search Engine Optimisation" ? C'est donc important d'inclure les 2 dans son contenu, cela permettra aussi d'éviter les pénalités dus à une densité de mots clés élevée.

Autre variations amusantes, les languages "texto": BBQ, Barbeq, pour "Barbecue", par exemple.

#### **▼ 2.2.1.3 Le trait d'union**

L'exemple du "Pay-Per-Click" peut être repris dans ce cas: "Pay per click", "Pay-per-click", ou "Pay per-click"? Et que dire de la variation entre deux mots espacés et ceux constituant un seul mot? est ce qu'on utilise "selfservice" ou "self service"? N'oubliez pas, Google Trends est votre ami!

# **▼ 2.2.1.4 Les synonymes**

Très importants à cerner, ces mots qui ont la même signification ou sont très proches. Exemple: "Soda" ou "Limonade"; attention toutefois à la signification géographique de certains synonymes.

#### ▼ 2.2.1.5 Mots vs. verbes d'actions

"Achat de voitures", "Acheter votre voiture"... Analysez ces diverses possibilités sur Google Keyword Sandbox.

#### **₹ 2.2.1.6 Pluriel vs. singulier**

Le singulier et le pluriel sont à prévoir dans tous les cas, pour toutes vos variations. Analysez ces diverses possibilités sur Google Keyword Sandbox, et priorisez vos choix **en fonction des niveaux de concurrence** déjà établis.

#### **▼ 2.2.2 L'impact d'une campagne de liens sponsorisés sur le SEO**

Beaucoup de rumeurs ont circulé sur Internet, véhiculant l'idée que, le fait d'acheter des liens commerciaux sur Google notamment, allait optimiser le référencement naturel du site concerné. Cela peut s'avérer effectivement positif, mais Google ne prendra pas en considération ce facteur comme étant le plus important. A ces yeux, la qualité d'un lien réside d'abord dans le format du **BL "backlink"**, qui est un lien provenant d'un site extérieur et qui fait une référence à votre site.

Un lien sponsorisé, c'est un lien qui n'est pas "naturel" ; Google ne peut donc considérer ce lien comme étant un lien fiable concernant la qualité de votre contenu. D'autant que ces liens, fournis par un tiers de confiance (Google, Yahoo, etc.) renvoient rarement vers le site, mais bien vers la plateforme qui redirige ensuite vers votre site.

#### **▼ 2.2.3 Le split testing**

Par ce que d'infimes changements peuvent provoquer des effets spectaculaires sur votre le comportement de vos internautes, il est crucial de mesurer l'impact de ces petites modifications, positives ou négatives (changement du titre d'une fiche produit, changement de couleur d'un bouton "ajouter au panier", changer de la police de caractère, de la taille du texte, d'une image, etc.). La saisonnalité, vos sources de traffic ou bien d'autres facteurs temporels peuvent influencer vos résultats. C'est pour cela qu'il est impératif de pouvoir mesurer de façon simultanée l'efficacité ou non d'une modification sur votre site. C'est d'autat plus important sur des sites à objectifs de transformation / conversion: le trafic doit se transformer en vente sur un site d'ecommerce, ou en lead, sur un advergame, par exemple.

Il existe pour cela 2 méthodes de test:

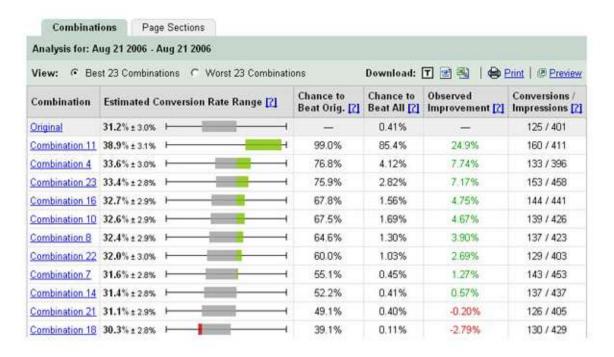
- le Test A/B, ou Split A/B Testing,
- l'analyse multivariée MVT "Multi-Variables Testing"

L'avantage d'un Split Test A/B, c'est que le nombre de visiteurs nécessaires pour avoir un résultat significatif est largement inférieur au nombre nécessaire à une MVT. Si vous avez moins de 2000 visiteurs/jours, étudiez plutôt la méthode Split Test A/B.

Et contrairement à une approche MVT, le Split Test A/B ne mesure pas les interactions entre plusieurs élèments différents (par exemple, est ce que l'expérience 1 à un effet sur l'expérience 2), il peut arriver dans certains cas que 2 résultats de Split Test A/B soient incompatibles entre eux. Attention donc aux conclusions hâtives. Coupler ses données avec un outil d'analyse d'audience comme Google Analytics peut donner une vision plus exhaustive de la pertinence des "landings pages".

# **▼ 2.2.3.1 Google Website Optimizer**

Ces méthodes d'analyse comportementales consistent à afficher deux versions différentes de vos pages. Pour réaliser ces tests, Google a mis en place, depuis 2008, le Google Website Optimizer, disponible à l'adresse: <a href="https://www.google.com/analytics/siteopt/splash?hl=fr">https://www.google.com/analytics/siteopt/splash?hl=fr</a>. Pour l'instant, l'outil est gratuit, il sera amené à ne plus l'être bientôt...! D'autres outils existent, mais souvent très onéreux, c'est le cas d'Amadesa par exemple: <a href="https://www.amadesa.com/">https://www.amadesa.com/</a>



Google Website Optimizer vous permet de tester différents scénarios de pages pour un même site et d'en tester l'impact sur le taux de réussite d'un objectif, typiquement, le taux de conversion d'un site e-commerce.

Par exemple, vous souhaiteriez mettre en ligne une nouvelle homepage de votre site, mais vous avez peur de perdre des clients? Créez alors votre nouvelle homepage, indiquez son URL à GWO et commencez les tests. GWO affichera à tour de rôle votre ancienne et votre nouvelle page (méthode A/B) et, **jour après jour**, vous indiquera celle qui donne le **meilleur taux de conversion**. Et si résultat ne vous satisfait pas, vous pouvez recommencer les tests autant de fois que vous voulez avec autant de modèles que vous voulez.

Google Website Optimizer permet également de **tester des sections de page** (méthode MVT), c'est à dire d'effectuer des tests de mises en page, de wording (="rédactionnel"), ou de photo à l'intérieur même des pages et de tester leur impact.

# **□** 2.2.4 Google AdWords

Google AdWords <a href="http://adwords.google.fr">http://adwords.google.fr</a> est un programme de publicité en libre-service à la performance qui permet aux annonceurs de poster leurs liens sponsorisés sur des domaines web tels que Google, AOL, Ask Jeeves, etc. Utilisant essentiellement le format texte, le programme fonctionne au PPC, et les enchères peuvent être calculées en fonction de critères démographiques, désormais disponibles sur une sélection de sites du réseau de contenu Google.

Note Si vous êtes une agence, vous devez aussi vous prémunir de mandats qui vont vous permettre d'agir en nom et place du client. Sur ce dernier, vous pouvez être désigné comme mandataire non payeur, ou mandataire payeur. Ces mandats vont aussi vous permettre d'obtenir les "remises agences" accordées par les moteurs. Vous devez aussi vous occuper des **Ordres d'insertions** qui vont définir le budget pour votre campagne. Ces derniers ne sont plus nécessaires chez Google, mais ils sont obligatoires chez Yahoo et MSN. Une fois tous ces papiers envoyés par télécopie, vous devez patienter pour l'ouverture du compte.

> Un extrait de la **Loi Sapin** <a href="http://www.lexinter.net/JF/loisapin.htm">http://www.lexinter.net/JF/loisapin.htm</a> du 29 janvier 1993 stipule justement les droits et obligations relatives à "la prévention de la corruption et à la transparence de la vie économique" pour toutes **prestations de publicité**. La loi Sapin a notamment transformé les règles économiques du marché publicitaire, régissant de façon très stricte les relations entre annonceurs, agences et médias et bouleversant le modèle d'entreprise des agences de publicité qui jusque-là se rémunéraient partiellement grâce à des marges inconnues des annonceurs et rétrocédées par les médias.



Note Une bonne structure de compte va vous permettre de mieux suivre votre campagne, en facilitant vos optimisations futures. En premier lieu, prenez l'habitude de travailler avec une seule campagne, découpée en différents adsgroup (="groupes d'annonces") qui eux sont composés de vos mots clés et de vos annonces. Mais pour mieux répartir son budget en fonction de la campagne et de votre attente en terme de ROI, vous pouvez découper votre travail en différentes campagnes afin de pouvoir fixer un budget quotidien pour chacune d'entre elle. Ainsi vous découpez votre offre en fonction du budget alloué à chaque produit ou groupe de produit et votre campagne pourra fonctionner au mieux.

N'hésitez pas à vous accompagner de Google Analytics, d'autant que l'interface de gestion de Google AdWords prévoie déjà d'y lier les annonces! Et surtout, n'oubliez pas de déposer tous les tags de performance qui vont vous permettre de remonter les ventes, le CA, les inscriptions newsletter, etc. directement au sein de votre campagne Search.

A ses débuts, Google AdWords, et son programme AdSense, qui permet aux webmasters de déployer les liens sponsorisés sur leur site et de gagner ainsi de l'argent, a rencontré quelques faiblesses, notamment en termes de clics incorrects ou frauduleux. Ces problèmes n'ont pas disparus pour autant, et un annonceur risque toujours qu'un de ses concurrents se soit amusé à cliquer des centaines de fois sur ses liens payants. Google prévoit tout de même une procédure à suivre, si une adresse IP particulière semble indiquer une activité incorrecte:

- Comment gérer un rapport clics incorrects ? <a href="https://adwords.google.com/support/bin/answer.py?answer=44008">https://adwords.google.com/support/bin/answer.py?answer=44008</a>
- Plusieurs clics issus de la même adresse IP indiquent-ils une activité incorrecte ? https://adwords.google.com/support/bin/answer.py?answer=74435&cbid=-suj455qp2ywn&src=cb&lev=answer

Les liens sponsorisés Google AdWords se distinguent en 2 branches:

- les top sponsored links,
- et les right sponsored links

Leur différence: Le coût! Rappellez-vous, le Triangle d'Or schématisant les zones préferentielles des internautes. Mais quel que soit l'emplacement choisi (top ou right), vous devrez respecter ces standards de rédaction:

• Titre : 25 caractères,

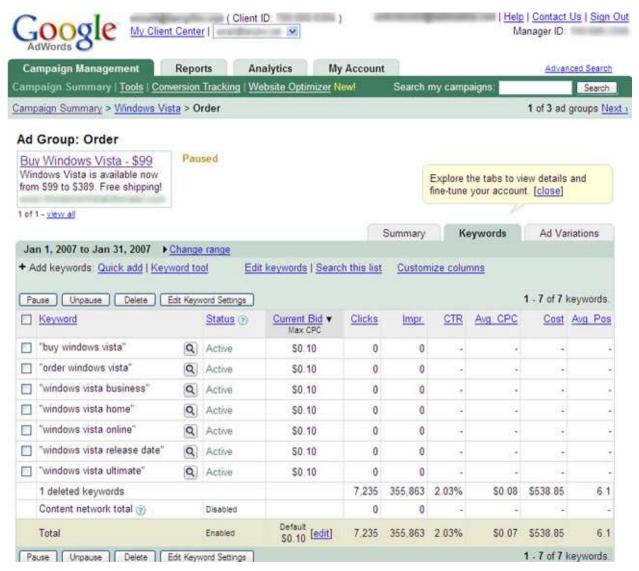
• Ligne 1 : 35 caractères,

• Ligne 2 : 35 caractères,

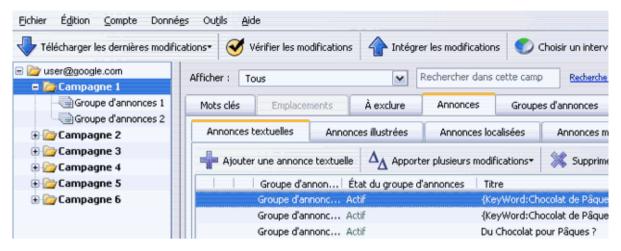
• URL affichée : 35 caractères.

# **▼ 2.2.4.1 Gérer ses campagnes AdWords**

• L'espace client web, qui permet d'administrer toutes ses campagnes:



• Google AdWords Editor, dont l'intérêt principal de est de pouvoir rapatrier son compte AdWords sur son ordinateur, travailler ses campagnes sans être connecté avant de les republier en ligne, éditer les mots clés, les CPC, les titres, les textes, les liens des annonces mais aussi détecter les erreurs des annonces (mots interdits...), tout ceci est accessible depuis ce logiciel. Outre le fait que l'on puisse travailler sans être forcément connecté à Internet, il peut être intéressant d'utiliser des outils dont on ne dispose que sur son ordinateur, et qui facilitent ainsi les traitements, comme Excel par exemple. http://www.google.com/intl/fr/adwordseditor/



#### La planification avancée des campagnes AdWords

Sortie en Octobre 2008, la dernière évolution de Google AdWords vous permet de réaliser **une planification avancée des annonces**, avec la capacité d'enchérir ou de sous-enchérir en fonction d'une planification temporelle. Cela vous permet donc de définir des tranches horaires sur lesquelles vous tenez à être présent, mais à moindre en coût, et des tranches sur lesquelles vous voulez être en "top position".

Vous pouvez ainsi maximiser votre présence sur les périodes qui marchent le mieux pour vous, sans laisser de coté les périodes creuses de votre business. Les périodes que vous laissez vides en temps normal (de minuit à 5 h par exemple) sont laissées vides par les autre annonceurs **dans la majeure partie des cas** ; donc, une bonne manière de se positionner à moindre coût !

Rendez-vous sur "Paramètres de votre campagne > Activer la planification des annonces":

#### Récapitulatif des campagnes > categories > Modifier les paramètres de campagne > Planification des annonces

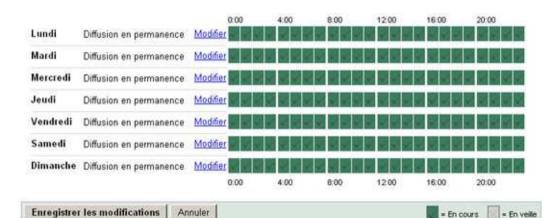
#### (9) Planification des annonces

Modifiez le jour et l'heure et vérifiez la mise à jour de la planification. Lorsque la planification vous satisfait, cliquez sur « Enregistrer les modifications » pour activer la planification des annonces. Pour ajuster votre prix au cours d'une période, passez en mode avancé. Vous pouvez toujours revenir en arrière. (9)



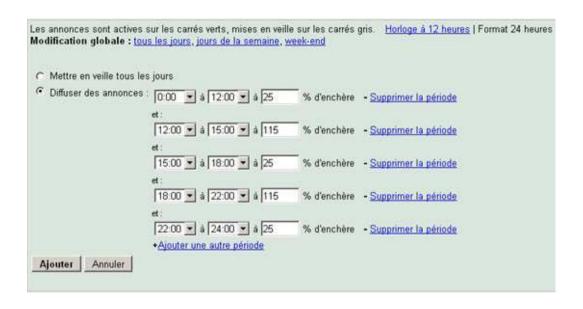
Q Par défaut, vos annonces sont maintenant diffusées tous les jours et à toutes les heures indiquées par les carrés verts ci-dessous.

Les annonces sont actives sur les carrés verts, mises en veille sur les carrés gris. Horloge à 12 heures | Format 24 heures Modification globale : tous les jours, jours de la semaine, week-end



Par exemple, vous pouvez définir la chose suivante, sur la base d'une enchère par défaut de 1€:

- de minuit à 10h du matin, travailler avec une enchère de 10%, soit consommer 10cents,
- de 18h à 20h, travailler avec une enchère de 150%, soit consommer 1,5€.





#### **▼ 2.2.4.2 Google AdSense & Google Affiliate Network**

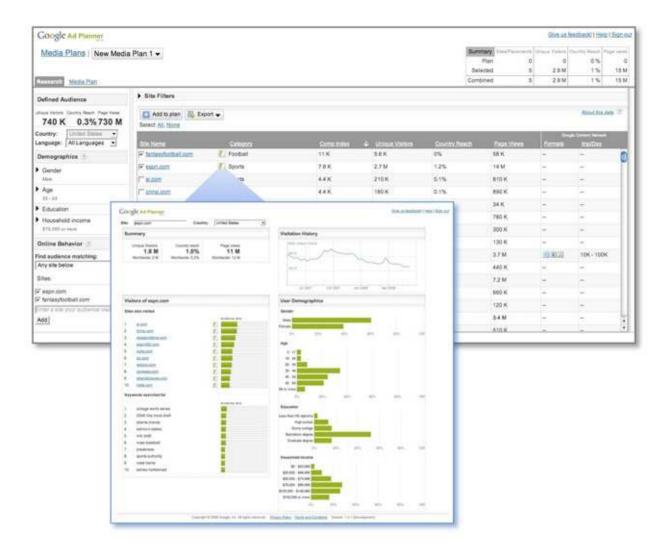
L'affiliation est le principe par lequel un site marchand ou commercial propose à un réseau de sites partenaires affiliés de promouvoir par le biais de bandeaux, ou de liens textes, ses produits ou ses services.

Google AdSense <a href="https://www.google.com/adsense/">https://www.google.com/adsense/</a> est la régie publicitaire de Google utilisant les sites Web comme support pour ces annonces. Il permet donc aux souscripteurs AdWords de s'afficher, et de communiquer hors des frontières du moteur. Ce système fonctionne d'ailleurs sur un système donnant-donnant, le site affichant les annonces étant rémunéré selon un ratio pré-défini. Suite au rachat de DoubleClick, Google a annoncé la fin du parrainage AdSense pour la fin Août 2008. Ce programme est désormais remplacé par la plateforme d'affiliation sous l'appellation Google Affiliate Network <a href="https://www.google.com/ads/affiliatenetwork/">http://www.google.com/ads/affiliatenetwork/</a>.

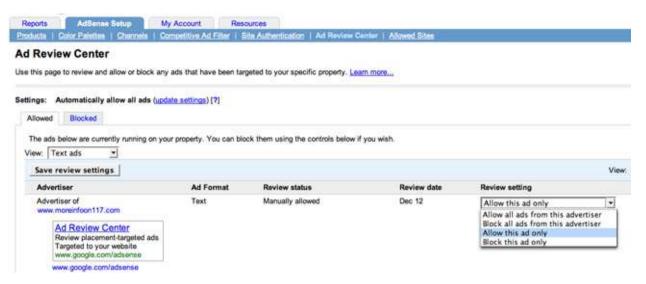
### **▼ 2.2.4.3 Gérer ses campagnes AdSense**

• Google AdPlanner <a href="https://www.google.com/adplanner/">https://www.google.com/adplanner/</a> est ciblé vers les media planner et les agences de communication qui peuvent entrer un profil démographique d'audience désiré ainsi que des exemples de sites sur lesquels leurs publicités s'affichent déjà. L'outil va alors utiliser ces données pour proposer d'autres sites web intéressants par rapport à la cible visée. Un outil que proposent déjà bon nombre de sociétés spécialisées (Comscore, Nielsen, etc.) celui-ci étant gratuit. L'outil complète la gamme AdReview Center et AdManager.

Plus d'infos ici: http://www.conseilsmarketing.fr/referencement/faites-de-la-veille-concurrentielle-avec-google-ad-planner



• Google AdReview Center est en réalité une section de l'espace de gestion Google AdSense, située dans le "Competitive Ad Filter" (="Filtre des annonces de la concurrence"), et informe des publicités qui visent directement les sites avec des "placement-targeted ads" (="ciblage par emplacement"), ce qui donne un certain contrôle sur ses campagnes publicitaires, permettant de les bloquer ou de les autoriser.



Note Google recommande d'autoriser automatiquement les annonces et de seulement de les bloquer pour des raisons justifiables. Car Google vous demandera la raison de votre action, et utilisera cet argument pour communiquer avec l'annonceur en question, en vue d'optimiser ses campagne publicitaires.

• Google Analytics, déjà cité plus haut, intègre depuis la v3 des statistiques très détaillées sur l'évolution des campagnes AdSense:



Grâce à ces nouvelles données comportementales de vos visiteurs, vous serez en mesure de prendre des décisions plus éclairées sur la façon d'améliorer l'expérience des utilisateurs sur votre site et d'optimiser vos ensembles d'annonces AdSense pour augmenter votre chiffre d'affaires potentiel.

#### **2.2.4.4** L'impact de la suppression des cookies

Les cookies sont devenus depuis des années indispensables à la bonne gestion d'un site internet, et ce notamment pour de la mesure d'audience. On sait qu'une partie non négligeable des utilisateurs possède, consciemment ou pas, un outil de suppression automatique, ou régulière des cookies (Firefox >> "Effacez mes traces", CCleaner, etc.). D'après Comscore, 30% des internautes suppriment leurs cookies au moins une fois par mois, ce qui multiplierait jusqu'à 2,5x l'audience réelle d'un site web. Nielsen s'accord également à dire que les "site centric" surestiment toujours les vrais chiffres. l'IAB Interactive Advertising Bureau ne l'entendant pas de cette oreille a demandé que ce problème d'envergure soit rapidement étudié.

