



Référencement de site Internet

Moteurs & Annuaire : référencement de site Web sur internet

Michel Cabaré – Ver 2.0 – Janv 2008-

<http://WWW.CABARE.NET> ©

Référencement sur Internet

Michel Cabaré– Ver 2.0 – Janv 2008

www.cabare.net ©

www.lallias.com ©

TABLE DES MATIÈRES

POURQUOI SE REFERENCER.....	4
ORIGINE DU TRAFIC SUR UN SITE.....	4
REFERENCEMENT - POSITIONNEMENT	5
ANNUAIRES & MOTEURS.....	6
LES ANNUAIRES:	6
LES MOTEUR DE RECHERCHE.....	7
FONCTIONNEMENT D'UN ROBOT.....	8
GOOGLE COMME EXEMPLE	8
COLLECTE – ANALYSE INDEXATION - RECHERCHE.....	8
COLLECTE SUR GOOGLE.....	9
INDEXATION SUR GOOGLE.....	9
BACKLINKS.....	10
DEFINITION :	10
TEXTE DES LIENS :	10
LIENS ABSOLUS - RELATIFS ET AUTRES:	10
AUTRES LIENS INTERDITS !:	11
LIENS ET FRAMES:	11
LIENS ET FLASH:	11
PAGERANK.....	12
DEFINITION.....	12
EVALUER SON PAGERANK – GOOGLETOLBAR:	12
LE PAGERANK DE LA GOOGLETOLBAR:	14
AUGMENTER SON PAGERANK ?:.....	15
L'EXEMPLE DU GOOGLE BOMBING:	15
RECHERCHE GOOGLE DANS UNE PAGE	16
PLUSIEURS MOTS CLES ET +:.....	16
CASSE ET ACCENTS:.....	16
MOTS DITS "STOP WORD":	16
ENLEVER UN MOT CLE SAUF -:	16
RECHERCHE GOOGLE SPECIFIQUES	17
DANS UN TITRE INTITILE - ALLINTITLE:	17
RECHERCHER DANS UN SITE:	17
RECHERCHER LES BACKLINKS:	17
RECHERCHER DES TYPES DE FICHIERS:	17
LES PAGES VUES PAR UN ROBOT	18
PRINCIPE.....	18
SOUMETTRE LES PAGES (A 'GOOGLE).....	20
GOOGLE VOIT IL VOTRE SITE ? ?	20
COMBIEN DE PAGES DE VOTRE SITE SONT REFERENCEES ?	20
QU'EST-CE UN PLAN DU SITE :	20
SITEMAP DE GOOGLE	21
CREATION D'UN PLAN DE SITE	22



CHOIX DES MOTS CLES	23
QUE TAPE L'INTERNAUTE ?	23
LE POSITIONNEMENT EST IL TECHNIQUEMENT POSSIBLE ?	23
KSE - RAPPORT REQUETE & RESULTATS :	24
ADWORD DE GOOGLE – POUR CHOISIR LES MOTS CLES	25
URL – DOMAINE	27
COMPOSITION D'UNE URL	27
EVOLUTION :	28
SOUS DOMAINES - REDIRECTIONS - DOSSIERS:	28
URL ET REFERENCEMENT	30
TITRE - TITLE.....	31
NATURE DU TITLE HTML	31
BALISES META.....	33
BALISE META GENERALITES.....	33
BALISE META DESCRIPTION.....	34
BALISE META KEYWORDS	34
BALISE META HTTP-EQUIV="CONTENT-LANGUAGE"	34
REVISIT AFTER	34
BALISE META ROBOTS	35
FICHER ROBOTS.TXT	36
LE FICHER ROBOTS.TXT	36
LA STRUCTURE D'UN FICHER ROBOTS.TXT:	36
UN FICHER ROBOTS.TXT A LA RACINE :	37
CORPS DE LA PAGE	39
CONTENU IDEAL :	39
PETITS PLUS ... :	39
OUTIREF :	40
CERTIFICATION DES PAGES W3C.....	41
POURQUOI CERTIFIER UNE PAGE :	41
CERTIFICATION HTML XHTML:	41
CERTIFICATION CSS:	42
LES SITES DYNAMIQUES	43
PROBLEMES DE REFERENCEMENT DES URL :	43
PROBLEMES DES SESSIONS PHP :	43
DES SOLUTIONS COMPLEXES URL REWRITING - :	43
ANNUAIRES IMPORTANTS	44
GOOGLE DIRECTORY – OPEN DIRECTORY PROJECT :	44
QUELQUES ANNUAIRES :	44
GOOGLE ANALITICS	45
PRINCIPE:	45
INSERER LE CODE DE SUIVIT:	46
VERSION DE ANALITICS:	47
RÉSULTAT DE ANALITICS:	47



POURQUOI SE REFERENCER

Origine du trafic sur un site

On peut se demander dans quelles proportions se décompose le trafic sur un site web. Et ce de manière générale.

En effet si sur votre propre site vous pouvez peut être activer une certaine traçabilité via des « fichiers log », il existe des études sérieuses pour évaluer la provenance des accès sur un site web.

Selon le site

<http://barometre.secrets2moteurs.com/>

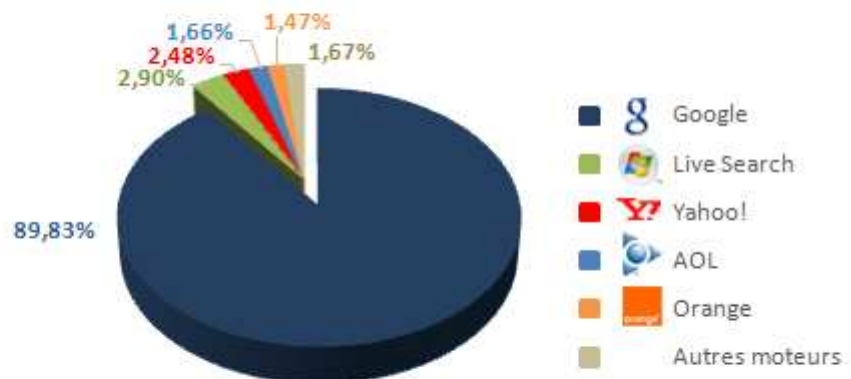
le partage est effectué de la manière suivante :

1/3	Outils de recherche (tous confondus)
1/3	Saisie de l'URL directement « a la main »
1/3	Lien hyper texte « au gré » de la navigation

Si on affine le détail en ce qui concerne les outils de recherche, on tombe alors sur la répartition suivante :

TOP 5 des moteurs de recherche en parts de visites

(France* - Avril 2009)



* Visites générées en France sur les sites audités par une solution AT Internet.



On peut dire que 10 moteurs génériques répondent à la totalité des recherches en France

Parmi ces moteurs, Google se taille une part plus que prépondérante, à tel point que aux Etats-Unis le verbe « to google » existe, pour signifier effectuer une recherche sur le web !!!!

Référencement - Positionnement

Précisons un peu le sens que l'on donne aux termes suivants:

Référencement :

cela indique la présence dans une base annuaires ou moteur de recherche

Indéxé :

Synonyme de référencement

Positionnement :

Faire en sorte que votre URL de site soit bien placée dans les résultats, suite à des saisies de mots clés...

C'est toute la difficulté de