Manifestement, il existe donc un **décalage** entre le besoin des annonceurs de profiter de toute la puissance d'Internet pour segmenter et affiner leurs campagnes de liens sponsorisés, et l'impossibilité relative qu'ont les fournisseurs de mesures d'audience de leur donner des chiffres **fiables à 100**%. Pour ceux qui examinent leurs taux de conversion, cela signifie que la partie "analyse des data" (trafic quantitatif, etc.) ne doit pas être la base exclusive de leur travail, et qu'il doit se coupler à d'autres analyses plus fines: "audit de parcours client", "split a/b testing", etc.

# ¬ Chapter 2: Optimiser la pertinence de son site web

Avoir beaucoup de contenu, c'est mieux que quelques lignes de description sur une page.

Il est évident que, pour qu'un site soit bien positionné dans les moteurs de recherche:

- il doit avoir beaucoup de contenu, intéressant, structuré, et mis à jour régulièrement.
- il ne doit pas contenir de "liens cassés" (="broken links").
- il doit être facilement accessible, et éviter d'utiliser des technologies non-indexables par les moteurs.

# ▼ 1 Choisir son hébergement

Google aime les sites qui se chargent rapidement. Assurez-vous donc que votre hébergement tient la route et que vos pages web ne sont pas trop chargées (évitez la surabondance des rich-media, de multiplier les widgets, images, plugins et autres scripts, allégez le nombre de requêtes d'une page pour optimiser la bande passante et la charge du serveur, etc.). Cela ravira également vos visiteurs et devrait améliorer les statistiques de trafic de votre site.

Note Sur un petit site, il est intéressant de **regrouper les scripts** javascript, **et les styles** CSS **dans des fichiers communs**. Moins il y aura de fichiers, moins il y aura de requête. C'est moins conseillé sur de gros sites, car on risque d'y rencontrer des conflits de version, et surtout d'obtenir des fichiers scripts et de style gigantesques!

La vraie solution consiste à **regrouper les scripts intelligement**. Par exemple, on s'aperçoit que le script 1 et le script 2 sont souvent utilisés ensemble, on va les mettre dans un même fichier. Idem pour les CSS.

L'idéal est de disposer d'un **serveur dédié**, si vous avez pas suffisamment de budget pour un dédié, choisissez un **serveur mutualisé de qualité**. Pensez également à mettre en place un système de cache pour améliorer les performances de votre site, et afin de réduire la consommation de celui-ci sur votre hébergement.

Si vous utilisez WordPress, vous pouvez activer le cache par défaut dans le fichier de configuration, et il existe également de très bons plugins pour cela, notamment **WP-SUPER-CACHE** for WordPress.

# ▼ 2 Optimiser la structure interne de son site web

Une page type est assez facile à décrire: il suffit de la composer naturellement telle que vous l'auriez fait si vous ne saviez pas qu'il fallait l'optimiser! Cela peut paraître stupide, mais les meilleures

pages sont celles contenant des phrases bien construites, avec bien sûr vos mots-clés. Par contre, évitez les listes de mots-clés, cela ne fait pas très sérieux pour le visiteur, et si un robot d'indexation tombe dessus, vous risquez d'être détecté et par conséquent blacklisté!

#### Evitez:

- d'écrire 3 fois à la suite vos mots-clés,
- 1/3 du contenu ne doit pas être dédié exclusivement à vos mots-clés, mais composez des textes humainement lisibles.

Enfin sachez que la longueur de la page est importante. En général, il ne faut pas concevoir des pages trop longues:

- essayez de ne pas dépasser les 20 Ko (sans compter les images), les pages trop longues, et trop lourdes, ne sont pas lues par les internautes qui sont toujours pressés!
- 300 mots par page minimum, pour créer un contenu intéressant à indéxer (si vous avez trop de texte, le poids des mots-clés diminue).

# **▼ 2.1 La proéminence des mots-clés**

Sans oublier d'écrire tout d'abord pour ses lecteurs, il est bon de leur être un peu infidèle en essayant d'optimiser le contenu pour les moteurs de recherche. Il faut essayer, sans exagérer, de glisser des mot-clés importants tout au long d'une page. Les mots importants du site doivent être systématiquement stockés dans:

- le nom de domaine.
- l'URL,
- et le titre de la page.

Pour les bots, le contenu présent au début de la page HTML générée a plus d'importance que le reste de la page. Il est donc important d'avoir un template qui charge d'abord le contenu des articles avant de charger la ou les barres latérales (comme dans le cas de WordPress par exemple). Bien sûr, si vous avez une template qui valide le test du W3C <a href="http://validator.w3.org/">http://validator.w3.org/</a>, c'est mieux mais ce n'est pas indispensable. Preuve en est, le site google.com ne respecte même pas ces standards!

Pour vous en rendre compte, calculez votre **proéminence**. La proéminence d'un mot est une mesure de sa **distance relative par rapport au début du contenu texte**. Elle fait partie des critères d'analyse de Google, et s'exprime en pourcentage: plus le mot-clé est situé dans l'en-tête du texte, plus ce pourcentage sera élevé. Il est donc important d'essayer dans la mesure du possible de positionner les mots les plus stratégiques d'une page plutôt vers le début.

Note Les moteurs semblent ignorer les positionnement des contenus depuis les feuilles de style CSS. Vous pouvez donc optimiser le positionnement du texte dans votre page HTML, en saisissant les textes dans un ordre optimisé pour le référencement et en les rapprochant un maximum de la balise <body>; puis afficher ce contenu dans le bon ordre, pour les visiteurs humains, les styles CSS se chargeant de faire apparaître les bons blocs aux bons endroits.

#### **2.1.1** L'écriture par pyramide inversée

Essayez d'utiliser la technique journalistique de l'écriture par pyramide inversée, ou inverted pyramid writing: utilisez le premier paragraphe pour résumer l'ensemble de l'article de la page, ce qui vous permettra d'y glisser la plupart de vos mots clés stratégiques.

Note Plus d'infos ici: http://en.wikipedia.org/wiki/News writing.

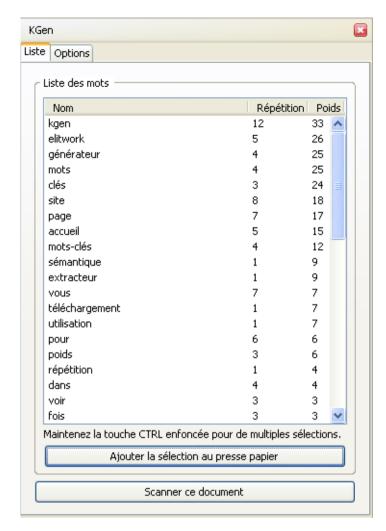
L'autre avantage est que Google utilisera souvent ce premier paragraphe pour constituer le snippet de ses SERP, si la balise meta description n'a pas été saisie. Si vous avez une page qui liste le début de chaque article, comme cela est souvent le cas sur un blog, ce paragraphe sera riche en mots-clés et par conséquent ce sera également le cas de cette **page sommaire**, dont le poids sémantique sera très fort!

# **▼ 2.1.2 Calculer votre proémicence**

Il existe différents outils dédiés vous permettant de calculer cette donnée ; c'est notamment le cas de KGen, plugin Firefox, disponible gratuitement ici: http://kgen.elitwork.com/.



Voici un exemple de scan effectué:	



# ▼ 2.2 l'URL-Rewriting, ou comment définir des règles de réécriture

#### **▼ 2.2.1 Définition**

L'URL-Rewriting, parfois noté UR, est une technique consistant à faire réécrire, par le serveur web, sous forme plus simple des URL complexes. Ainsi, en apparence, les URL deviennent lisibles pour les utilisateurs, et surtout pour les bots et les moteurs de recherche classiques.

Pour améliorer son réferencement de façon conséquente, il faut vraiment adopter cette technique de façon quasi systématique pour toutes les pages contenant des arguments, et profiter d'y écrire des mots-clés importants. Même si l'article ci-joint nous fait croire le contraîre: <a href="http://googlewebmastercentral.blogspot.com/2008/09/dynamic-urls-vs-static-urls.html">http://googlewebmastercentral.blogspot.com/2008/09/dynamic-urls-vs-static-urls.html</a>.

L'adresse <a href="http://www.monsite.com/page.php?nom=fiche-produit&titre=table-bois&id=19824&page=2">http://www.monsite.com/page.php?nom=fiche-produit&titre=table-bois&id=19824&page=2</a>
deviendrait <a href="http://www.monsite.com/table-bois-19284">http://www.monsite.com/table-bois-19284</a>

#### **▼ 2.2.1.1 Quelques erreurs à ne pas commettre**

Les moteurs et navigateurs ont tendance à se limiter à 255 caractères pour ce qui est de la taille des URL. Généralement, il faut travailler entre 50 et 200 caractères, pour fournir de la matière à travailler suffisante pour les bots. Plus les mots-clés seront à gauche dans la liste des caractères, en raison du sens de lecture, plus ils seront pris en compte par les moteurs. Et donc, meilleur sera le positionnement du site sur les SERP.

Il faut donc préférer une adresse de ce type: <a href="http://www.site.com/produit-lambda-1.htm">http://www.site.com/produit-lambda-1.htm</a> à <a href="http://www.site.com/index.php?id=1&titre=produit-lambda">http://www.site.com/produit-lambda-1.htm</a> à <a href="http://www.site.com/index.php?id=1&titre=produit-lambda">http://www.site.com/produit-lambda-1.htm</a> à <a href="http://www.site.com/index.php?id=1&titre=produit-lambda">http://www.site.com/index.php?id=1&titre=produit-lambda</a> même si les mots importants apparaissent dans le "Query String".

L'exemple le plus pertinent, c'est de se rendre compte que http://www.produit-lambda.com sera encore mieux réferencé que ces 2 URL.

Note Même si la plupart des browsers vont traduire les caractères spéciaux en codes ASCII, il faut écrire des caractères non accentués, sous peine d'être pénalisés en terme de réferencement naturel.

#### **▼ 2.2.2 Les étapes à suivre pour mettre en place l'UR**

Tous ces points sont listés en détail dans la suite du document.

- Attention, cette technique ne fonctionne pas chez tous les hébergeurs, même payants: Free, Le Relais internet, etc. Il convient donc de se renseigner auprès de l'hébergeur auparavant.
- Identifier les pages dynamiques dont l'URL comporte des paramètres, et choisir un nouveau schéma d'URL "propre".
- Ecrire les règles de réécriture dans le fichier .htaccess adéquat.
- Changer tous les liens vers chaque fichier dont l'URL a changé.

```
      Réécrire
      <a href="index.php?nom=<?=$var?>&id=<?=id?>">...</a>

      en
      <a href="page-<?=$var?>-<?=id?>">...</a>
```

Mettre à jour le site et vérifier que tout fonctionne!

## ▼ 2.2.3 Les avantages de cette technique

Cette technique cumule de nombreux avantages:

optimisation du réferencement naturel dans les moteurs de recherche

car on peut y inclure des mots clés importants, et éviter des URL un peu trop similaires (en parallèle, il faut également saisir des <title> et <description> différents d'une page à l'autre)

URL propres qui ne sont pas parasitées par des variables

on se débarasse des éléments "?" et "&"

Meilleure sécurité, si l'URL est bien choisie, un internaute ne peut pas savoir que la page est dynamique

on peut saisir des extensions neutres .html ou .htm

Possibilité de changer les adresses physiques des pages tout en gardant la même URL virtuelle

évolutibilité du site plus aisée

Eviter les pages d'erreur 404

pour éviter de perdre à la fois en "bonus" de réferencement et en trafic

Note Attention, si une même page est traduite plus de 2 fois via un fichier .htaccess, cela est considéré comme une technique de spam-indexing, et peut vous désindexer totalement de Google!

#### **▼ 2.2.4 Vérifier la compatibilité avec votre hébergeur 1/2**

La première chose à faire est bien évidemment de s'assurer que le serveur qui héberge votre site permet d'utiliser la réécriture d'URL. Tout dépend, dans un premier temps, du type de serveur utilisé. Voici un résumé des possibilités de réécriture d'URL sur les 2 serveurs web les plus courants:

APACHE	Géré par le module mod_rewrite, un module standard d'Apache à partir de la version 1.3.27	Le module mod_rewrite doit être actif. Le fichier de configuration d'Apache httpd.conf doit contenir cette ligne:
		LoadModule rewrite_module libexec/mod_rewrite.so
		ainsi que celle-ci:
		AddModule mod_rewrite.c
IIS Microsoft	En ASP: réécriture possible par des filtres ISAPI, commercialisés par diverses sociétés (payants). Il existe une version gratuite light de ISAPI. Attention, ISAPI consomme environ 20 à 30% de CPU.	Le paramétrage des règles de réécriture est spécifique à chaque composant.
IIS Microsoft	En ASPX (.NET), sur tous les serveurs supportés : des fonctions sont disponibles comme RewriteURL(), etc. qui prennent en charge la réécriture d'URL.	Des codes prêts à compiler pour exploiter ces capacités sont fournis par Microsoft ou via des projets open source comme:  codeproject.com. Aucune méthode standard n'a été prévue pour définir les règles de réécriture mais une utilisation pratique consiste à les paramétrer directement dans le web.config (fichier de configuration de l'application ASP.Net, présent notamment à

la racine du site), qui est standardisé en XML.

Si votre site est hébergé sur un serveur dédié,

Vous avez accès vous-même à la configuration du serveur. Dans le cas d'un serveur Apache, vous pouvez donc modifier le fichier de configuration afin d'activer le support de la réécriture d'URL. Pensez à redémarrer Apache après avoir modifié le fichier de configuration.

Si votre site est hébergé sur un serveur mutualisé,

Il n'est pas garanti que votre hébergeur ait activé le support de la réécriture d'URL, principalement pour des raisons de sécurité. Parfois, cette activation change même d'une offre d'hébergement à l'autre, chez un même fournisseur, comme chez OVH par exemple.

Si votre site est fourni par un hébergeur gratuit,

Il y a peu de chances que la réécriture d'URL soit possible. Il vaut mieux investir dans un hébergement payant en plus d'un nom de domaine adéquat, les avantages sont réellement nombreux pour effectuer un bon référencement.

Note parfois, lors du dépôt du fichier .htaccess sur le serveur, selon sa configuration, il peut disparaître. Parfois, il est simplement invisible, parfois, il est supprimé automatiquement.

Cela peut prêter à confusion, faites en part à votre hébergeur. N'oubliez pas non plus de vérifier que votre client FTP ne vous empêche pas de voir les fichiers cachés!

#### **▼ 2.2.5 Vérifier la compatibilité avec votre hébergeur 2/2**

L'exemple qui suit se focalise exclusivement sur Apache. Pour vérifier si le module mod rewrite d'Apache est activé, il vous suffit de suivre les points suivants:

1. Créez un répertoire nommé test que vous placerez à la racine de votre site, donc accessible via l'adresse <a href="http://www.site.com/test/">http://www.site.com/test/</a>

et créez-y une page index.html avec le code HTML suivant:

```
<html>
<head>
<title>Index</title>
</head>
<body>

La redirection fonctionne
</body>
</html>
```

2. Dans ce même répertoire, créez un fichier nommé .htaccess contenant les lignes suivantes:

```
Options +FollowSymlinks
RewriteEngine on
RewriteRule ^test\.html$ /test/index.html [L]
```

3. Rendez-vous à l'adresse: http://www.site.com/test/test.html

Que s'affiche t-il à l'écran?

#### Une Erreur 404?

Si le navigateur affiche un message d'erreur indiquant que le fichier nommé test.html n'existe pas à cet endroit sur votre site, alors votre hébergeur n'autorise sans doute pas la réécriture d'URL : contactez-le pour lui demander !

#### Une Erreur 500?

Votre site est totalement bloqué, aucune page ne peut être affichée, et vous avez un message indiquant "Erreur 500". Dans ce cas, il vous suffit de retirer le fichier .htaccess qui est incompatible avec votre hébergeur.

#### La redirection fonctionne?

Sinon, vous devriez voir le texte "La redirection fonctionne", ce qui signifie qu'en demandant à voir le fichier test.html, qui n'existe pas physiquement sur le serveur, le serveur vous affiche le contenu du fichier index.html, qui, lui, existe bien. C'est le principe même de la réécriture d'URL et donc la preuve que votre serveur gère bien la réécriture d'URL.

#### **₹ 2.2.6 Définir les schémas d'URL**

Voici un exemple de pages avant la redirection:

```
produit.php?id=387&cat=5&promo=1
produit.php?id=12&page=2&cat=8
```

Le principe de l'UR consiste à trouver les schémas des URL à partir de leurs formes communes.

Dans notre exemple, les produits sont accessibles selon 3 types d'URL:

```
id + cat
id + cat + page
id + cat + promo
```

A partir du moment où vous avez identifié ces "schémas d'URL", vous devez choisir un nouveau format d'URL "propre". En général on fait apparaître un nom de fichier avec l'extension .html ou .htm mais sachez que vous pouvez mettre ce que vous voulez, cela n'a aucune incidence sur la prise en compte des pages par Google. En effet, quelle que soit l'extension que vous aurez choisie, la page restera une page respectant la norme HTML.

Le nom du fichier sera formé d'un préfixe et/ou d'un suffixe, et des valeurs des variables, que ce soient des chiffres ou des lettres.

Note Profitez de cette étape pour bien réfléchir en fonction du référencement, car vous pouvez utiliser ici des mots-clés intéressants dans les URL de vos pages, qui soient plus parlants pour les internautes et donc pris en compte par les moteurs de recherche.

Voici des proposition de réécriture:

```
produit-387-5-promotion.html
```

Note Pour séparer les différentes parties de l'URL, vous devez choisir un séparateur, comme le tiret dans notre exemple. Il est plus efficace pour le référencement de choisir un caractère qui soit considéré comme un séparateur de mots par Google. Ainsi, vos URL pourront contenir des mots-clés, ce qui est pris en compte sans soucis par Google.

Quelques caractères sont acceptés, le tiret étant le plus utilisé:

- Le tiret -
- La virgule ,
- Le point .
- La barre oblique "slash" /

Malgré ce qui peut etre lu ici et là sur Internet, Matt Cutts nous confirme également d'autres choses mais pour ma part il ne s'agit pas de nouveautés

• La barre verticale "pipe" |

On peut tout a fait cumuler ces caractères, libre au réferenceur de déterminer les choix les plus lisibles:

produit.387-5|promotion.html

Attention, les caractères suivants sont déconseillés, principalement car ils ne permettent pas à des moteurs comme Google de discerner des séparations entre les mots:

• Le tiret bas ou "underscore" \_

A ce sujet, Matt Cutts a annocé courant 2007 pendant le WordCamp, que Google allait prochainement considérer l'underscore comme un séparateur. Ce qui ne pourra que profiter à Wikipedia, qui étrangement abuse des underscores.

- Le signe dièse #
- Le plus +
- · L'esperluette &
- . L'arobase @
- Le point d'interrogation ?
- Le signe dollar \$
- Les caractères accentués et l'espace

En résumé, il est beaucoup plus simple d'utiliser le simple tiret "-", la barre oblique pouvant parfois porter à confusion avec les répertoires, quant à la barre verticale "pipe" n'est pas très connue des internautes. Enfin, l'underscore pose des soucis avec Google.

Autre information, l'extension n'impacte en rien le référencement: .htm, .html, .php, .aspx, même s'il n'y a pas d'extension.

Note Attention aux répertoires virtuels. Si l'URL apparente aurait la forme /article/8126/2.html au lieu de /article-8126-2.html, dans ce cas, le navigateur "estime" que la page se trouve dans un répertoire /article/8126 qui n'a pas d'existence réelle sur votre site. Toute tentative de résolution de liens relatifs se fera donc à partir de ce répertoire inexistant et sera vouée à l'échec. Pour éviter cela, deux solutions se présentent:

- Utiliser des liens absolus,
- ou, faire usage de la balise <base> en HTML

```
<head>
    ...
    <base href="http://www.votresite.com/repertoire/" />
</head>
```

Quid des mots de liaison?

#### 2.2.6.1 Un cas concrêt avec WordPress

WordPress est un moteur de blog open-source en Php/MySQL qui permet de mettre en place un blog facilement.

Dans son panneau d'administration, dans "options > permaliens", on va pouvoir définir des règles de réécriture permanentes. Cette option de WordPress offre un large éventail de possibilités:

#### http://www.site.com/?p=123

C'est le type d'URL appliquée par défaut sur une nouvelle installation de la plateforme et certainement l'une des plus mauvaises à utiliser. Si le format permet d'avoir une adresse courte, il n'a aucun intérêt pour un moteur. 123, 1052 ou 4 ne sont pas des informations pertinentes et ne permettent pas de "renforcer" le poids du contenu de l'article.

#### http://www.site.com/categorie/123

Ici la structure est déjà plus intéressante dans le sens où le terme du dossier ("categorie" dans le cas présent) va déjà donner une information sur le type de contenu présent (à moins que vous ne classiez des recettes de cuisines dans une catégorie auto-moto). Mais comme dans le premier exemple, le format numérique derrière ne servira à rien d'un point de vue des moteurs.

#### http://www.site.com/yy/mm/dd/titre du billet

Un autre des formats souvent rencontré est celui basé sur la date et le titre. Le titre va aider fortement à apporter de l'info pertinente pour le moteur, quant à la date elle ne sera pas forcément d'un grand intérêt sauf si vous prévoyez la publication de séries d'articles portant le même titre (cela permettra de différencier chaque billet de la série). D'un point de vue "humain", la date pourra aussi influencer la personne qui effectue une recherche: si vous apparaissez en 1er résultat mais que dans le lien apparait la date de votre article qui a été rédigé en 2000, peut-être que vous ne serez pas jugé comme pertinent par rapport à une situation actuelle.

#### http://www.site.com/dossier blog/categorie/titre du billet

Placer la catégorie contenant l'article + son titre est certainement une bonne chose pour renforcer la densité autour d'un sujet. Mais vous pouvez très bien avoir votre blog ailleurs que sur la racine de votre hébergement, c'est même très souvent le cas. Et là il faut un petit peu voir la longueur de l'url finale sachant que dans la limite du possible mieux vaudra rester sous la barre des 100 caractères (on peut aller plus loin mais il y aura alors une certaine dilution du contenu).

#### http://www.site.com/titre du billet

Sûrement le format qui possède le meilleur ratio lien court/pertinence pour un moteur, surtout si vous êtes focalisé sur une niche bien précise.

En conclusion, si vous avez un nom de domaine court, représentatif de votre activité, et que votre blog est à la racine, peut-être qu'ajouter les catégories renforcera votre positionnement (pour autant que le terme utilisé pour la catégorie soit en rapport étroit avec l'article). Si par contre vous avez besoin d'appliquer de longs titres à vos articles il sera peut-être préférable de mettre le moins de dossier possible entre la racine du domaine et ceux-ci.

Il existe des plugins des redirections intéressants.

Google conseille d'éviter une profondeur de 4 max. Donner un exemple.

#### **₹ 2.2.7 Rédiger les règles de réécriture**

Maintenant que nous avons déterminé les différents schémas d'URL, il reste à écrire les règles de réécriture qui vont indiquer au serveur comment interpréter chacun de ces schémas.

En reprenant l'exemple précédent, voici le contenu du fichier .htaccess situé à la racine du dossier "produits" étudié:

```
# Répertoire : /produits/
# Le serveur doit suivre les liens symboliques
Options +FollowSymlinks
# Si un internaute visite une page qui n'existe pas, il est redirigé
ErrorDocument 404 http://www.site.com/index.htm

# Activation du module de réécriture d'URL
RewriteEngine on

# Produit simple
RewriteRule 'produit-([0-9]+)-([0-9]+)\.html$ /produits/produit.php?id=$1scat=$2 [L]

# Produit avec page
RewriteRule 'produit-([0-9]+)-([0-9]+)-([0-9]+)\.html$ /produits/produit.php?id=$1spage=$3scat=$2 [L]

# Produit info promo
RewriteRule 'produit-([0-9]+)-([0-9]+)-promotion\.html$ /produits/produit.php?id=$1scat=$2spromo=$3 [L]
```

Note Il ne doit pas y avoir de retour chariot sur une ligne de règle de réécriture.

Les lignes commençant par le signe dièse # sont des commentaires. N'hésitez pas à en ajouter pour rendre vos fichiers plus compréhensibles: ces lignes sont totalement ignorées par le module de réécriture d'URL.

Chaque fichier .htaccess est spécifique à un répertoire ; nous avons pris l'habitude d'indiquer en haut de ce fichier l'emplacement du répertoire sur le site. Chaque répertoire de votre site devra donc proposer son propre fichier .htaccess.

Bien évidemment, on peut n'utiliser qu'un seul fichier .htaccess à la racine de son site, qui définisse les règles de réécriture de l'ensemble du site. Mais cela risque parfois d'être peu pratique dans l'organisation technique du site.

PAS de redirection invisible (cf. hébergeur) ou de pages sattelites!!

### **▼ 2.2.7.1 Explications du contenu du fichier .htaccess**

### **Explications**

- 1. Les deux premières instructions "Options +FollowSymlinks" et "RewriteEngine on" ne doivent être présentes qu'une seule fois par fichier, avant toute règle de réécriture. "RewriteEngine on" peut être extrêmement pratique, car vous pouvez désactiver en quelques secondes la réécriture d'URL le temps de comprendre le problème: il vous suffit d'écrire "RewriteEngine off" à la place de "RewriteEngine on".
- 2. Il arrive que vous soyiez obligé de supprimer ou de renommer une page, ce qui n'est pas conseillé car tous les moteurs de recherche auront gardé l'ancienne adresse dans leur base. Vous aurez donc des erreurs 404 et une perte de visiteurs. ErrorDocument est donc une methode très importante qui va éviter une perte de trafic, et un problème de réferencement.

Voici les codes erreurs les plus communs:

401	Mot de passe requis (Authorization required)
403	Accès interdit (Forbidden)
404	Page inexistante (Page not found)
500	Erreur interne au serveur (Internal server error). Le plus souvent du à une erreur d'execution
	d'un script

3. La suite du fichier est constituée d'une série de règles de réécriture. Sauf règles complexes, chaque règle est écrite sur une seule ligne et respecte le format suivant:

RewriteRule URL\_REECRITE ANCIENNE\_URL

#### **▼ 2.2.7.2 Les expressions régulières**

Pour écrire convenablement une nouvelle URL, il convient de connaître la base d'écriture des "expressions régulières".

^	Indique le début de l'URL à récrire. Ce caractère est facultatif mais il est plus rigoureux de l'utiliser.
produit-	Cette partie de l'URL n'est pas utilisée directement. Nous aurions pu écrire prod à la place de produit. Ce préfixe peut servir à différencier différents schémas d'URL, et il permet à l'internaute de mieux comprendre l'objet de la page.
()	Les parenthèses servent à encadrer une variable dont la valeur est récupérée dans la 3ème partie de la ligne, ANCIENNE_URL
[0-9]+	Indique que la variable est composée d'un ou plusieurs chiffres.

Ş	Cette partie de l'URL n'est pas utilisée directement. Nous aurions pu écrire prod à la place de produit. Ce préfixe peut servir à différencier différents schémas d'URL, et il permet à l'internaute de mieux comprendre l'objet de la page.
/produits/	Cette partie est parfois facultative (cela dépend de la configuration du serveur). En général il suffit d'indiquer l'emplacement du fichier de manière relative au répertoire dans lequel est situé le fichier .htaccess (donc on peut se passer de cet élément). Attention, chez certains hébergeurs mutualisés (OVH, Sivit, etc.), vous devez indiquer le chemin complet vers le fichier, à partir de la racine du site, comme dans notre exemple.
produit.php	C'est le nom du fichier que le serveur doit utiliser pour afficher la page. C'est le nom d'un fichier qui existe physiquement et qui contient un script, PHP dans notre exemple, de gestion de la page dynamique.
?	Caractère obligatoire précédant la série de variables passées dans l'URL réécrite.
id=\$1	Indique que la variable nommée id prendra la valeur située dans la première paire de parenthèses.
&	Caractère utilisé pour séparer 2 variables dans l'URL réécrite.
cat=\$2	Indique que la variable nommée cat prendra la valeur située dans la deuxième paire de parenthèses.
[L]	Drapeau (option) signifiant "Last", indiquant au module de réécriture qu'il doit s'arrêter. Plus précisément, si l'URL de la page demandée par le visiteur correspond au schéma défini par cette règle, alors le module de réécriture d'URL ne doit pas examiner les autres règles situées dans le reste du fichier .htaccess. Il n'est pas toujours obligatoire mais il permet un léger gain de temps.

Il existe d'autres règles plus complexesen voici quelques unes:

# ([0-9]{1,2})

autorise un ou deux chiffres

# ([0-9]\*)

autorise tous les chiffres, autant de fois qu'on veut

# ([\d]\*)

autorise tous les chiffres, autant de fois qu'on veut

# (-[a-z]\*)

autorise toutes les lettres et tirets, autant de fois qu'on veut

# (-[a-zA-Z]\*)

autorise toutes les lettres majuscules et minuscules

([\w])

correspond à une lettre, accentuée ou non, à un chiffre ou au à "underscore".

# **▼ 2.2.8 Quelques flags utiles**

Les flags sont les éléments déposés en fin de ligne du fichier .htaccess.

[L]	Déjà vu, indiquant au module de réécriture qu'il doit arrêter sa boucle de lecture.
[R=301]	Indique une redirection permanente. La redirection temporaire [R=302] existe également mais promet un moins bon réferencement. Attention, une redirection différente de 301 ou 302 risque une désindexation de la page, voire un blacklistage complet du site. On peut également écrire [R=permanent]. Ce flag nécessite de préciser une nouvelle URL absolue (http://site.com/)
[F]	C'est un "flag d'erreur". Car il affiche une erreur 403 "Forbidden". ex: RewriteRule ^secret.html\$ - [F] ( pas de réécriture vu le deuxième argument - ) Attention, cela désindexe la page
[G]	C'est un "flag d'erreur". Car il affiche une erreur 410 "Gone" Attention, cela désindexe la page
[QSA]	Le "Query String Append" indique que toutes les adresses comportant des variables seront redirigées intégralement. En somme, il va réécrire les paramètres en entrée sur l'URL de sortie.  Utilisée le plus souvent avec le flag [L], comme dans [QSA,L]
[NC]	le "NoCase" indique que l'on fait les comparaisons sans considérer la casse.  RewriteRule ^script\.php\$ programme.php [NC,L]  Cet exemple s'appliquera aussi bien à script .php, SCRIPT.PHP ou ScRiPt .PhP
[N]	A l'inverse de [L], on continue à parcourir entièrement le fichier.Le principe, dès qu'une réécriture est executée, le fichier .htaccess est reparcouru entièrement, avec, comme paramètre d'entrée la nouvelle adresse réécrite. Attention donc à ne pas produire une boucle infinie. On appelle cela le "Redirect Permanent".

[C]	Un peu comme l'idée du [N], on continue à parcourir le fichier. Sauf que cette fois, si une
	réécriture se vérifie, on ne retourne pas au dbut du fichier mais à la règle de redirection suivante.
	Si la règle ne se vérifie pas, toutes les règles qui suivent jusqu'à la première ne comportant pas le
	flag[C]ne sont pas appliquées.

#### **₹ 2.2.9 Les réécritures conditionnelles**

Dans les quelques exemples qui précèdent, nous n'avons vu que des réécritures d'URL inconditionnelles, c'est à dire s'appliquant indépendamment du navigateur, de l'adresse IP ou du domaine émettant la requête. Nous allons maintenant passer à l'étape suivante, à savoir la réécriture sous conditions, à travers quelques exemples concrets.

#### **▼ 2.2.9.1** Une page d'accueil différente selon le navigateur

Voici par exemple un code qui va afficher une page d'accueil différente, selon le navigateur de l'internaute, déterminé avec l'identifiant HTTP USER AGENT.

Ce code va s'avérer beaucoup plus pratique qu'une redirection en JavaScript par exemple, car plus rapide et plus sûr, le JS pouvant être désactivé sur le browser de l'internaute.

```
RewriteCond %{HTTP_USER_AGENT} ^Mozilla.*
RewriteRule ^/$ /complexe.html [L]

RewriteCond %{HTTP_USER_AGENT} ^Lynx.*
RewriteRule ^/$ /simple.html [L]

RewriteRule ^/$ /standard.html [L]
```

Un nouveau mot-clé fait son apparition ici : RewriteCond ou "condition de réécriture". La syntaxe est simple et de la forme:

- si l'internaute est sur Mozilla, alors on affiche complexe.html
- s'il est sur Lynx, on affiche simple.html
- sinon, sur Internet Explorer par exemple, on affiche standard.html

#### **▼ 2.2.9.2 Protéger ses images**

La récriture conditionnelle peut s'avérer intéressante pour, par exemple, protéger les images de son site web. Un tel procédé peut éliminer des liens « sauvages » et économiser de la bande passante.

```
RewriteEngine On

RewriteCond %{HTTP_REFERER} !^$

RewriteCond %{HTTP_REFERER} !^http://www.votredomaine.net/.*$ [NC]

ReWriteRule .*\.(gif|png|jpe?g)$ - [F]
```

En mettant plusieurs conditions à la suite, un "ET" logique est effectué entre elles. Pour que la règle de réécriture soit effectuée, il faut donc que toutes les conditions soient vraies prises isolément. A la première condition "FAUSSE", le moteur de réécriture pointe directement après la règle et ne teste pas les conditions suivantes. Si un "OU" logique est nécessaire, on rajoute le flag [OR] en fin de

ligne, en le combinant aux autres le cas échéant [NC,OR].

Le HTTP\_REFERER permet de récupérer l'adresse de provenance de l'internaute. Mais cela ne fonctionne pas systématiquement à 100%, car certains navigateurs permettent de le masquer, et certains proxies ou firewall ne transmettent pas cette référence.

#### **▼ 2.2.9.3 Bloquer l'accès au site aux visiteurs indésirables**

Tous les robots ne sont pas bénéfiques pour votre sites. Certains d'entre-eux sont des aspirateurs de site, d'autres collectent les adresses email et finissent par remplir votre boîte aux lettres de courrier non-sollicité (spam). Ils ont tous une caractéristique commune: utiliser les resources de votre serveur sans vous apporter aucun visiteur "utile".

Tous ces robots "indélicats" ne respectent pas le protocole d'exclusion représenté sous la forme du fichier "/robots.txt".

#### teste le nom d'un ordinateur hôte spécifique

```
RewriteCond %{REMOTE HOST} ^badhost\.baddomain\.com$
```

### teste le domaine complet (se termine par..., notez l'absence du caractère ^)

```
RewriteCond %{REMOTE HOST} baddomain\.com$
```

#### teste le nom du robot indésirable (HTTP\_USER\_AGENT commence par la chaîne "Bot")

```
RewriteCond %{HTTP USER AGENT} ^Bot.*
```

#### teste une plage d'adresses IP (de 123.45.67.125 à 123.45.67.129 inclus)

```
RewriteCond %{REMOTE ADDR} ^123\.45\.67\.12[5-9]$
```

Attention, sovez très attentifs dans l'écriture de vos règles d'exclusion.

par exemple la condition:

```
RewriteCond %{HTTP_USER_AGENT} Bot
```

est beaucoup trop générique et vous priverait du passage de GoogleBot, ce qui n'est pas le but ici!

Voici un exemple plus concrêt, qui interdit toute visite en provenance de laurion.com et laurion.net. Ce robot est réputé pour ne pas respecter le protocole d'exclusion. Quant à la seconde ligne, elle empêche par exemple toute visite depuis la Chine:

```
RewriteCond %{REMOTE_HOST} \.laurion\.(com|net)$ [OR]
RewriteCond %{REMOTE_HOST} \.cn$ [OR]
RewriteRule ^.*$ - [F]
```

L'exclusion de laurion.com pourrait s'écrire autrement, en bloquant l'accès à son bot:

```
RewriteCond %{HTTP USER AGENT} ^IPiumBot [OR]
```

Certains sites web permettent de vérifier les entêtes reçues très facilement, par exemple: <a href="http://www.wannabrowser.com/">http://www.wannabrowser.com/</a>. Ce site, combiné avec une analyse approfondie de vos fichiers logs, vous permettra de mettre au point vos conditions de réécriture pour les différents visiteurs de votre site.

### **▼ 2.2.9.4 Des règles différentes selon les répertoires**

Un fichier .htaccess placé dans un répertoire régit l'accès à ce répertoire ainsi qu'à tous les sous-répertoires et fichiers de celui-ci. Vous pouvez bien sûr avoir plusieurs fichiers .htaccess dans des répertoires différents, selon les différentes protections ou réécritures que vous désirez appliquer.

Dans le cas d'un fichier .htaccess situé dans un sous-répertoire du site, les règles et conditions remplacent celles définies à l'échelon supérieur. Si votre souhait est d'ajouter des règles de réécriture à celles du niveau supérieur au lieu de les remplacer, ajoutez la ligne suivante juste après le "RewriteEngine on":

```
RewriteOptions inherit
```

Cette instruction spécifie que toutes les règles et conditions définies au niveau supérieur sont héritées, en supplément à celles que vous rajouterez dans le fichier .htaccess.

#### **▼ 2.2.10** Interdire de réferencer tout ou partie du site

Il est parfois nécessaire de protéger l'accès à un répertoire sur un serveur web (ex : répertoire d'administration, contenant des données sensibles) afin d'éviter que n'importe qui puisse y accéder. Autre objectif: cacher l'existence d'une partie du site, et ne pas la réferencer dans Google. Il faut donc avoir recours à un fichier .htpasswd.

Voici un exemple du contenu du fichier .htaccess:

```
AuthUserFile /home/login/admin/.htpasswd
AuthGroupFile /dev/null
AuthName "Veuillez vous identifier"
AuthType Basic

<Limit GET POST>
require valid-user

</Limit>
```

#### et du fichier .htpasswd:

herve:x310HLu5v6mOF

### **▼ 2.2.10.1 Explications sur le couple .htaccess et .htpasswd**

#### **Explications:**

AuthUserFile	C'est le nom et le chemin d'accès du fichier qui contiendra les noms des utilisateurs et les mots de
	passe associés. Ce chemin doit partir de la racine du site. Ici, les mots de passe seront dans

	/home/login/admin/.htpasswd. On peut, et il est même conseillé de choisir un autre nom que .htpasswd pour le fichier qui contiendra le couple utilisateur - mot de passe. Il est également recommandé de mettre le fichier des mots de passe en dehors de l'arborescence du site si l'on en a la possibilité.
AuthGroupFile	Permet de définir un droit d'accès à un groupe d'utilisateur. Cette solution n'est que rarement utilisée pour un site Web. Le reste du temps il pointe vers /dev/null. Il faut que cette ligne soit présente.
AuthName	C'est le texte qui apparaîtra dans la fenêtre demandant le mot de passe.
AuthType	L'authentification est en générale "basic". Les mots de passe sont alors envoyés en clair sur le réseau. Pour sécuriser davantage l'accès, on peut utiliser la méthode d'authentification "digest" qui crypte les mots de passe en MD5. Ce système n'est supporté que par certains navigateurs.
Limit	c'est ici qu'on va indiquer ce qui est autorisé et interdit dans le répertoire. Les commandes GET et POST indiquent la récupération de pages web et la réponse à certains formulaires. POST est utilisé pour autoriser l'upload de fichiers sous le protocole http.
Require valid-user	accepte tous les utilisateurs qui ont un login et un mot de passe dans .htpasswd

On peut remplacer "Require valid-user" par "Require herve jacques", ce qui limite l'accès à un ou plusieurs utilisateurs précis, ici herve et jacques. A noter que les utilisateurs sont séparés par des espaces.

Il existe de nombreux outils sur internet qui permettent de crypter le mot de passe du user, il suffit de googler "generateur+htpasswd".

#### **▼ 2.2.11 Tester les URL**

Modifier les liens de son site après avoir mis en place une technique de redirection d'adresses, peut s'avérer long et fastidieux. Bien entendu, l'idéal est bien sûr de s'attaquer au problème dès la création du site en indiquant de suite les bons liens.

Dans tous les cas, nous vous conseillons d'utiliser un logiciel de vérification des liens à l'intérieur de votre site, comme Xenu's Link Sleuth disponible ici: <a href="http://home.snafu.de/tilman/xenulink.html">http://home.snafu.de/tilman/xenulink.html</a>, sous Windows. Ce type de logiciel agit comme Googlebot: il parcourt vos pages en suivant tous les liens qu'il trouve.

# **▼ 2.3 Les balises HTML**

Lorsqu'un bot va parcourir une page, il va prioriser certaines zones du contenu pour en déterminer l'impact sémantique.

Pour optimiser le poids des mots-clés, il convient donc de les encadrer correctement avec certaines balises clés.

#### **₹** 2.3.1 La balise <title>

Google et les autres bots vont parcourir le contenu du site dans le sens de lecture, les premiers mots auront donc plus d'importance que les derniers.

Cette logique est vraie pour tout le contenu de la page, mais également au sein de chaque balise du code.

La balise <title> étant la première ligne possédant un contenu intéressant, il est donc primordial de la remplir correctement:

```
<html>
<head>
<title>Titre de la page</title>
</head>
...
```

- elle doit contenir des mots très importants, 5 à 8 mots, et peut contenir jusqu'à 100 caractères maximum. Essayer de se limiter à 65 caractères.
- il faut éviter d'y saisir des mots-clés classiques, ou génériques: "accueil, "homepage", etc.
- s'il y a besoin de le spécifier, il faut inscrire le nom du site à la fin,
- et pour éviter de référencer toujours les mêmes mots, il doit être différent du <h1> principal. Il faut essayer de varier le champ lexical.

Note Attention, car cette dernière règle n'est pas souvent appliquée correctement par défaut par les CMS open-source, WordPress ou Joomla! pour ne citer qu'eux.

Voici quelques exemples à suivre ou éviter:

#### Accueil

Trop générique, aucun élément lexical intéressant

#### Blog de David

Mieux, mais pas forcément plus intéressant, car on n'a pas beaucoup plus d'informations lexicales à proposer aux bots.

#### Cours de réferencement web

Un peu plus de termes, et surtout on sait cette fois de quoi il s'agit rééllement.

#### David - Cours de réferencement web

Cette fois, on perd la valeur du titre précédent. Le nom de l'auteur est généralement une information secondaire, à déplacer donc.

#### Cours de réferencement web < internet - David

Le meilleur exemple: des termes correctement hiérarchisés, des informations précises.

#### **2.3.2** Les meta tags

#### **₹** 2.3.2.1 les meta name

Chaque page web peut disposer d'attributs invisibles aux visiteurs, permettant toutefois une qualification du contenu qu'elle contient, notamment pour les bots.

Il existe plusieurs balises <meta> pour une page, qui se situent toutes dans l'entête de la page (X)HTML:

```
<html>
<head>
<title>Titre de la page</title>
<meta name="description" content="description de la page" />
</head>
...
```

La liste qui suit diffuse la plupart des <meta> existants, bien que très peu sont utilisés, puisque les bots ne détermineront que le contenu de la <meta> "description":

```
<meta name="title" content="Titre du site" />
```

C'est le titre de la page. Cette balise est devenue inutile, à l'instar de <title>.

```
<meta name="author" content="David, Alain" />;
```

L(es) auteur(s) de la page.

```
<meta name="category" content="cyclisme, vélo de course" />;
```

Définit l(es) catégorie(s) du contenu de la page, généralement utilisé par certains annuaires.

```
<meta name="copyright" content="Société & cie" />;
```

L(es) proriétaire(s) du contenu de la page.

```
<meta name="subject" content="Le sujet de votre site" />;
```

Définit le sujet du contenu de la page, généralement utilisé par certains annuaires.

```
<meta name="note" content="quelques mots-clés supplémentaires" />;
```

Quelques notes pour compléter la <meta> "description"

```
<meta name="language" content="fr|en|de|it|es|pt|etc." />;
```

Il s'agit de la langue du contenu de la page.

```
<meta name="description" content="texte de description du site ou de la société" />;
```

#### Très important pour le référencement !

Saisir entre 100 à 150-200 caractères maximum, pour afficher un snippet convenable. Pour le rédiger, inspirez-vous des recommandations pour le <title> (ordre des mots importants, etc.). Par contre, essayez de ne pas répéter les termes du "title".

En l'absence de ce <meta>, le bot stockera le début du contenu de la page dans le snippet.

Tout comme la balise <title>, saisissez des descriptions uniques pour chaque page de votre site, afin d'obtenir un bon "snippet", et un affichage précis pour créer du trafic qualifié.

Quelques recommendations:

- Ne pas y saisir un texte de nature commerciale
- Ne pas y mettre une liste de mots-clés sans
- Ne pas tout mettre en majuscule
- Ne pas faire de faute de frappe / d'orthographe
- Ne pas utiliser de superlatifs, "le meilleur site ..."
- Ne pas utilier des articles vou désignant, ou désignant l'internaute ("nous vous proposons", "on fait", "vous pourrez", "tu trouveras", etc.)

Et globalement, tout ce qui s'écarte d'une description objective et soignée d'un site.

Découvert en Septembre 2008, Google serait en train de faire un test sur son SERP, permettant à l'internaute de modifier la taille des snipet, habituellement limité à 2 lignes.

## gnuplot homepage

Gnuplot is a portable command-line driven interactive data and function plotting utility for UNIX, IBM OS/2, MS Windows, ... Gnuplot supports many types of plots in either 2D and 3D. It can draw using lines, points, boxes, contours, ... Gnuplot supports many different types of output: interactive screen terminals ( with mouse and hotkey ... Gnuplot is easily extensible to include new output modes. Recent additions include interactive terminals ... Development version is gnuplot 4.3 (SourceForge CVS). New features are being added regularly. You are welcome to build gnuplot from the CVS source code or download a testing binary. ... www.gnuplot.info/ - 6k - Cached - Similar pages - Note this

Download new in version 4.2

Documentation How to get CVS source and apply ...

FAQ Gnuplot-en

More results from gnuplot.info »

# gnuplot tips (not so Frequently Asked Questions)

Plotting Numerical Data in a Data File · What is the format which **gnuplot** can recognize? 2-dimensional data · 3-dimensional data · matrix · How do I plot several data sets in a single file? I want to modify values in my data file when plotting. I want to put some plotting commands in a data file. I want to eliminate some data points. How do I plot a part of data in a file? How do I use UNIX commands inside **gnuplot**? sort · paste · sed · awk · Date/Time Data · plotting time-sequence data · to change output text · After Plotting · I want to put a figure in my TeX document. I want to merge several figures into one figure . ...

t16web.lanl.gov/Kawano/gnuplot/index-e.html - 16k - Cached - Similar pages - Note this

Les liens S, M et L permettant respectivement:

- de supprimer la description de chaque résultat ("Small"),
- de laisser l'affichage par défaut ("Medium")
- d'allonger la taille de la description ("Large")

Dans le cas des descriptions étendues, Google semble continuer d'utiliser un mélange entre la balise meta description, si elle est définie, et le texte trouvé sur la page, s'il existe et s'il est pertinent par rapport à la requête. On peut s'inquiéter de ce snipper "géant", car l'information cherchée par l'internaute s'y retrouvait directement, ne rendant plus nécessaire la visite du site et faisant donc chuter le trafic issu de Google...!

<meta name="distribution" content="global|local|iu" />;

A qui s'adresse le contenu de cette page ? ("iu", pour "intranet").

```
<meta name="generator" content="Notepad++">; <meta name="formatter" content="Notepad++" />;
```

Le logiciel utilisé pour créer la page.

```
<meta name="identifier-url" content="http://www.exemple.com" />;
```

Cette balise est censée fournir aux moteurs de recherche l'adresse URL complète du site. Elle doit figurer uniquement sur la page d'accueil et ne peut contenir qu'une seule adresse, celle correspondant à l'entrée du site.

```
<meta name="keywords" content="liste de mots et expressions-clés, mot, mot, mot" />;
```

La balise Meta Keywords fait l'objet de nombreux débats dans la sphère des référenceurs, sert-elle à quelque chose ou est-elle tout simplement ignorée par les moteurs de recherches suite à la surexploitation de celles-ci pour faire du spam-indexing ? Elle est en tous cas totalement ignorée sur Google.

Si vous la saisissez, limitez-vous à 10/15 expressions-clés, séparées par des virgules

```
<meta name="publisher" content="David" />;
```

Qui publie le site?

```
<meta name="reply-to" content="webmaster@exemple.com" />;
```

L'email du webmaster responsable de cette page web.

```
<meta name="revisit-after" content="30 days" />;
```

Délai minimum souhaité entre 2 visites des bots. Cela ne change rien sur la véritable mobilisation des bots.

```
<meta name="robots" content="all" />;
```

Equivalent du fichier robots.txt

Par défaut, si la balise n'existe pas, les robots travailleront comme si elle comportait la valeur "all".

index;

Cette page sera indexée

noindex;

Cette page ne sera pas indexée.

Souvent utilisé pour les pages de connexion, de mentions légales, etc.

#### follow;

Les robots suivront les liens hypertextes pour indexer les autres pages

#### nofollow

Les robots ne suivront pas les liens hypertextes pour indexer les autres pages.

Attention, cela n'empêche toutefois pas une indexation dans Google!

#### none;

= noindex, nofollow soit l'inverse de all (=index, follow)

Souvent, on peut être confronté à une page peu intéressante à indexer, mais qui contient des liens vers des pages qu'il faut indexer; ce cas peut se concrétiser dans le cas des blogs, avec une page qui liste des catégories d'articles. Dans ce cas, pour ne pas référencer un contenu pertinent qui risque de défavoriser son référencement, il vaut mieux choisir "noindex, follow".

#### noimageindex;

Interdiction d'indexation des images.

#### noarchive;

Ne met aucun élément indexé en cache dans les serveurs du moteur.

#### nosnippet;

Le moteur n'affichera pas de description sous les titres indexés, sur les SERP.

```
<meta name="googlebot" content="noarchive" />;
```

Spécifique à Google. Et fonctionne comme "robots".

#### **2.3.2.2** les meta http-equiv

Les balises meta se composent en meta "NAME" comme nous venons de le voir, mais également en meta "HTTP-EQUIV", mieux interprétées par les bots:

```
<meta http-equiv="expires" content="Mon, 27 Sep 2007 14:00:00 GMT">;
```

Définit la validité de la page, date au format US.

```
<meta http-equiv="pragma" content="no-cache">;
```

Pour éviter une mise en cache chez l'internaute, en cas de mise à jour régulière du contenu par exemple.

```
<meta http-equiv="window-target" content=" top">;
```

A utiliser si votre utilise des frames. Rappelons que les frames, et les iframes, sont à proscire, car elles ne respectent par la loi du 1 URL = 1 PAGE.

```
<meta http-equiv="set-cookie" content="name; path=/; expires=Thursday, 20-May-07 00:15:00 GMT">;
```

Pour placer des cookies chez l'internaute.

```
<meta http-equiv="PICS-label" content="(pics-1.1 "http://images-webpages.com/VWP1.0/" l gen true comment Tol 0 ))>;
```

Ce tag permet de donner aux visuels de votre site une valeur. Développé par le World Wide Web Consortium (W3C), ce standard est devenu la référence pour la sélection du contenu Internet (PICS "Platform for Internet Content Selection"). Il permet d'étiqueter le contenu (évaluation du site, respect de la vie privée, droits de la propriété intellectuelle, etc ...) d'un site de deux façons:

- Soit les étiquettes seront stockées sur le serveur web du service qui a évalué le contenu de votre site.
- Soit vous indiquez vous même (à l'aide de balise meta pics-label) votre étiquette par l'intermédiaire d'un générateur de meta-tags.

Plus d'informations sur ce tag sur http://www.w3.org/PICS

```
<meta http-equiv="refresh" content="30, url=http://www.google.fr">
```

Attention, danger de blacklistage si cette page est réferencée par les bots !

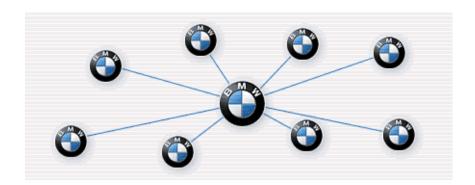
Comme vu dans la partie "URL-rewriting", on ne doit utiliser que des **redirections permanentes**, et aucune redirection invisible ou temporaire. Cette meta doit être utilisée par exemple pour rediriger un ancien site vers un nouveau site.

Toutes les redirections temporaires, en javascript, en php, ou quelle que soit la méthode d'écriture sont **PROSCRITES**. La méthode .htaccess, dans le cas d'Apache par exemple, est la meilleure solution pour placer une redirection.

#### Le cas des pages satellites

Une page satellite (="doorway page") est une page web qui était destinée à améliorer la place d'un site donné sur les moteurs de recherche en proposant de nombreux liens vers le site en question, associées à des combinaisons de mots clés conçues pour obtenir un score élevé, lorsqu'elles sont évaluées par les algorithmes des moteurs. Le visiteur qui atterrit sur une telle page se verra le plus souvent redirigé automatiquement. Ces pages étaient composées de texte souvent incohérent, avec une forte concentration des mots-clés, 1 page étant dédiée à chaque mot clé choisi.

Cette technique est depuis quelques années reconnues comme étant **abusive**. L'exclusion pure et simple de la base de données du moteur a fait de gros dégâts dans le monde du référencement dans les années 2004, 2005. Le plus gros scandale connu est celui de la marque BMW, dont le site a été banni de la base de données du moteur Google pour l'utilisation de ce procédé.



#### **2.3.2.3** Le cas du moteur Wordpress

Sur le moteur de blog Wordpress, il existe des plugins qui vous permettent de gérer les métaéléments de façon avancée, par exemple "All-In-One SEO Pack". Ce plugin permet de générer automatiquement des Meta tags uniques en utilisant, entre autres, les titres et tags de vos articles.

Par défaut, Wordpress diffuse des titres de pages sous la forme TITRE DE L'ARTICLE | NOM DU BLOG. All in One SEO Pack permet d'inverser cet ordre.

Egalement, ce plugin permet d'activer l'option qui permet d'utiliser "no index" sur les pages "catégories", "archives" et les pages de "tags". Cela vous permettra d'éviter tout risque de Duplicate Content. Pour ceux qui voudrait un plugin avec encore plus de possibilités, il est bon de jeter un oeil du coté de **wpSEO**.

D'autres CMS proposent de gérer de façon native des méta uniques, c'est le cas notamment de Joomla!.

#### **2.3.3 L'attribut "nofollow"**

Il arrive souvent que dans une page, on fasse des liens vers des sites externes. Le problème est que chacun de ces liens fait "fuir" une partie du PageRank de la page.

Il est donc conseillé de mettre l'attribut rel="nofollow" sur les liens n'ayant pas une grande importance, afin de limiter cette perte:

```
<a href="http://www.google.fr" rel="nofollow">lien externe</a>
```

Utilisez-les néanmoins avec modération:

- Activez l'attribut "nofollow" vers des domaines n'ayant pas de PR.
- Activez l'attribut "nofollow" vers des domaines trop loin de la thématique de votre site,

ex: si vous utilisez Feedburner, agrégateur RSS, n'oubliez pas de mettre un "nofollow" sur le lien de votre flux, vous garderez ainsi un peu plus de votre « google juice ».

```
<a href="http://feeds.feedburner.com/Pressecitron" rel="nofollow">RSS</a>
```

- Pour mieux orienter le robot, par exemple, si une page de tag doit bien être crawlée (pas de rel nofollow, pas de blocage par robots.txt), elle ne devrait pas forcément être indexée (robots:follow, noindex) Et inversement.
- Si vous travaillez sur un blog:

- ☐ réduisez le nombre de liens sortants dans votre "blogroll",
- □ activez cet attribut sur les commentaires laissés par les internautes.
- □ utiliser le sur des liens non pertinents: infopublicité, site de positionnement, concours, etc.

Attention néanmoins, cela n'empêche toutefois pas une indexation dans Google!

#### **2.3.3.1 Déterminer ses "Juicy Links"**

Pour ne vous tromper dans l'édition de vos "follow" ou "nofollow", n'hésitez pas à utiliser un outils qui va auditer les liens de votre site, et déterminer les liens qui vous donnent du PR, et ceux qui vous en font perdre, par exemple: Juicy Link Finder SEO Tool, disponible ici: <a href="http://www.seomoz.org/link-finder">http://www.seomoz.org/link-finder</a>

```
a[rel~="nofollow"] {
  border: thin dashed red! important;
  background-color: #ffc ! important;
}
```

#### **▼ 2.3.4 Les tableaux**

De moins en moins utilisés, notamment depuis l'avènement du XHTML, les tableaux sont devenus trop rigides et ne permettent pas de s'adapter à différents formats d'écran, de façon aussi dynamique que des blocs <div>. Toutefois, ils restent faciles à manier, et sont encore très utilisés.

Pour les optimiser, on oublie souvent que les tableaux contiennent des balisages spécifiques qui ne se limitent pas aux récurrents et .

Voici un exemple de tableau avec des titres de colonnes notamment, qui permettent ainsi de mieux hiérarchiser le contenu:

```
Titre colonne 2
   Titre colonne 3
  <t.r>
   Titre rangée n°1
   contenu B
   contenu C
  <!-- début footer -->
 <tfoot>
  Titre colonne 1
   Titre colonne 2
   Titre colonne 3
  </tfoot>
 <!-- fin footer -->
```

# Titre du tableau

Titre colonne 1	Titre colonne 2	Titre colonne 3
contenu A	contenu B	contenu C
Titre rangée nº1	contenu E	contenu F
Titre colonne 1	Titre colonne 2	Titre colonne 3

# **▼ 2.3.5** La structure des contenus et des pages

Nous avons vu que l'ordre des mots avait un impact sur le référencement. Les premiers mots, et les premiers paragraphes seront les mieux analysés.

Dans un paragraphe de texte, un bot va également cibler plus facilement un mot sur lequel un style particulier sera appliqué.

Autrement dit, les balises HTML qui définissent une mise en valeur, ou qui désignent précisément un type de contenu, seront prioritaires dans l'analyse.

Voici un aperçu global de certaines de ces balises:

# <strong>mot</strong>

Permet de mettre un mot en gras. L'avènement du XHTML insiste sur cette utilisation, à défaut du <b>mot</b>, désuet.

La balise <strong> sert véritablement à marquer un texte sur lequel on veut insister.

#### <u>...<u>

Le soulignement est aussi une marque en mise en valeur d'un mot ou expression-clé.

#### <i></i>, <em></em>

Pour une mise en italique.

### 

les listes sont très intéressants pour la conception des menus"

etc.

D'autres balises très utilisées ne fournissent aucune information sémantique, c'est le cas de:

#### <div></div>

Pour un encadrement d'un bloc de texte, préférez .

Même si vous utilisez un style CSS pour agrandir un texte, ou obtenir un équivalent graphique avec une autre balise forte, cela ne changera pas rien en terme de référencement.

# <span></span>

Même chose, le <span> n'a aucun poids.

Si vous diffusez un back-office de gestion, qui permet à votre client de créer lui-même le contenu de ses pages, il vaut mieux dans ce cas qu'il utilise **un WYSIWYG intuitif**, et surtout conforme avec les attentes des moteurs. Un outil gratuit répond à ce besoin, **BB Composer**: http://bbcomposer.elitwork.com/

### **2.3.5.1** Les headings

Encore plus forts que les autres balises fortes HTML, les headings créent un avantage très important sur le poids du référencement de votre contenu.

Les headings sont les balises <h1> ... <h6> qui permettent de dresser une hiérarchie de votre contenu.

Quelques précautions:

- 1 seul <h1> par page,
- ne pas sauter les étapes, et utiliser un <h4> sans avoir écrit un <h3> au préalable.
- entre 2 et 5 mots maximum par heading

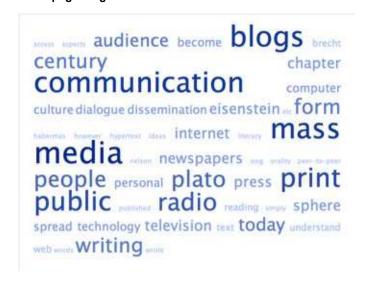
Comme vu dans la création du <titre> de la page, essayez d'utiliser des mots différents entre le titre et le h1.

Pour tester les headings de votre page, n'hésitez pas à télécharger le navigateur Lynx http://csant.info/lynx.

Ou utiliser la Web Accessibility Toolbar: http://www.visionaustralia.org.au/info.aspx?page=614.

#### ▼ 2.3.5.2 Les nuages de mots clés, ou "tag-clouds"

L'omniprésence des nuages de tags n'est plus à démontrer à l'heure actuelle. Nous en mangeons à toutes les sauces: sur les blogs, sur les sites, les réseaux sociaux, etc ... ils sont partout, parfois ajoutés par simple réflexe "2.0". Certaines études de type **eye-tracking** démontrent que leur présence n'améliore rien en ce qui concerne le trafic "humain" ; d'un point de vue robot par contre, c'est une belle section à posséder pour l'aider à crawler plusieurs dizaines de pages régulièrement.



## **2.3.5.3** L'auto-linking

L'auto-linking est un système qui permet de créer directement dans un texte un lien vers une page précise sur un mot précis, sans pour cela devoir créer le lien manuellement à chaque fois.

Concrètement, si votre blog, par exemple, propose une catégorie "cuisine", vous allez pouvoir décider qu'à chaque apparition de ce mot dans vos billets, un lien soit créé automatiquement vers les archives de cette même catégorie. Il est bien sûr possible de renvoyer vers ce que vous voulez: une page statique, les archives d'un mois précis, les autres billets d'une même catégorie, ceux ayant un tag précis (on parlera dans ce cas d'auto-tags), etc.

Divers outils peuvent très facilement vous aider sur ce point, c'est le cas avec le récent plugin **Keywords Autolink** sur Wordpress. Attention toutefois à ne pas en abuser. Comme toujours, pensez que le contenu est avant tout proposé à des êtres humains, pour qui des liens tous les 4 mots, ou vers des résultats peu pertinents, ne seront pas intéressants.

# **▼ 2.3.5.4 Les liens hypertextes <a href="#">**

L'intérêt d'un tag-cloud repose sur les headings utilisées pour déterminer la force des mots-clés, mais aussi sur le fait que ces headings sont accompagnés de liens hypertextes.

Le lien <a></a> est un autre point-clé d'une force du réferencement. Nous avons déjà vu comment l'utiliser à bon escient, pour ce qui est de son attribut "nofollow". Le bon usage des liens et de l'attribut "follow" dans l'HTML sert à offrir plusieurs façon d'arriver sur la page finale.

Les liens créent des "bonus", car s'il y a lien, c'est qu'il y a intérêt de contenu. Lorsque ces liens sont placés entre différents sites, on appelle cela le Net-linking.

Les liens sont d'autant plus fort s'il sont accompagnés d'attributs <title>, pour augmenter davantage le poids sémantique du lien:

<a href="http://www.google.fr" title="D'autres mots clés importants">mots clés importants</a>

Au sein d'un site, le maillage interne définit également un système hiérarchique entre les pages. La multiplicité de liens va engendrer une force de contenu, pour chaque page. Par exemple, sur un blog, composer une page regroupant les statistiques de vos billets les plus vus, les plus commentés, les plus populaires, concevoir un pied de page dans lequel vous avez accès aux X derniers articles publiés, ou rédiger un article "best-of" listant une série de liens vers d'autres pages, tout cela va entrainer une toile de liens très intéressante.

Note Evitez de mettre le lien en début d'article, et qu'il s'affiche avant la césure, pour que ce lien n'apparaisse pas sur la page d'accueil.

#### **2.3.5.5** Les liens relatifs

Ce type de liens que l'on rencontre souvent à la fin d'un billet, sur les blogs, rend 2 services: non seulement ils permettent de **proposer au lecteur d'élargir ses lectures autour d'un sujet particulier** (les choix sont souvent basés sur des mots-clés commun) mais ils **aident également les moteurs à passer de billet en billet** et ainsi de continuer à prendre en considération des articles datant parfois de très longtemps.

Dans le domaine de l'ecommerce, on retrouve un peu ce schéma, grâce au "cross-selling" (="vente croisée"). Lorsque vous êtes sur un fiche produit, vous verrez un lien vers un produit complémentaire: "Nos clients qui ont déjà acheté cette casserole, achètent régulièrement tel ou tel produit".



#### **2.3.5.6** Le fil d'ariane, ou "breadcrumb"

Le fil d'Ariane est une solution esthétique qui répond aux besoins des visiteurs humains et des bots, pour créer une hiérarchie de contenu entre les pages, et en terme de référencement. On obtient ainsi un lien optimisé vers la page principale depuis toutes les pages du site et un lien vers les pages parentes.



# **2.3.5.7** Composer un "plan du site"

Le plan du site, ou sitemap, à ne pas confondre avec le sitemap xml (dans la suite du cours), est une autre solution qui répond au besoin pour créer une hiérarchie de contenu entre les pages. Mettre un lien sur toutes les pages du site vers cette page, fait reconnaître aux bots cette page comme éant le "sommaire" du site.

Cette solution est toutefois efficace dans le cas d'un site "corporate". Lorsqu'il s'agit d'un site à très fort contenu, comme un blog (qui contient des billets hiérarchisés par date) ou un site e-commerce (qui contient des produits hiérarchisés par arborescence catalogue), il vaut mieux travailler sur des pages dédiées, par catégorie par exemple, pour ne pas fournir des pages à trop fort contenu, et ainsi perdre en intensité sémantique. Dans ce cas, il serait de faire une page "plan des catégories et des sous-catégories", en hiérarchisant le contenu par headings et listes ordonnées, et des pages connexes, dédiées à chaque branche de l'arborescence. On parle alors plutôt de "Toile du site".

N'oubliez pas que cette page, et toutes les pages composant cette toile, doivent avoir un lien vers la page d'accueil.

# OutilsRéférencement Outils dédiés au référencement de sites Web Vous êtes ici : Outils référencement > Plan du site Plan du site - Outils Référencement Outils • Infos · Politique d'accessibilité Outils Technique o Contacter Outils Référencement o Plan du site Entête HTTP Lexique Balises HTML Analyse de pages web Rendu moteur o Outils Contenu Générateur de balises Meta Nambre de mots clés Densité de mots clés Similarité de contenu Prominence o Outils Popularité Indexation de pages ■ Backs Links PageRank Positionnement de mots clés o Outils Choix de Mots clés Suggestion de mots clés · Croisement de mots clés Affiner ses mots clés

### Le cas de Kelkoo

Kelkoo est le meilleur exemple possible d'intégration de ce schéma sitemap. Chaque produit est classé et accessible depuis des pages reproduisant l'arborescence de son catalogue, selon

o Outils Sitemaps

diférents critères: type de produit, marque, marchands, etc. Le tout accessible dès l'accueil et de façon ergonomique, avec en plus un système ordonné de breadcrumb. Découvrez cette bonne idée ici: kelkoo.com



## **2.3.5.8** Diffuser un glossaire

Un peu dans l'esprit d'un plan de site, le glossaire répond cete fois à 2 critères:

- Optimiser le réferencement naturel d'un site, en créant du contenu.
- Car dans chaque page, ou article, vous aurez quelques mots-clés, avec un lien vers la définition de ce terme dans le glossaire. Une page globale avec tous les mots-clés mais aussi une page dédiée à chaque mot-clé optimisera efficacement votre référencement.
- S'adresser à un auditoire moins expert dans son domaine, et ne pas perdre de trafic (au début du cours, nous avons vu Alexa, dont les résultats sont pris en compte pour le calcul du SERP de Google!)

Prenons le cas d'un blog. La mise en place d'un glossaire sur un blog pourrait paraître étrange car ce n'est pas vraiment commun ou habituel sur la **blogosphère**. Néanmoins, on réalise rapidement que cela complète très bien une stratégie d'ouverture de son blog. Il faut faire attention à ne pas perdre des visiteurs et donc des lecteurs potentiels quand on écrit un blog sur un thème précis, et **nécessitant un minimum de connaissances** pour le comprendre. Avec un glossaire regroupant l'ensemble des mots-clés de base à connaître, et leurs définitions, un utilisateur débutant, ou souhaitant en savoir plus, pourra dès lors ne pas rester sans comprendre ou – mieux encore – ne pas quitter votre site pour Wikipedia à la recherche d'une définition et quitter le site sans y revenir.

Sur Wordpress, il existe pour cela le plug-in **IMM-Glossary**, qui va créer automatiquement une page "Glossaire". Vous pouvez y paramétrer l'affichage de votre contenu: sous forme d'une pop-up, d'une info-bulle ou d'une page dédiée (c'est mieux pour le référencment), accompagné d'une icône ou pas, mot-clé lié pour chaque occurrence dans un article ou seulement à la première, etc.

#### **2.3.5.9** Les "Top recherches"

Pour aller plus loin encore que le principe des nuages de mots clés, composés notamment grâce aux tags attribués aux billets de blogs, le "top recherche" est plus dynamique, et se base sur la mémorisation des recherches effectuées par les internautes, et sur les résultats d'un moteur de recherche interne d'un site pour créer automatiquement des **pages de résultats optimisées et référençables**.

Cette technique est très importante sur les sites à gros contenu et à gros trafic. Plusieurs milliers de requêtes peuvent être tapées sur les site marchands par exemple. Et il s'agit d'enregistrer les comportements similaires des internautes, pour leur fournir les résultats les plus pertinents, le plus rapidement possible, en **réduisant le nombre de clics**. C'est un peu comme reproduire un SERP, au sein de son site.

Il faut donc déterminer une arborescence dans le contenu, en profitant du balisage HTML cité auparavant, avec un contenu le plus dense possible. Concrêtement, il faut créer une page d'entrée "index", qui ensuite classera toutes les recherches des internautes suivant un algorithme optimal de classification, basé sur les mathématiques et les arbres complets à N niveaux.

Par exemple, si nous avons un total de 1.000.000 de recherches distinctes, il faut classer ces 1.000.000 de recherches en 100 lots de niveau 1. Ces 100 lots de niveau 1 seront accessibles par la page d'accueil du top recherche indexée par ordre alphabétique. Ensuite, chaque lot de 10 000 recherches sera ensuite découpé en 100 lots de 100 recherches, triés par ordre arbitraire. Voilà

comment on peut classifier 1.000.000 de recherches en 10.000 lots de niveau 2 eux même classifiés en 100 lots de niveau 1. **N doit être choisi en fonction de la taille maximale estimée du top recherche**, à raison de 100 liens par page (si nombre de recherches = 1.000.000, alors n=3, car 100^3 = 1.000.000).

Cette technique est actuellement utilisée par un gros site sportif français. La hausse de trafic apportée par cette technique est estimée à environ **200.000 visites de plus par mois**, grâce à des dizaines de milliers de pages référencées en plus dans Google.

### **2.3.5.10** Les erreurs à ne pas commettre

Nous avons vu que le code HTML a une importance sur le contenu. Plus il y aura de contenu dans votre page, plus important sera la hiérarchisation à pratiquer sur ce contenu.

Toutefois, ne vous laissez par emporter par des excès de techniques "exotiques", dont voici quelques exemples des méthodes les plus répandues, et douteuses:

#### saisir un texte blanc sur un fond blanc

C'est considéré comme une technique de spam-indexing, et vous risquez le black-listage. Et guelle que soit la méthode d'encodage (hml, css, etc.)!

#### stocker du contenu-clé en bas de page

C'est inutile, et risque même de diluer le contenu du reste de la page.

# écrire un style "display: none;" sur un contenu

C'est considéré également comme une technique de spam-indexing, et vous risquez le black-listage. De nombreux sites se font faits blacklistés simplement pour cette raison!

### les pages satellites

Nous avons déjà vu que cela créait un contenu incohérent, et surtout des redirections non-permanentes rédhibitoires.

#### font: 0

Google ne lira pas le contenu, et risque également de vous blacklister!

Pour en savoir plus: http://www.google.com/support/webmasters/bin/answer.py?answer=66353

#### **₹ 2.3.5.11 Les éléments inutiles**

Pour compléter la partie précédente, voici 3 fausses idées, qui n'amélioreront pas votre référencement!

#### inscrire des commentaires <!-- --> avec des mots clés dedans

Inutile, et risque même d'alourdir la page...!

## dupliquer des liens hypertextes dans la même page

Inutile, ne va pas augmenter le PR de la page concernée

### écrire un lien vers la page courante

Inutile!

# **▼ 2.4 Le fichier sitemap**

Un sitemap est un fichier XML très simple, avec une structure stricte pour organiser les données de manière homogène. Il décrit l'ensemble des pages d'un site en ajoutant quelques informations:

- Date de dernière modification
- Priorité
- Fréquence de mise à jour

Il doit décrire tout le contenu du site afin de faciliter le réferencement par le bot. Il n'est pas obligatoire de le mettre à la racine du site, mais dans ce cas il faut créer un sitemap "index". Par contre, ne seront parcourues par les robots que les url des fichiers contenues dans le dossier du fichier sitemap.

Dans le cas d'un blog, on trouvera donc la liste des URL des articles, la liste des URL des catégories, des archives, etc.

Ce fichier est très important pour le référencement, car vous pouvez ainsi indiquer à Google et aux autres moteurs de recherche l'ensemble des pages à référencer, au lieu de leur laisser la peine d'essayer de toutes les trouver tous seuls.

Limites techniques:

- 50 000 URL maximum
- le fichier ne doit pas excéder 10Mo

Note Si le fichier est trop gros, on peut le compresser: http://www.site.com/sitemap.gz ou http://www.site.com/sitemap.gz

Pour soumettre un fichier sitemap sur Google, il faut un compte Google (ou Gmail par exemple). Google propose le "Google Sitemap" au sein du GWT "Google Webmaster Tool", un outil très intéressant qui a pour but de recenser toutes les URL de votre site et de suivre en temps réel leur réferencement. Yahoo diffuse également son YSE "Yahoo site explorer".

Voici un exemple de fichier sitemap:

# Quelques explications:

#### <url>...</url>

Il faudra dupliquer pour chaque page de votre site le bloc <url> et y spécifier le contenu adéquat. 2048 caractères maximum. Les URL doivent respecter l'**encodage UTF-8**, et pout cela il faut convertir les caractères spéciaux des URL, dans la même norme qu'en HTML:

- & devient & amp;
- ' devient '
- " devient "
- > devient >
- < devient &lt:</p>

#### <|astmod>...</|astmod>

représente la date de dernière modification de l'URL (format de date validé par le W3C: YYYY-MM-DD). L'heure est optionnelle. Pour plus d'infos sur ce format: <a href="http://www.w3.org/TR/NOTE-datetime">http://www.w3.org/TR/NOTE-datetime</a>

# <changefreq>...</changefreq>

fréquence prévisible de mise à jour du site:

- always, à utiliser pour les pages qui changent à chaque affichage (une page "top recherche" par exemple)
- hourly
- daily
- weekly
- monthly
- vearly
- never, à utiliser pour des pages "archivées"

Notez bien que ces informations ne modifieront pas grand chose dans les crawls des bots. ce ne sont pas des commandes, mais des indices.

## <priority>...</priority>

priorité de l'URL, si plusieurs URL (de 0.1 à 1.0, par défaut 0.5). Cette information ne sera prise en compte par les moteurs que pour des pages qu'ils estiment de "même niveau", et les aidera à départager les contenus indexés. Cela ne changera en rien que la page d'accueil demeurera par exemple la page n°1.

Lorsque vous avez de nombreuses URL à saisir, cela devient vite fastidieux. Heureusement, au sein des CMS open-source, il existe des plug-ins qui permettent d'automatiser leur création; Wordpress jouit du **Google (XML) Sitemaps Generator for WordPress**, Gsitemap pour DotClear1, Sitemaps pour Dotclear2, quant à Joomla!, il utilise le module **Joomap**; Pour Blogger ou d'autres plateformes de blog auto-hébergées, le sitemap est en général automatique et vous n'avez donc à vous en occuper.

Pour obtenir de la documentation précise sur les sitemap, rendez-vous à http://www.sitemaps.org/.

# **▼ 2.4.1 Plusieurs sitemap sur son site**

Si vous avez de nombreux fichiers sitemap au sein de votre site, il est possible de créer un sitemap index, et y lister vos fichiers sitemap:

## Quelques explications:

#### <sitemap>...</sitemap>

Il faudra dupliquer pour chaque sitemap de votre site le bloc <sitemap> et y spécifier le contenu adéquat.

#### <loc>...</loc>

Cela peut être un fichier sitemap, un flux RSS ou Atom, ou un simple fichier texte.

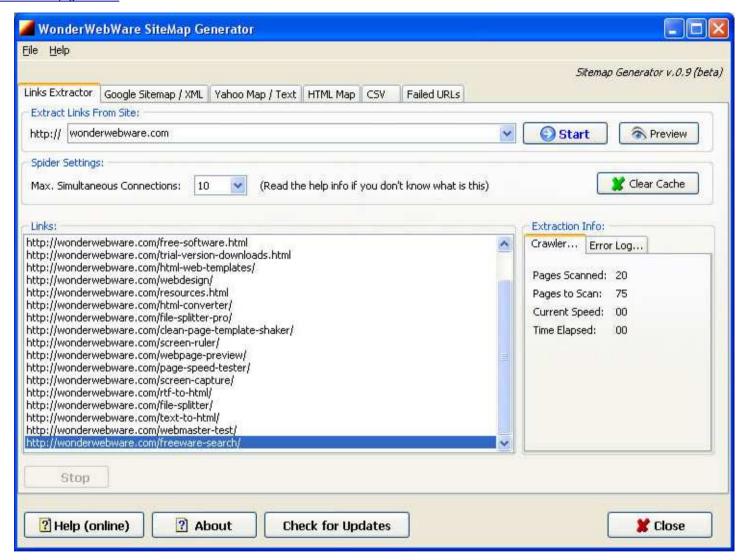
### <lastmod>...</lastmod>

En définissant ainsi des dates de mise à jour, par tranches de votre site, vous permettez aux bots de définir de véritables hiérarchies de contenu au sein de votre site.

## **▼ 2.4.2 Générer automatiquement les sitemap**

Pour les sites aux nombreuses pages, parfois dynamiques, il est vrai que lister l'ensemble des pages manuellement s'avérera très vite fastidieux.

N'hésitez pas à utiliser des Sitemap Generator, comme WWW Sitemap generator, qui va vous permettre de générer des sitemap.xml de différents formats:



# **▽ 2.5 Le fichier robots.txt**

Le fichier robots.txt est un petit fichier texte qui va donner des indications aux bots sur ce qu'il peuvent parcourir et ne peuvent pas parcourir sur le site.

A savoir:

• il ne peut exister qu'un seul fichier robots.txt sur un site,

- il doit absolument se trouver à la racine du site.
- le nom du fichier doit toujours être écrit en minuscules,
- si le fichier n'existe pas, le bot va "aspirer" la totalité du site, considérant que rien ne lui est interdit.

Par exemple, il peut être intéressant de bloquer le parcours des archives des publications par mois, sur un blog, pour éviter un problème de **Duplicate Content**.

Voici un exemple de contenu, qui va permettre à tous les bots de parcourir tout le contenu du site:

User-Agent: \*
Disallow:

Il est conseillé d'utiliser un éditeur de texte tel que le Bloc-notes pour le saisir. Des programmes comme Word sont utilisables, mais avec précaution, car le fichier sauvé sur le serveur peut parfois ne pas être au format ASCII pur.

### **▼ 2.5.1 Quelques exemples de structure**

Voici quelques exemples de structure d'un fichier robots.txt:

# User-agent: \*

signifie que l'accès est autorisé à tous les bots

# User-agent: Googlebot, Scooter

signifie que l'accès est autorisé à qqs bots uniquement (Googlebot, le robot de Google, Scooter, celui d'Altavista). Le site ne sera donc pas réferencé sur Live Search, Yahoo, etc.

#### Disallow:

le robot n'exclura aucune page du serveur, comme si le fichier était totalement vide

#### Disallow: /

le robot exclura toutes les pages du serveur

# Disallow: /cgi-bin/

le robot n'ira pas explorer le dossier cgi-bin

#### Disallow: /dossierexemple/test.html

le robot n'ira pas explorer le fichier test.html

Bien entendu, on peu cumuler ces informations,

dans le cas présent, on indique que le bot "fast" de Altheweb est le seul autorisé à indéxer le site, et que google doit tout parcourir, sauf 3 fichiers html

User-Agent: fast
Disallow:

```
User-Agent: googlebot
Disallow: /repertoire/chemin/page.html
Disallow: /repertoire/chemin/page2.html
Disallow: /repertoire/chemin/page3.html
User-Agent: *
Disallow: /
```

# Quelques noms de bots connus:

#### Altavista

Scooter

#### **Excite**

ArchitextSpider

# Google

Googlebot

Googlebot-Image, etc.

## HotBot

Slurp

## InfoSeek

InfoSeek Sidewinder

## Lycos

T-Rex

## Voila

Echo

# A retenir:

- le fichier ne doit contenir aucune ligne vièrge,
- l'étoile (\*) n'est acceptée que dans le champ User-agent,
- # permet d'écrire un commentaire.

# **▼ 2.5.2 Les cas particuliers de Google**

Google permet quelques cas particuliers, pour le "User-Agent: Googlebot":

```
User-Agent: Googlebot-Image
Disallow: /dossier/
Disallow: /*?
Disallow: *.gif$
Disallow: /*.jpg$
Allow: /annexes/images/tatouage-moto.jpg
Allow: /annexes/images/24h-mans-moto.jpg
```

# Explications:

- Google permet d'éviter aux URL contenant un point d'intérrogation ? d'être référencées.
- Google accepte l'indexation de fichier d'un type particulier, dans notre exemple on lui interdit de réferencer les fichiers .gif et .jpg.
- Google accepte la permission "Allow", pour gérer des exceptions à des interdictions générales.

Voici un exemple extrait du robots.txt de <a href="http://www.peperuka.com/">http://www.peperuka.com/</a>, un blog conçu sur Wordpress ; on voit qu'un effort précis a été fait pour éviter de réferencer des "trackback" et autres doublons (cf. partie "sitemap").

```
User-agent: *
Disallow: /wp-content/uploads/
Disallow: /album/
# If bots like wildcard
Disallow: */feed/
Disallow: */trackback/
Disallow: /*.php$
Disallow: /*.htm$
Disallow: /*?
User-agent: Googlebot
Disallow: /test/
Disallow: /wp-*
Disallow: */feed/
Disallow: */trackback/
Disallow: /2008/
Disallow: /2008/*/
Allow: /2008/*/*/$
```

```
Disallow: /2007/
Disallow: /2007/*/
Allow: /2007/*/*
Disallow: /*.css$
Disallow: /*.tms
Disallow: /*?
Allow: /wp-content/photos/

User-agent: Googlebot-Image
Allow: /*

User-agent: Mediapartners-Google
Allow: /*

User-agent: ia_archiver
Disallow: /

User-agent: duggmirror
Disallow: /
```

# **2.5.3** La gestion du sitemap

Enfin, le fichier robots.txt peut également insérer le(s) lien(s) vers des fichiers "sitemap".

```
User-Agent: *
Disallow:
Sitemap: http://www.example.com/sitemap.xml
```

Cette information reste optionnelle, toutefois elle peut s'avérer intéressante lorsque vous gérez plusieurs sites, car cela vous permet de **centraliser** sur un même site tous les sitemaps de ces différents sites. Cela est utile lorsque vous travaillez avec un outil comme le GWT, ainsi vous gagnez du temps lors de la soumission de vos sites auprès de Google car plus besoin de procéder à diverses vérifications des sites.

En conclusion, si vous gérez les sites www.site1.com et www.site1.com et que vous souhaitez centraliser les sitemap sur www.site1.com, voici ce que vous écrirez dans le fichier robots.txt du site n°2:

```
User-Agent: *
Disallow:
Sitemap: http://www.sitel.com/sitemap-www-site2.xml
```

# **▼ 2.6 Les images**

Les images sont des éléments fondamentaux dans l'aspect graphique d'un site web.

Elles peuvent avoir 2 utilités:

- soit compléter le contenu d'un article ou d'un contenu particulier.
- soit venir dessiner des effets graphiques pour mettre en valeur le contenu.

Il semble que ces 2 cas doivent être traités différement lors de l'encodage HTML/CSS d'une page. En effet, les moteurs d'images sont très friands des images indexées au sein des balises <img>. Il convient donc de traiter les véritables images comme il se doit, par l'utilisation de <img>, plutôt qu'avec un "background" en CSS.

Concrêtement, la balise img permet d'ailleurs d'optimiser l'accessibilté d'un site, en diffusant un texte alternatif au sein de la balise:

```
<img src="image" alt="Ceci est un texte alternatif" />
```

Suivez les mêmes conseils que pour la partie <title> pour constuire ce contenu. Gardez bien en tête que les textes alternatifs sont très importants pour un référencement correct de votre contenu.

#### **2.6.1** Quelques astuces d'optimisation

Conseil n°1: utilisez des images esthétiques !

Voici quelques autres optimisations possible pour optimiser le référencement des images importantes affichées par les <img>:

# Intégrez chaque image dans une balise <div>

Ce n'est pas vraiment prouvé, mais la communauté web s'entend à dire que cela pourrait augmenter le taux de précision.

#### Faites précéder l'image d'un titre situé dans un heading de niveau 3 ou supérieur

Nous avons vu l'importance d'un heading. L'utiliser pour encadrer une image augmentera donc forcément le contenu qu'elle diffuse.

#### Utilisez des images qui occupent environ 35-40% maximum de la surface de la page web

Pour des raisons évidentes de poids, et donc de chargement de la page.

#### Mettez les images en haut des pages web

Dans un billet de blog par exemple, il est préférable de diffuser ces contenus en haut de la page, puisque les bots parcourent les pages de haut en bas.

# Mettez des mots clés dans le nom du fichier image

Par exemple, cours-referencement.jpg se réferencera mieux que test01.jpg

#### Ajoutez une légende

Profitez d'encadrer votre image d'une balise <div>, <a>, ou de headings, pour y greffer un texte, qui sera considéré comme une légende de l'image.

```
<div>
  <img src="image" alt="Ceci est un texte alternatif" /><br />
```

```
<strong>Ceci est une légende</strong>
</div>
```

Vous pouvez tout autant augmenter davantage le poids du texte grâce aux balises de forme.

### Mettez des mots-clés dans l'attribut ALT

Déjà, vu. Par contre, pas besoin d'utiliser les attributs "TITLE" et encore moins le "LONGDESC", ignoré.

### Faites des liens vers votre image depuis d'autres pages de votre site

Les liens internes ne font pas de mal!

## Pour l'instant, les informations EXIF sont ignorées mais elles pourraient très bien être utilisées plus tard

Il s'agit des éléments inscrits dans le fichier image. Qui peuvent être gérés par Photoshop par exemple. Tenez compte du fait que les moteurs améliorent leur algorithmes pour tenter d'identifier le contenu des photos (ça fonctionne déjà bien pour la reconnaissance de visages dans Google Images ou plus récemment dans les albums Picasa Web).

Tous ces points peuvent vous aider sensiblement.

Toutefois, une astuce testée, et qui ne fonctionne pas, faire un sitemap qui renvoie vers les sources .jpg:

# **▼ 2.6.2 Les images-map**

Le principe de l'image map, ou image "réactive": mettre plusieurs liens sur la même image, pour ne pas avoir à la redécouper par exemple, ce qui peut s'avérer utile pour des menus, ou des blocs d'image où il y a plusieurs liens. Et ainsi, éviter de référencer dans les moteurs dédiés des images inintéressantes.

```
</map>
<img src="image" alt="Ceci est un texte alternatif" usemap="#test" />
```

## **▼ 2.6.3 Les sprite CSS**

Autre idée, pour éviter de redécouper une image: gérer l'affichage d'une image par CSS, en affichant spécifiquement une zone particulière d'une image contenant des nombreux éléments picturaux. On affiche donc l'élément en question en "background", en créant une fenêtre et une vue spécifique sur l'élément qui nous intéresse. Par exemple:

```
/* l'image 500x10px contient de nombreuses petites "puces", mais on ne va afficher que l'une d'entre elles */
img {
    width: 10px;
    height: 10px;
    background: #fff url('image.jpg') left top no-repeat;
    /* il suffit de remplacer left et top par les coordonnées en pixels de la puce, dans l'image */
}
```

Des outils gratuits vous aident dans la création de sprites: spirtegen website-performance.org ou smartsprites.osinski.name

# **₹ 2.7 Les animations Flash**

Comme dans les cas des images, les bots ne peuvent pas lire le contenu d'une animation flash.

Toutefois, la meilleure solution, appellée Flash Satay, bien que limitée, est de stocker du contenu alternatif dans le balise <object>, et parfois <noembed> en HTML:

Attention à ne pas dupliquer le contenu entre ces zones alternatives. Préférez la seconde zone, la balise <noembed> étant très rarement utilisée. D'autant que l'inconvénient majeur de cette méthode est que la balise <embed> et <noembed> n'est pas conforme aux recommandations du W3C.

Avoir une intro flash ou **un site full-flash**, va compléxifier les optimisations possibles. Dans la plupart des cas, le manque de mise à jour notoire d'un site Flash sera, à force, un "malus" aux yeux des bots. Pourtant, Google a annoncé au 1er semestre 2008 que son robot arrivait désormais à lire un contenu flash... Affaire à suivre donc!

Il existe néanmoins quelques autres solutions, plus ou moins efficaces, pour réussir à indexer efficacement du flash:

- dupliquer le site Flash en une version HTML, parfois plus facile à dire qu'à faire,
- générer le contenu Flash au moyen d'un fichier XML, lequel pourra être crawlé. Généralement, ce fichier XML contiendra du texte hiérarchisé (arborescence produit, etc.),
- créer une liaison avec des ancres HTML:

le code actionscript suivant,

```
on (press) {
   getURL("page.htm#inscription","_self");
}
```

mènerait donc à une ancre nommée "inscription" sur la page page.htm:

```
<a name="inscription">ancre HTML</a>
```

• utiliser la méthode SWFObject en JavaScript. SWFObject est un petit script JavaScript utilisé pour inclure un contenu Flash dans une page HTML. Le script détecte le plug-in Flash dans tous les principaux navigateurs et est conçu pour rendre l'intégration des animations SWF aussi simple que possible. Il résout notamment le problème de double-clic rencontré parfois sur Internet Explorer. L'intégration dans la page est compatible avec les normes HTML et XHTML 1.0.

Auparavant, SWFObject s'appelait FlashObject, mais il a dû être renommé pour des raisons de copyright.

```
<script type="text/javascript" src="/swfobject.js"></script>
<div id="flashcontent">texte alternatif & mots clés</div>
<script type="text/javascript">
   var so = new SWFObject("fichier", "animation", "100", "7", "#fff");
   so.write("flashcontent");
</script>
```

La balise <div> contenant le texte alternatif doit toujours comporter l'attribut "flashcontent". Ce <div> s'affichera lorsque le plug-in Flash Player sera absent.

Note Vous pouvez télécharger SWFObject ici: http://code.google.com/p/swfobject/.

Une autre solution de deep-linking: SWFAdress <a href="http://www.asual.com/swfaddress/">http://www.asual.com/swfaddress/</a>

# **2.8 Les autres contenus multimedia**

#### **₹ 2.8.1 Les vidéos**

Depuis 2007, Googlebot reconnait certains formats de vidéos:

- .mpg
- mpeq
- .mp4

- mov
- wmv
- .asf
- .avi
- --
- .ra
- .ram
- .rm

Et pour optimiser leur crawl, vous pouvez créer un sitemap vidéo, ce qui vous permet d'optimiser les éléments alternatifs qui accompagnaient jusqu'alors les contenus vidéos.

Note Google Webmaster Tool est prévu pour accueillir des sitemap vidéo.

#### Quelques explications:

#### <url>...</url>

Il faudra dupliquer ce bloc pour chaque page de votre site contenant des vidéos.

#### <loc>...</loc>

L'adresse de la page hébergeant l(es) vidéo(s).

### <video:video>...</video:video>

Il faudra dupliquer ce bloc pour chaque vidéo de votre page.

#### <video:title>...</video:title>

Vous pouvez désormais donner un titre à votre vidéo. Respectez les prérogatives d'un <title> de page HTML.

#### <video:description>...</video:description>

Vous pouvez désormais donner une description à votre vidéo. Respectez les prérogatives d'un <meta> description de page HTML.

Pour créer facilement vos sitemap vidéo, rendez-vous ici: http://www.videositemappro.com/

N'oubliez pas que la meilleure méthode reste le dépôt de vos vidéos sur des hébergeurs dédiés: Youtube, Dailymotion, etc.

### **▼** 2.8.2 Les applets java

Les applets Java sont aussi compliqués à référencer que des animations flash. La seule option est de stocker un texte alternatif, au cas où l'utisateur n'utilise pas le Java Runtime:

```
<applet code="programme.class" width="100" height="100">+
    texte alternatif & mots clés
</applet>
```

# **2.9 Le "Duplicate Content"**

On appelle aujourd'hui "contenu dupliqué", ou **duplicate content**, un contenu identique qui peut être accessible via plusieurs URL. D'un point de vue des moteurs de recherches, les contenus dupliqués sur plusieurs URLs peuvent s'apparenter à des techniques frauduleuses de spam-indexing.

Un débat fait rage enre les référenceurs, avec d'une part des partisans expérimentés de la blogosphère du "il faut optimiser l'accès au plus petit nombre de page possible". Leur explication: Google n'a pas envie de perdre son temps à crawler des pages par mot-clés ou catégories qui retournent les mêmes articles. En faisant cela, Google va certes indexer et votre page aura plus de chance d'être en première page, mais:

- trop de pages à indexer peut s'avérer néfaste, tout dépend leur structure,
- les doublons sont considérés comme une technique de spam-indexing, appelée "Duplicate Content"

Conclusion, il est préférable de n'indexer que votre page d'accueil et vos pages par article (et surtout ne pas indexer les pages par article, mots clés, dates, auteur, etc.).

Avec trop de duplicate dans le contenu du site, celui ci ne sera pas totalement **crawlé**. Ce qui explique un peu mieux le problème de crawl des annuaires qui ont tous 2 fois au moins le même contenu. (page de resultats et fiche).

Si vous vous rendez compte qu'un contenu dupliqué a été indexé dans Google, vous pouvez le supprimer grâce au Google URL Removal, intégré dans la suite Google Webmaster Tool.

#### ■ 2.9.1 Indexer son site avec et sans "www"

Avez-vous testé ce qui se passe si on ne tape pas www dans votre URL?

C'est-à-dire qu'on accède à votre site en tapant <a href="http://example.com">http://example.com</a> au lieu de <a href="http://www.example.com">http://example.com</a>. Si votre site reste accessible sans que l'on soit redirigé vers la version officielle, celle qui contient www dans l'URL, alors votre site risque d'être indexé 2 fois par les moteurs, et vous aurez des problèmes de **contenus dupliqués**.

Le remède consiste à inscrire un code dans son .htaccess:

```
RewriteEngine On

RewriteCond %{HTTP_HOST} !^www\.example\.com [NC]

RewriteRule (.*) http://www.example.com/$1 [QSA,R=301,L]
```

Selon les cas, vous devrez peut-être retirer le / à la fin du domaine, comme ceci:

```
RewriteEngine On

RewriteCond %{HTTP_HOST} !^www\.example\.com [NC]

RewriteRule (.*) http://www.example.com$1 [QSA,R=301,L]
```

Au contraire, pour supprimer le sous-domaine www, et interdire l'indexation du site avec ce sous-domaine, mettez le code suivant en haut de votre fichier .htaccess situé à la racine du site:

```
RewriteEngine On
RewriteCond %{HTTP_HOST} !^example\.com [NC]
RewriteRule (.*) http://example.com/$1 [QSA,R=301,L]
```

Attention aux cas particuliers, notamment si vous avez des sous-domaines, autres que www. Remarquez que l'exemple inverse revient au même.

N'oubliez pas d'ailleurs que Google permet aux webmasters d'indiquer eux-mêmes quelle est la version officielle de leur site (avec ou sans www), ce qu'on appelle l'**URL canonique**. Il suffit d'aller dans son compte Google Webmaster Tool.

## **▼ 2.9.2 Avoir 2 URL pour sa page d'accueil**

Presque tous les sites sont conçus avec un lien sur chaque page pour retourner à la page d'accueil (en général ce lien est situé sur le logo du site). Avez-vous vérifié que ce lien pointe bien vers l'URL précise de votre nom de domaine et non pas autre chose ?

Exemple: la page d'accueil est index.php mais tous les liens pointent vers <a href="http://www.example.com/">http://www.example.com/</a> et non pas vers <a href="http://www.example.com/index.php">http://www.example.com/index.php</a>, sinon il y a encore un problème de contenu dupliqué, et notamment une dilution du PageRank et des autres effets liés aux backlinks. Si vous désirez optimiser votre site à 100%, il faut tester sur chaque page importante si l'URL demandée est bien l'URL officielle, et dans le cas contraire rediriger de façon permanente (=301) vers l'URL officielle.

Il est possible de régler cela par .htaccess, à condition d'y avoir accès. Sinon, en Php, il suffit de quelques lignes. Voici une fonction d'exemple à appeler sur chaque page du site:

```
function redirection_301_si_besoin($url_attendue)
{
   if ($_SERVER['REQUEST_URI'] != $url_attendue)
   {
     header("Status: 301 Moved Permanently", false, 301);
     header("Location: http://www.example.com".$url_attendue);
     exit;
```

## **2.9.3** Avoir 2 URL pour la page n°1

Imaginons que vous ayiez un forum et que, dans chaque discussion, vous listiez les messages en limitant l'affichage à **15 messages par page**. Vous avez également un système de pagination pour voir les messages des pages 2 et suivantes. Ces pages-là ont certainement dans **leur QueryString** un paramètre qui indique le n° de la page. Avez-vous vérifié que sur les pages 2 et suivantes, le lien vers la page 1 pointe bien vers la même URL que la page par défaut de la discussion ?

Par exemple:

- la page par défaut de la discussion est http://www.example.com/topic.php?t=456
- ◆ la page 2 de la discussion est http://www.example.com/topic.php?t=456&p=2
- le lien depuis la page 2 vers la page 1 doit pointer vers http://www.example.com/topic.php?t=456 et surtout pas vers http://www.example.com/topic.php?t=456&p=1

## **▼ 2.9.4 Avoir 2 types d'URL indexables (réécrites et classiques)**

Imaginons que vous veniez de mettre en place l'URL Rewriting sur votre forum. Pour reprendre l'exemple précédent, les URL de pages de discussion sont passées de <a href="http://www.example.com/topic.php?t=456">http://www.example.com/topic.php?t=456</a> à <a href="http://www.example.com/topic.php?t=456">http://www.example.com/topic.456.html</a>. Avez-vous vérifié que vous interdisiez l'indexation des pages avec l'ancien format d'URL ? La meilleure solution dans ce cas est même de rediriger de façon permanente (=301) chaque page à l'ancien format vers la page équivalente avec le nouveau format.

Si l'URL-rewriting est défini de manière à traduire plusieurs fois une même page, même si elle a un contenu légèrement différent (bannière de pub aléatoire par exemple), vous risquez de vous faire blacklister, car cette technique est considérée comme du spamdexing. L'objectif n'est pas indéxer plusieurs fois une même page, mais fournir à Google plusieurs moyens d'accéder à cette page.

#### **2.9.5** Avoir plusieurs noms de domaine indexés pour un même site

Sans doute avez-vous acheté plusieurs noms de domaine pour votre site ?

- par prévention pour éviter que d'autres achètent des noms de domaine très proches du vôtre,
- par souci pratique pour les internautes qui tapent directement l'adresse en inversant .fr et .com par exemple,

Si vous réalisez des rediretions permanentes, assurez-vous qu'un seul site est référencé sur Google! Et méfiez-vous des **redirections "invisibles"** proposées par hébergeurs comme OVH, qui créent des pages "exotiques" contenant des iframe, en déposant le contenu des sites dedans.

Utilisez la requête "site:" dans Google, pour vous assurer qu'un seul nom de domaine est indexé!

#### **▼ 2.9.6 L'ordre des paramètres d'un QueryString**

Si vous avez un site dynamique et que vous n'avez pas encore mis en place la réécriture d'URL, vous avez peut-être des URL qui contiennent plusieurs variables comme:

page.php?t=2534&postdays=0&postorder=asc&start=15

Le problème est que cette page est accessible également aux URL suivantes:

page.php?t=2534&postorder=asc&postdays=0&start=15

page.php?postorder=asc&start=15&t=2534

L'URL-rewriting est essentiel pour corriger le tir, et obtenir des URL statiques!

#### ▼ 2.9.7 Les balises <noscript> et <noframes>

Ces balises servent à l'origine à saisir un contenu alternatif, en cas de désactivtion du javascript, ou dans le cas où on utiliserait un browser d'ancienne génération. Ces balises peuvent donc servir à écrire un contenu qui sera parcouru par les bots, et très rarement vu par les humains. Mais attenton à ne pas tomber dans le piège du "Puisque ces balises ne seront jamais exploitées par 99,99% des internautes, autant y greffer le même contenu que ma page visible". Evitez cette erreur, car les bots le perçoivent comme une tentative de spam-indexing. Ecrivez-y un contenu différent, autant que possible!

#### **2.9.8** Le cas des blogs

Dans le cas des blogs, évitez donc au maximum de publier vos articles dans plusieurs catégories.

Par exemple, vous avez 5 articles, tous classés sous la catégorie Cat-A et la catégorie Cat-B avec les mots-clés Tag-A et Tag-B. Les URL <a href="http://www.blog.com/categorie/cat-A/">http://www.blog.com/categorie/cat-A/</a> et <a href="http://www.blog.com/categorie/cat-B/">http://www.blog.com/categorie/cat-A/</a> et <a href="http://www.blog.com/tag/tag-A/">http://www.blog.com/tag/tag-A/</a> et <a href="http://www.blog.com/tag/tag-B/">http://www.blog.com/tag/tag-A/</a> et <a href="http://www.blog.com/tag/tag-B/">http://www.blog.com/tag/tag-B/</a> auront donc un contenu presque similaire, car ils contiennent les mêmes articles. Solution: il ne faut pas indexer ces pages. Utilisez des métaélément "noindex" pour cela!

Pour encore optimiser, faites en sorte que les articles sur la page d'accueil soient différents de la page de l'article elle-même, pour cela 2 moyens:

- le plus pratique, car diffusant un nouveau contenu: on écrit un extrait compact de l'article,
- le moins pratique: on crée une césure / coupure dans l'article, pour n'obtenir en page d'accueil qu'une indroduction au contenu global de l'article.

Surtout, évitez que votre page d'accueil comporte des articles non modifiés ou non coupés, qui ont eu lien "zoom" avec le même contenu!

### ▼ 2.9.9 Le cloaking

Le cloaking est une technique visant à faire indexer par les moteurs de recherche un contenu spécifique et non visible par les internautes afin d'obtenir un bon positionnement sur des mots-clés donnés. Cela consiste donc à créer des pages très (trop ?) fortes en contenu, qui seraient devenues illisibles par les humains. Le serveur web va comparer l'adresse IP, ou le user-agent du spider, et lui fournir une page en attente avec ses critères de jugement.

Cette technique est utilisée en positionnement sur des mots-clés ultra-concurrentiels, afin d'éviter que le code d'une page ressortant en bonne position ne soit par exemple réutilisée

frauduleusement par un concurrent. Toutefois, son usage est controversé, même si l'on ne peut pas l'assimiler véritablement à du spamdexing. En effet, suite à de nombreux cas d'atteintes au copyright par des sites utilisant des pages de leurs concurrents, à des fins de positionnement, en **cloakant** celles-ci, certains moteurs de recherche sont devenus très méfiants vis à vis de cette technique.

Google interdit explicitement cette technique. De toutes façons, le recours au cloaking est inefficace sur ce moteur puisqu'il permet à l'internaute d'accéder à une version de la page **mise en cache**. D'autres moteurs ont une politique plus pragmatique, et ne sanctionnent en général que les abus liés à cette technique.

Pour tester cette technique, utilisez le pluq-in UserAgentSwitcher sous Firefox: http://extensions.geckozone.org/UserAgentSwitcher/.

## **▼ 2.9.10 Référencer des sites multilingues**

Lorsque vous développez un site en plusieurs langues, 4 solutions majeures se présentent:

- Réserver un nom de domaine par langue. ex: www.example.fr, www.example.it
- Créer un répertoire par langue. ex: example.com/fr, example.com/it
- Créer un sous-domaine par langue. ex: fr.example.com, it.example.com
- Afficher plusieurs langues sur la même page.

Cela est typique dans les CMS open-source (Typo3, Joomla!, etc.), le passage d'un paramètre "lang" est monnaie courante. Gâre aux **duplicate content**: <u>index.php</u>, <u>index.php?lang=1</u> ou encore index.php?lang=0, si 0 appelle la langue par défaut.

Le changement de langue par cookie est également à proscrire, car les bots ne verront pas les cookies, mais seulement la langue par défaut!

L'idéal reste l'achat de noms de domaine séparés. Ainsi, on peut obtenir des **backlinks dédiés** à chaque version du site, et cela permet d'optimiser l'indexation par les bots, puisque les moteurs de recherche identifient la localisation géographique des sites avec l'emplacement du serveur et l'extension du nom de domaine.

#### **▼ 2.9.11 Le First Click Free**

Le principe du système **First Click Free** de Google est de permettre aux sites ayant un contenu de type extranet, accessible uniquement aux membres ou aux clients, de le faire indexer par Google, afin d'obtenir en retour du trafic issu des recherches Google. Concrètement, Google propose ni plus ni moins de faire du cloaking:

- sur vos pages à accès restreint, vous mettez en place une détection du type de visiteur:
  - □ si c'est Googlebot, vous lui laissez l'accès pour qu'il indexe tout le contenu, et lister les résultats dans son SERP,
  - □ si c'est un internaute, vous ne lui donnez l'accès que s'il a payé, par exemple.

Cela peut freiner certains webmasters, puisque, avec ce système, Google a accès gratuitement à 100% d'un contenu que l'éditeur a pourtant choisi de faire payer. Google peut indexer l'ensemble du contenu et non pas seulement un titre et un chapeau d'article, ce qui lui permet de faire apparaître les pages concernées pour un très grand nombre de requêtes.

• si un internaute arrive sur votre site (rubrique payante) en provenance de Google, Google exige que vous laissiez l'internaute consulter votre page de contenu. Bien entendu cet internaute ne pourra pas consulter d'autres pages de votre site sans payer : seule la première page consultée en provenance d'une recherche Google est gratuite, d'où le terme "First Click Free".

Se pose aussi la question économique: un internaute qui aura acès à une page gratuitement, et s'il y trouve le contenu recherché, paiera t-il pour obtneir le reste du contenu du site?

Par ailleurs, à l'heure actuelle, impossible de dire si les sites ne respectant cette requête de "First Click Free" seront blacklistés. Et on peut également se poser la question de l'optimisation du

positionnement de ces pages dans les SERP, puisqu'elles n'auront certainement pas de backlinks!

En réalité, le principe du First Click Free peut être contourné facilement avec un autre système assez proche, consistant à faire indexer le titre et l'introduction de chaque article payant : c'est le principe des **archives payantes partiellement indexées**, et pour lire l'article en entier, l'internaute doit payer.

# **▽ 2.10 Cas particulier: optimiser un blog**

La manière dont vous allez bloguer est aussi un facteur important en terme de référencement.

Pensez aussi à ceci:

- Publiez des articles pertinents avec des liens internes vers vos pages. Attention, les articles doivent être uniques assez long, et une limite raisonnable de 3 liens vers votre site.
- Mettez en avant le flux RSS ou Atom et facilitez les moyens de s'y abonner. Il existe désormais des icônes officielles pour symboliser le flux (visitez <u>SmashingMagazine</u>). Utilisez-les pour créer des boutons d'abonnement aux flux via les principaux outils.
- N'oubliez pas de prévoir une balise < link /> dans vos pages qui fasse référence à votre flux, afin de permettre aux navigateurs ou aux outils spécialisés de détecter le(s) flux associé(s) à votre site. Firefox et Internet Explorer 7 les détectent (et affichent le logo du flux) et sans doute aussi d'autres navigateurs! Profitez également du texte alternatif pour diffuser un titre de flux.

```
<link rel="alternate" type="application/rss+xml" title="Presse-Citron RSS Feed" href="http://www.presse-citron.net/feed" />+
```

Voici un exemple de contenu de fichier RSS 2.0. Chaque article est associé à un titre, un descriptif, et à une URL, la plus courte possible!

```
<?xml version="1.0" encoding="iso-8859-1"?>
<rss version="2.0">
    <channel>
       <title>Mon site</title>
       <description>Ceci est un exemple de flux RSS 2.0</description>
       <lastBuildDate>Wed, 27 Jul 2008 00:30:30 -0700/lastBuildDate>
       <link>http://www.site.com</link>
       <item>
           <title>Article N°1</title>
           <description>Ceci est ma première actualité</description>
           <pubDate>Tue, 19 Jul 2008 04:32:51 -0700</pubDate>
           <link>http://www.site.com/actu1</link>
       </item>
       <item>
           <title>Article N°2</title>
           <description>Ceci est ma seconde actualité</description>
           <pubDate>Tue, 19 Jul 2008 04:32:51 -0700
```

Pour savoir si votre flux est lu par beaucoup de monde, et quel trafic il vous génère, il vous faut des **agrégateurs de flux**. Certains outils de mesure d'audience proposent d'analyser le trafic de votre flux, c'est le cas par exemple de Xiti, mais vous pouvez utiliser également des outils spécifiques comme:

☐ FeedBurner feedburner.com/

Note FeedBurner réalise même un "hot list" des meilleurs flux du moment.

- □ SimpleFeed simplefeed.com/
- □ Nooked nooked.com/
- ☐ MeasureMap, de Google measuremap.com/
- ☐ SiteMeter <u>sitemeter.com/</u>

Note Vous voulez plus d'info sur l'augmentation d'abonnées de vos flux RSS ? rendez-vous sur le blog de Techtrends:

http://techtrends.eu/blog/2008/10/13/une-idee-pour-augmenter-le-nombre-dabonnes-rss-des-petits-blogs/

Le seul inconvénient important de ces outils est que votre flux est géré en dehors de votre site, alors qu'il est préférable d'avoir un flux dont l'URL fait partie de votre nom de domaine.

Note Dérivé du flux RSS traditionnel, le flux mRSS ("media-RSS") est un format de syndication de contenus multimédias : vidéos, images etc. Voici un exemple de fichier mRSS:

```
<?xml version="1.0" encoding="utf-8"?>
<rss version="2.0" xmlns:media="http://search.yahoo.com/mrss/" xmlns:av="http://www.searchvideo.com/schemas/av/1.0">
 <channel>
   <title>Site.com</title>
   <link>http://www.site.com/rss/videos/mrss.xml</link>
   <description>Description des vidéos du flux</description>
   <item>
     <link>http://www.site.com/video.avi</link>
     <media:title>Titre de la vidéo</media:title>
     <media:description>Texte de description</media:description>
     <media:player url="http://www.site.com/video.avi" />
     <media:thumbnail url="http://www.site.com/thumbnail.jpg" width="50" height="50" />
     <media:keywords>Liste de mots clés</media:keywords>
    </item>
 </channel>
</rss>
```

Le mRSS permet, comme le sitemap vidéo, de définir un contenu de type "titre" et "descriptif" pour vos contenus multimedia. Respectez les prérogatives générales des métaéléments pour les saisir correctement.

- Utilisez une feuille de styles XSL (=transformateur XML) pour votre flux afin de rendre votre flux visible par tous.
- Proposez plusieurs flux (articles, commentaires, par catégorie, etc.)
- Dans vos flux, mettez un nombre important d'articles, par exemple 20.
- Faites-vous connaître de la blogosphère en tissant des liens avec d'autres blogs de votre domaine d'activité: installez un blogroll.
- Mettez en évidence l'URL permanente de chaque billet.
- Proposez à vos lecteurs d'ajouter votre billet dans leurs outils de bookmarks sociaux comme del.icio.us, etc.
- Proposez à vos lecteurs de voter pour votre billet dans leurs outils du type Digg (en France on pourrait citer Fuzz, Scoopeo, TapeMoi ou Wikio).
- Ajoutez des tags à vos billets, et pourquoi pas un nuage de tags, selon vos envies.
- Définissez vos articles principaux en tant que référence. Ces articles sont listés sur la page d'accueil, ou toutes les pages de votre blog selon la construction établie, et sont donc mises en avant, et obtiennent un meilleur PageRank.

### **▼ 2.10.1 Optimiser son référencement avec un blog WordPress**

Liste de plugins utiles:

# Adhesive + WP Sticky

http://www.hongkiat.com/blog/adhesive-wordpress-sticky-plugin/ + http://www.lesterchan.net/wordpress/readme/wp-sticky.html

Met en avant certains billets sélectionnés, en les affichant en premier sur la page d'accueil ou sur les listes d'articles par catégorie.

#### WordPress Canonical URL

markjaquith.wordpress.com/2007/09/25/wordpress-23-canonical-urls/

Fait en sorte que vos billets soient tous obtenus avec la bonne URL, pour éviter les problèmes inhérents au Duplicate Content.

#### **Contextual Related Posts**

http://weblogtoolscollection.com/archives/2005/11/27/wp-plugin-contextual-related-posts/

Liste les billets similaires grâce aux tags, sur les billets en cours.

#### **WordPress Related Posts**

http://wordpress.org/extend/plugins/wordpress-23-related-posts-plugin/

Liste les billets similaires grâce aux tags, et les ajoute sur les flux RSS.

## **Google XML Sitemaps Generator**

http://www.arnebrachhold.de/projects/wordpress-plugins/google-xml-sitemaps-generator/

Crée automatiquement un fichier Google Sitemap à partir de votre blog.

# All in One SEOPack + wpSEO

http://wp.uberdose.com/2007/03/24/all-in-one-seo-pack/ + http://www.wpseo.us/

Quelques options SEO pratiques, qui permettent notamment de gérer certains problèmes de contenu dupliqué, causés par l'utilisation des archives, tags ou catégories.

#### **Permalink Redirect**

http://scott.yang.id.au/code/permalink-redirect/

Fait en sorte que vos billets soient tous accédés avec la bonne URL, sinon effectue une redirection 301, pour éviter à nouveau les problèmes de Duplicate Content.

## SEO Title Tag + Headspace 2 + Add meta tags

http://www.netconcepts.com/seo-title-tag-plugin/ + http://urbangiraffe.com/plugins/headspace2/ + http://www.g-loaded.eu/2006/01/05/add-meta-tags-wordpress-plugin/

Optimise la balise <title> et les métaéléments description et keywords partout dans votre blog.

# Google Analytics and Feedburner Reports plugin for WordPress

http://tantannoodles.com/toolkit/wordpress-reports/

crée des rapports de statistiques sur le trafic de votre blog, basés sur FeedBurner et Google Analytics.



# SH-Autolink + WP Keywords Autolink

http://www.sethson.com/super-version-of-the-sh-autolink-plugin-for-wordpress/ + http://www.webinventif.fr/keywords-autolink-2/

Transforme automatiquement une liste de mots pré-définis en liens hypertextes, les faisant pointer vers l'intérieur du blog.

# **Ultimate TagWarrior**

http://lorelle.wordpress.com/2005/10/23/the-ultimate-tag-warrior-wordpress-plugin/

Pour mettre en avant vos tags.

#### **Breadcrumb Nav Xt**

http://mtekk.weblogs.us/code/breadcrumb-nav-xt/

Pour optimiser ses fils d'Ariane.

#### Link to me

www.jimwestergren.com/link-to-me-textbox-wordpess-plugin/

Ajoute un texte à la suite du billet, suggérant aux lecteurs de faire un lien vers ce billet.

# SearchPages + Search meter

http://www.internetofficer.com/wordpress/search-pages/ + http://www.thunderguy.com/semicolon/wordpress/search-meter-wordpress-plugin/

Permettant à la fonction "recherche du blog" de scanner, non seulement les articles, mais aussi les pages statiques. Recherches ensuite analysées par Search meter.

# Simple Tags + WP Cumulus

http://wordpress.org/extend/plugins/simple-tags/

Pour faciliter la création des nuages de mots-clés en headings, ou avec du WPCumulus, si l'aspect visuel priorise sur les objectifs de réferencement:



# **Popularity Contest**

http://alexking.org/projects/wordpress/readme?project=popularity-contest

Pour afficher les articles les plus populaires.

#### Nofollow + Dofollow

http://www.michelem.org/wordpress-plugin-nofollow-free/ + http://kimmo.suominen.com/sw/dofollow/

Partager la popularité de son blog avec les commentateurs: liens des commentaires avec nofollow ou follow!

# Bling This + ShareThis

http://www.blingblog.info/2006/09/18/plugin-wordpress-bookmarking-v12/ + http://alexking.org/blog/2007/11/08/sharethis-20

Soumettre un article aux gestionnaires de favoris et Digg-likes.

# **Multi URI Ping**

http://mani.wordpress.com/2005/10/24/multi-uri-plugin/

Améliore la gestion native des pings, rendant plus efficace l'envoi de pings à un grand nombre de serveurs. Le ping détermine la qualité d'une connexion sur internet. Sa valeur en millisecondes doit être proche de 0. Voici un exemple de configuration du plugin Multi URI Ping:

```
http://rpc.technorati.com/rpc/ping
http://blogsearch.google.com/ping/RPC2
http://www.vizibox.com/ping/ping.php
http://rpc.weblogs.com/RPC2
http://ping.syndic8.com/xmlrpc.php
http://ping.blo.gs/
http://joueb.com/cgi-bin/rpc.pl
http://api.moreover.com/RPC2
http://www.blogpeople.net/servlet/weblogUpdates
http://pinger.blogflux.com/rpc/
http://www.blogoon.net/ping/
http://ping.feedburner.com/
http://www.holycowdude.com/rpc/ping/
http://xping.pubsub.com/ping/
http://topicexchange.com/RPC2
http://ping.weblogalot.com/rpc.php
http://bblog.com/ping.php
http://ping.bitacoras.com/
http://rpc.blogbuzzmachine.com/RPC2
http://www.blogdigger.com/RPC2
http://ping.cocolog-nifty.com/xmlrpc
```

# **▽** 3 Suivre son réferencement

# **▼ 3.1 Google Webmaster Tools**

Google Webmaster Tools <a href="https://www.google.com/webmasters/tools/">https://www.google.com/webmasters/tools/</a> aide à maîtriser son référencement, à décrypter le feedback du comportement de GoogleBot et peut être très utile pour trouver

des erreurs dans les liens entrants et sortants.

Dès l'acuel de l'outil, vous trouverez la liste de tous vos sites Internet. Si c'est la première fois que vous vous y connectez, il faut commencer par ajouter un site en entrant l'URL puis en cliquant sur "Ajouter un site". Vous arrivez alors sur une deuxième page qui sera la même pour tous vos sites: la "Vue d'ensemble". Avant de voir ensemble les différents menus de cet outil, il vous faut valider votre site c'est-à-dire prouver à Google qu'il s'agit bien d'un site dont vous êtes l'auteur, le propriétaire ou le webmaster. Cette étape est très importante, car sans vérification, vous n'aurez pas accès aux informations concernant le référencement de votre site.

Pour cela, cliquez sur "Valider votre site". Google vous propose alors 2 méthodes:

- En ajoutant une balise Meta définie par Google
- En transférant un fichier HTML vide sur le serveur de votre site Internet

Suite à cette vérification, Google vous indique que le site est maintenant validé. Vous avez alors accès aux informations le concernant:

#### Vue d'ensemble

Cette page vous donne **un aperçu global de l'indexation** de votre site web. Vous y apprenez par exemple **la date du dernier crawl de Googlebot**. Plus intéressant ensuite, vous avez une liste des erreurs rencontrées par Google lors de l'exploration de votre site Internet. Si vous avez des erreurs signalées par un petit point d'exclamation jaune, le lien correspondant vous renvoie vers une page d'explication.

## Diagnostic > Exploration du Web

Cette sous-rubrique est très importante car elle répertorie toutes les erreurs rencontrées par les robots de Google lors de l'analyse de votre site. Il faut donc prendre le temps de bien lire les erreurs et de voir comment les résoudre. Les erreurs sont classées par type (erreurs http, introuvable, inaccessibles, etc). Pour chaque erreur listée, Google vous indique la raison exacte du problème et une explication pour résoudre le problème.

### Diagnostic > Exploration du Web mobile

Cette sous-rubrique est identique à la précédente, mais ne concerne que les sites proposant une version pour les téléphones portables.

#### Diagnostic > Analyse du contenu

Cette dernière sous-rubrique est aussi très importante car elle répertorie les problèmes rencontrés au niveau des balises de vos pages ou sur le contenu de vos pages. Par exemple, Google vous informe des pages qui possèderaient toutes la même description ou des balises trop courtes. Google vous indique comment remédier au problème le cas échéant.

#### Statistiques > Vue d'ensemble

Quelques stats intéressantes qui vous donnent de très bonnes informations concernant le comportement des utilisateurs:

- Requêtes les plus fréquentes, répertorie les 20 requêtes pour lesquelles votre site est apparu le plus souvent dans les SERP,
- Requêtes suivies de clics les plus fréquentes, répertorie la liste des recherches à partir desquelles les internautes ont cliqué sur votre snippet,

Il est possible de trier les données par ancienneté des recherches (il y a une semaine, il y a 2 semaines, le mois précédent, etc) mais aussi par pays et même par type de recherche (Google Search, GoogleBlog, etc.).

#### Statistiques > Ce que voit Googlebot

Vous obtenez ici la liste de tous les mots ou phrases que Google à trouvé sur d'autres sites et comportant un lien vers votre site Internet. Vous pouvez ainsi vérifier tous les mots utilisés par les autres sites Internet qui parlent de vous, et connaître la véritable situation de vos backlinks!

### Statistiques > Statistiques relatives à l'exploration

Cette sous-rubrique n'est remplie que si votre site Internet possède déià un peu d'ancienneté, puisqu'elle indique le Pagerank de chaque page de votre site.

#### Statistiques > Statistiques relatives aux abonnés

Cette sous-rubrique est une des plus récentes ajoutée à l'outil. Elle vous indique, dans le cas où votre site propose un ou plusieurs flux RSS, comme souvent le cas sur un blog, le nombre de personnes qui se sont abonnés à vos flux RSS via un outil de la panoplie Google, comme Google Reader, iGoogle ou Orkut. Cette liste ne comporte pas les flux RSS générés par Feedburner.

### Liens > Pages avec des liens externes

Cette sous-rubrique est une analyse de votre deep-linking; elle répertorie la liste des pages de votre site Internet vers lesquelles d'autres sites Internet ont créé des liens. Vous avez ainsi connaître, pour chaque page de votre site, le nombre et la liste des autres sites pointant vers cette page. Comme cette liste est triée, vous pouvez facilement savoir quelles sont les pages vers lesquelles le plus de personnes pointent et donc sont potentiellement les plus intéressantes. Cela représente un très bon complément à la commande "link:monsite.com", qui ne liste que les sites pointant vers la page d'accueil de votre site.

## Liens > Pages avec des liens internes

Cette sous-rubrique complète l'analyse deep-linking, et indique pour chaque page de votre site, la liste des liens internes (donc, uniquement entre les pages de votre site) pointant vers cette page.

#### Liens > Liens de site

Dans le cas où votre site est bien indexé et a déjà un peu d'ancienneté, Google présentera des "sitelinks" dans sa liste de résultat. Si tel est le cas, cette sous-rubrique vous permettra de supprimer certains des liens présents dans la liste de résultats, et qui sont ajoutés par défaut par Google.

Les sitelinks ont évolué ces derniers temps, passant de 4 liens, à 8 liens:

# Décathlon - magasin de sports, article de sport, équipement sportif ...

magasin de sports, article de sport, équipement sportif, magasin sport, chaussure de sport, matériel de sport, vêtement de sport, maillot sport, maillot bains ...

www.decathlon.fr/ - 48k - En cache - Pages similaires

Vélo - www.decathlon.fr/.../cycle-and-rolle/bikes/index.html

ski-snowboard-luges - www.decathlon.fr/.../mountain/ski-snowboard/index.html

Nos magasins - www.decathlon.fr/.../homepage-mag/index.html

Randonnée - www.decathlon.fr/.../mountain/hiking/index.html

Autres résultats, domaine www.decathlon.fr »

# Cuisine AZ, Recettes de cuisine de A à Z

Cuisine AZ, 101 000 délicieuses idées et recettes de cuisine classées par thèmes ou par catégories , avec photos, faciles ou techniques, 100% testées et ...

www.cuisineaz.com/ - 42k - En cache - Pages similaires - À noter

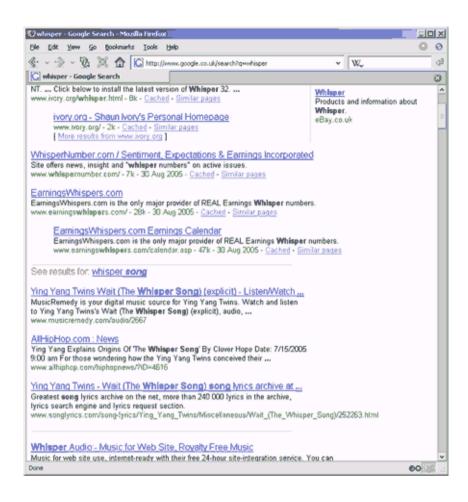
Rechercher une recette Affiner ma recherche

Nos recettes d'entrée Dossiers
Express Verrine
Minceur du monde

Autres résultats, domaine cuisineaz.com »

Pour obtenir des sitelinks, c'est assez flou. La communauté des référenceurs a déterminé toutefois les points suivants:

- Il ne dépend pas du PageRank,
- Il ne dépend pas d'un fichier Google Sitemap,
- Il ne dépend du nombre de pages indexées ni du nombre de backlinks,
- Il ne dépend pas de la présence du site dans DMOZ.
- Il faut être premier sur un résultat de recherche,
- Il faut avoir une structure interne de liens correcte (bon maillage de lien),
- L'âge du site est important,
- Le site doit être beaucoup plus pertinent que le deuxième résultat. D'où la concordance avec une recherche liée à votre nom de domaine,
- Il semble que ces sites bénéficient d'un bon TrustRank, qui détermine le site comme étant valide et de qualité. Et à 100% conforme avec la requête.
- Cet affichage fait perdre le clustering, un snippet inférieur décalé vers la droite, généralement une page très proche du premier lien:



# **Plans Sitemap**

Cette rubrique vous liste les sitemap XML que vous avez ajoutés pour votre site, et vérifier qu'ils sont bien parcourus par Google.

#### Outils > Analyser & Générer le fichier robots.txt

Cette sous-rubrique est un formulaire pour vous accompagner dans la création d'un fichier robots.txt correct. N'oubliez pas que cette option est très importante pour la gestion des URL à ne pas indexer, en cas de Duplicate Content. Utilisez la requête "site:" dans Google pour tester les doublons.

### Outils > Définir la vitesse d'exploration

Cette sous-rubrique vous indique à titre indicatif à quelle vitesse/fréquence Google indexe et parcourt votre site Internet. Vous pouvez diminuer cette vitesse. L'intérêt reste flou.

#### Outils > Définir la zone géographique ciblée

Cette sous-rubrique est très intéressante, car elle vous permet de définir le pays correspondant à votre site Internet. Google attribue en général automatiquement cette donnée à un site Internet en fonction de la **situation géographique de votre hébergeur**, ou de l'extension du nom de domaine. Donc si vous hébergez votre site aux USA mais que vous souhaitez toucher un public

français, cette option vous permettra de forcer Google à associer votre site au pays France. Ainsi quand quelqu'un cherchera quelque chose en spécifiant "France", voire "pages francophones" dans Google, vous apparaîtrez.

#### Outils > Définir le domaine favori

Cette sous-rubrique vous permet de forcer Google à utilise soit une URL avec les "www" devant votre nom de domaine soit sans les "www". Il ne faut pas oublier de configurer ce point, en faisant à nouveau attention aux Duplicate Content.

## Outils > Recherche d'images avancée

Cette sous-rubrique vous permet d'autoriser Google à mieux indexer les images de votre site pour son moteur de recherche par images. N'hésitez pas à cocher la case!

# Outils > URL à supprimer

Le Google URL Removal. Cette sous-rubrique présente une fonctionnalité pratique: elle vous permet de bloquer l'accès à une page spécifique. Google indique pour cela toute la marche à suivre pour réussir une telle opération.

### Outils > Gadgets

Cette dernière sous-rubrique vous propose d'installer un widget Google sur votre page iGoogle pour suivre les données de GWT directement depuis iGoogle.

Note Yahoo propose également sa suite d'outils: http://siteexplorer.search.yahoo.com/

# **▼ 3.2 Google Add URL**

Un peu désuet depuis la sortie de la suite Google Webmaster Tools et ses sitemap XML très efficaces, l'ancien outil d'ajout de site auprès de Google est toujours disponible ici: <a href="http://www.google.fr/addurl/?hl=fr&continue=/addurl">http://www.google.fr/addurl/?hl=fr&continue=/addurl</a>. Il vous suffit dans c cas de poster simplement la page d'accueil, Google se débrouillera tout seul avec le maillage interne.

Préférez la suite GWT, beaucoup plus efficaces, et bourrée d'informations capitale de suivi!

# ¬ Chapter 3: Optimiser la notoriété de son site web

Nous avons déjà vu, que pour générer du trafic, il faut optimiser la popularité de son site, et donc de son URL:

- 1. par des "backlinks" (ou BL, en somme des liens sur des sites web partenaires qui mènent vers notre site). L'objectif étant de dénicher des "natural backlinks", et pas seulement sur la page d'accueil de son site, mais bien à l'intérieur de son site (deep-linking).
- 2. en évitant d'avoir une URL trop compliquée, ... mais aussi en communiquant off-line!

# ▼ 1 Le Net-linking

Plus une page a de liens pointant vers elle, plus elle a de chance d'être bien placée dans les SERP. Pour avoir beaucoup de liens, faites du contenu original, intéressant, et unique!

Quelques prérogatives à suivre, certains points ont déjà été survolés dans ce cours:

- faites héberger des pages sur d'autres sites, et publiez des articles sur des sites tiers, cela créera facilement des backlinks,
- echangez des liens dans le corps des pages, et pas dans les liens bas de page ou pages partenaires, dévalorisés par Google,

échangez des liens avec vos fournisseurs, vos clients, etc.

et n'oubliez pas, travaillez sur un schéma de "deep-linking" (=liens profonds), pour profiter du maillage interne de votre site, et pas seulement de la page d'accueil!

Note Voici un outil basé sur l'API Yahoo qui va vous permettre d'analyser le taux de deep-linking: http://www.webrankinfo.com/outils/liens-profonds.php

Pourcentage de liens profonds pour cuisineaz.com :

- Liens (externes) vers tout le site : 104 000
- Liens profonds (externes): 22 000
- □ Taux de liens profonds (externes) : 21,2%

- inscrivez-vous dans des annuaires,
- diffusez des liens sociaux, et des liens vers des digg-like (scoopeo, delicious, etc.) pour gagner en backlinks depuis des plateformes web 2.0 gratuites

Les groupwares (YouTube, Flickr, etc.) et les social bookmarking (Digg.com, Del.icio.us) ont une grande place dans le SEO, car ils permettent justement le partage de données et la diffusion massive d'une URL. Un site conçu ainsi en "Rich-media" se verra indexé dans différents types de moteurs (Google Image, Google Vidéo, etc.), ce qui augmentera son pagerank.

N'oubliez pas de créer du contenu sur Wikipedia (même si les liens sont en nofollow, cela crée du trafic et indirectement d'autres liens), de stocker vos coordonnées sur GoogleMap, de créer des profils Twitter, Facebook, Viadéo, etc.

testez vos positionnements grâce aux requêtes Google et Yahoo:

### le requête "site:" pour tester l'indexation

Nous avons déjà vu qu'en tapant "site:monsite.com" dans Google, vous verrez exclusivement les pages indexées du site monsite.com. Pour affiner vos recherches, vous pouvez requêter:

- "site:monsite.com/dossier/2008/" et ainsi vérifier si le contenu d'un path est stocké dans l'index de Google.
- "site:domaine.monsite.com/dossier/2008/" et ainsi restreindre par sous-domaine,
- "referencement site:domaine.monsite.com/dossier/2008/" et ainsi lister les pages d'un path précis d'un sous-domaine précis, et liés au réferencement.

Attention, cette requête n'offre pas une précision infinie, en naviguant dans la pagination, on peut déjà voir le nombre de résultats totaux varier.

### le requête "link:" pour analyser les backlinks (Google)

la commande "link:" est bridée sur Google:

- elle est très loin d'être exhaustive : Google n'affiche parfois qu'1% des backlinks
- elle est rarement mise à jour: environ tous les mois voire moins souvent
- elle ne peut pas être combinée avec d'autres opérateurs ni même avec des mots-clés, comme la requête "site:"
- la commande "link:URL" ne renvoie que les pages faisant un lien vers la page URL. Etant donné que le travail de référencement se fait page par page, il faut en théorie utiliser la commande "link:" pour chaque page pour laquelle on cherche à optimiser le référencement. Un bon référencement est un travail de fourmi!

les requêtes "link:" & "linkdomain:"pour analyser les backlinks (Yahoo)

Yahoo, MSN et d'autres moteurs hormis Google diffusent la requête linkdomain, qui liste les pages faisant un lien vers n'importe quelle page d'un site donné. On peut bien mieux analyser le deeplinking d'un site.

Quelques exemples d'utilisation:

- "link:http://www.monsite.com -site:monsite.com", va lister les liens externes faisant un lien vers la page d'accueil,
- "linkdomain:http://www.monsite.com -site:monsite.com", va lister les liens externes faisant un lien vers toutes les pages du site,
- "referencement linkdomain:http://www.monsite.com -site:monsite.com", va lister les liens externes faisant un lien vers toutes les pages du site contenant "referencement"
- "linkdomain:http://www.monsite.com site:example.com", va lister les backlinks en provenance de example.com, faisant un lien vers toutes les pages du site.

Pour obtenir un résumé des meilleures tendances de link-building, rendez-vous ici: http://www.webrankinfo.com/actualites/200802-idees-de-link-building-en-2008.htm

# **▼ 1.1 Les annuaires**

Les annuaires sont de différentes formes, certains sont gratuits, d'autres payants.

L'annuaire payant le plus connu est "les pages jaunes". Le tarif évolue en fonction des options qui y figurent (affichage de l'URL de son site, etc.). Il intéressa toutefois beaucoup de PME pour qui les "pages jaunes" demeurent un pilier. D'autres annuaires payants travaillent sur un **système d'accroche commerciale**: les comparateurs de prix par exemple (Kelkoo, etc.).

D'autres annuaires gratuits, ne sont pas moins intéressants: Google Annuaire, DMOZ, etc. D'autres annuaires sont moins généralistes, et spécialisés par thématique. Il convient dans ce cas à bien définir son besoin.

Attention toutefois aux annuaires qui vous demandant un lien en retour, ils ne travaillent donc pas sur des "natural backlink" ce qui risque de nuire à notre PageRank. Certains blogeurs conseillent de boycotter ces annuaires.

Vous trouverez ici une liste régulièrement complétée, contenant des annuaires "de confiance", classés par thématiques (généralistes, ou spécialisés): http://directory-directories.rankbl.com/

# **□** 1.2 Le ping, pour les flux RSS

Pour informer certains sites que **vous venez de publier un nouveau billet** dans votre blog, et que ces billets se retrouvent dans vos flux RSS, il faut envoyer un **ping**. Nous avons déjà vu cela brièvement au chapitre précédent avec WordPress.

Certains sites proposent d'envoyer un ping à toute une série de sites d'un seul coup, ce qui fait gagner du temps. En voici quelques-uns:

- FeedPing feedping.com/
- Pingoat <u>pingoat.com/</u>
- Pingomatic <u>pingomatic.com/</u>
- PingShot feedburner.com/fb/a/publishers/pingshot
- Weblogs weblogs.com/

A utiliser abondamment!

# 

Le linkbaiting consiste à créer une très forte envie chez les bloggeurs et webmasters de créer un backlink depuis leur site, vers le votre. Cette idée s'assimile à l' "aimant à liens".

L'idée est de susciter une émotion particulière chez l'internaute (blagues, actualité, jeux, polémiques...) pour le motiver à stocker spontanément un lien. Cette mode se répand essentiellement avec les blogs.

C'est la meilleure forme de backlink "naturel", puisqu'elle témoigne d'un contenu riche et intéressant de votre site/blog.

# **▽ 2 Le Black Hat SEO**

Le principe du black hat, est souvent de court-circuiter un concurrent. En vue de diminuer sa visibilité sur les moteurs de recherche, et en agissant à sa place. Ces techniques sont bien sûr frauduleuses, et parfois prises en compte par des moteurs tels que Google. Toutefois, les situations sont souvent **subjectives**, et les cas peuvent être mal intéprétés.

Découvrons quelques cas partculiers.

# **▽ 2.1 Le Google Bombing**

Le Google Bombing, ou Bombardement Google, consiste à réaliser **un nombre de backlinks très conséquents** vers une page. Généralement, cette faille est exploitée des fins ludiques, politiques et économiques par les internautes.

Par exemple, pendant la campagne présidentielle en France en 2007, la blogosphère a créé de nombreuses pages, chacun ayant créé une simple page sur son nom de domaine, avec le texte "Programme Ségolène", avec un lien pointant vers l'article "Vide" de Wikipedia. Ce qui avait pour effet de spammer le SERP de Google de la candidate du même nom.

Depuis ces nombreuses manoeuvres douteuses, Google a déterminé que les liens devaient être obtenus régulièrement, et pas trop vite, sinon cela revient à considérer que la publicité du site est faite au détriment de son contenu!

D'autres moteurs, comme Yahoo, présentent toujours cette particularité.

# **▽** 2.2 Le spam referer

Le spam referer est une technique de spamdexing, comme son nom l'indique. Le principe: un site mal intentionné va stocker votre URL sur une de ses pages, le webmaster ou l'internaute va cliquer dessus, et vous obtiendrez donc le "referer" de ce site dans vos analyses de trafic, dans la partie "sources de trafic". C'est surtout un moyen de créer une publicité, plutôt que de gagner véritablement en PageRank. Cette technique est souvent utilisée par les sites pornographiques par exemple, pour faire croire au propriétaire du site que le site source est très générateur de trafic au point de favoriser la négociation de liens commerciaux.

Pour éviter de salir vos statistiques, n'hésitez pas à faire recours au .htaccess ; voici un exemple de code:

```
SetEnvIfNoCase Referer ".*(mot1|mot2|mot3|.....|motn).*" spammer=yes

Order allow, deny
allow from all
deny from env=spammer
```

Cela ne règlera pas le problème à 100% toutefois, mais éliminera déjà beaucoup de liens "parasites".

# □ 2.3 Les splogs (=spam blogs)

Le principe est double:

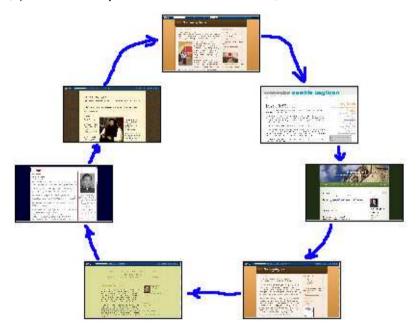
- dupliquer des mots clé et du contenu a forte valeur attractive depuis des blogs concurrents pour les insérer dans des blogs tiers et créer des backlinks vers des blogs cibles,
- insérer automatiquement des commentaires et des liens dans les bylines des blogs ou des forums, toujours dans cette optique de générer des liens vers des sites cibles.

Cette technique est souvent utilisée dans le cas d'un Google Bombing. blog a.k.a. splog

# **2.4 Le linkdoping**

Le principe: échanger et tisser un réseau de liens en grand nombre entre d'autre sites. Cela ressemble un peu à un système de Google Bombing, et risque plutôt de vous pénaliser pour les mêmes raisons déjà citées.

Toutefois, s'enregistrer auprès de web-rings, "anneau de site", qui relie des sites présentant un intérêt commun, tout en faisant un lien de l'un à l'autre, est une bonne solution.



# **▼ 2.5 302 Google jacking, ou page Hijacking**

La page hijacking, nommée aussi page hijack ou jacking, est une technique de spamdexing visant à dérober les positions d'autres sites sur les moteurs de recherche. Les sites visés sont souvent des sites populaires avec un trafic important et un grand nombre de pages. Elle se traduit en français par "détournement de page".

La methode est simple, il sagit de creer un site au contenu similaire de celui de votre concurent. Titles, meta tag, contenu, images, seul les liens internes changent. Cette technique de page jacking est possible gràce à la fonction "copy content" de Google, qui à pour but de supprimer toutes les pages au contenu similaire (Duplicate Content). Le moteur choisit une des deux pages, et place la

deuxiéme en "pages similaires" (non affiché dans les resultats de recherche). Cela n'est bien entendu possible qu'avec un nombre suffisant de pages car Google ne remplacera pas toutes les pages.

Les redirection 302 temporaires peuvent aussi être utilisées lors de la création de pages hijacking. Le but étant de faire croire à Google que la page n'existe plus et qu'elle a été temporairement déplacée à une autre URL. Un bon exemple peut être un annuaire qui utilise des redirections 302 sur les liens des sites qu'il référence. Dans ce cas les pages de redirections peuvent apparaître dans Google.

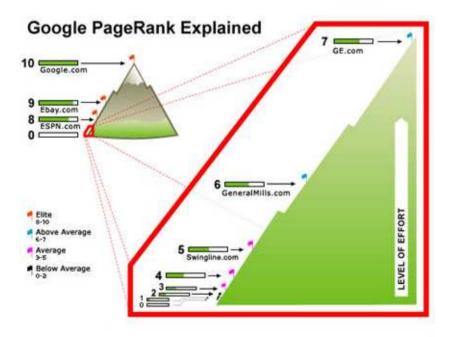
Pour plus d'informations: http://www.googlejacking.org/

# **▼ 3 Le Pagerank**

Le PageRank est un indice de popularité d'une page web, disponible avec la "toolbar" (TBPR) qui varie dans le temps.

Tout lien d'une page A vers une page B est considéré comme un "vote" en faveur de B. Si A est "sérieux", alors le PR de B va augmenter, ce qui va influer sur son "ranking". Chaque page a son propre pagerank.

Les 1ères marches sont les plus simples à gravir...



En réalité, l'indice TBPR est inexact, et est un condensé du "Real PR" qui va de 0 à l'infini. Algorithme de calcul du PR:

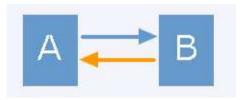
```
PR(A) = (1-d) + d(PR(T1)/C(T1) +...+ PR(Tn)/C(Tn))
--
PR(x): le PageRank de la page x
```

```
d: un coefficient d'amortissement égal à 0.85 C(x): le nombre de liens vers la page x
```

En réalité, l'indice TBPR est inexact, et est un condensé du "Real PR" qui va de 0 à l'infini.

# **3.1** exemple n°1: le PR tend toujours vers 1.0

CB(A)=CB(B)=1



# Si PR(A)=0

```
PR(A) = 0.15 + 0.85 * 0 = 0.15

PR(B) = 0.15 + 0.85 * 0.15 = 0.2775
```

```
PR(A) = 0.15 + 0.85 * 0.2775 = 0.385875
PR(B) = 0.15 + 0.85 * 0.385875 = 0.47799375
```

```
PR(A) = 0.15 + 0.85 * 0.47799375 = 0.5562946875
PR(B) = 0.15 + 0.85 * 0.5562946875 = 0.622850484375
```

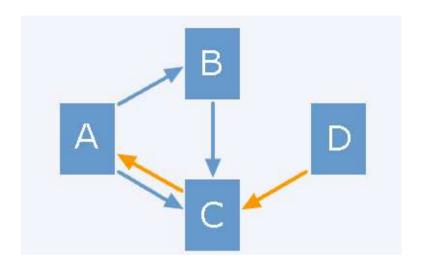
## Si PR(A)=2

```
PR(A) = 0.15 + 0.85 * 2 = 1.85
PR(B) = 0.15 + 0.85 * 1.85 = 1.7225
```

```
PR(A) = 0.15 + 0.85 * 1.7225 = 1.614125
PR(B) = 0.15 + 0.85 * 1.614125 = 1.52200625
```

```
PR(A) = 0.15 + 0.85 * 1.52200625 = 1.4437053125
PR(B) = 0.15 + 0.85 * 1.4437053125 = 1.377149515625
```

# **3.2** exemple n°2: le PR d'un site



# Au bout de 20 intérations:

• PR(A): 1.49

• PR(B): 0.78

• PR(C): 1.58

• PR(D): 0.15

Somme des PageRank 4.0

Moyenne du site 1.0

# **▽** 3.3 TBPR : Real PR

La traduction des véritables valeurs, entre le Toolbar PageRank, et le Real PageRank:

0/10:0.15-0.99

1/10 : 1.0 - 10

2/10 : 11 - 100

3/10:101-1000

4/10 : 1001 - 10 000

5/10 : 10 001 - 100 000

6/10 : 100 001 - 1 000 000

7/10 : 1 000 001 - 10 000 000

8/10:10:000:001-100:000:000

9/10 : 100 000 001 - 1 000 000 000

10/10 : 1 000 000 001 - ∞

# **3.4 Le TrustRank**

Le TrustRank (=indice de confiance) est défini par un homme et non un robot. Il peut influer sur le PR, si le PR est élevé, et que le TR est bas, le PR risque de décroître. Il est complémentaire au PR.

Google communique encore peu à ce sujet. Toutefois, nous avons vu les sitelinks dans le chapitre 2, qui sont le meilleur témoignage d'un bon TrustRank.

# **▼ 4 les concours de referencement**

Il existe des concours de réferencement, notamment à volonté ludique, ou pour vérifier ses véritables compétences en la matière, qui circulent sur le web.

Les concurrents doivent obtenir les meilleurs positionnements sur des termes tels que "Le mangeur de cigogne, "seraphim proudleduck", ou "sorcier glouton".

Pour en apprendre davantage sur le phénomène: http://www.noname.fr/concours-de-referencement/.

# **¬** Appendix A: Glossaire

# **▼ 1 Sources d'information**

Les sites ou blogs généralistes:

#### Abondance

http://www.abondance.com/

http://blog.abondance.com/

#### Webrankinfo

http://www.webrankinfo.com/

#### Zorgloob

http://www.zorgloob.com/

## Le Blog de SEO Camp, communauté francophone de réferencement

http://www.seo-camp.org/

## Le Blog de Matts Cutts, un membre de Google

http://www.mattcutts.com/blog/

Le Blog de SEO Blackout

# http://www.seoblackout.com/

# Search Engine Land

http://searchengineland.com/

# Le Blog de Sebastien Billard, membre actif de la communauté

http://s.billard.free.fr/referencement/

# Wikipedia

http://www.wikipedia.com

Les blogs officiels de Google

# **Google Analytics Blog**

http://analytics.blogspot.com/

# Google Inside AdWords Blog

http://adwords-fr.blogspot.com/

# Google Inside AdSense Blog

http://adsense-fr.blogspot.com/

# **▽ 2 Lexique**

Si des termes demeurent obscurs, utilisez la requête Google: "define:mot" pour obtenir une explication sur "mot".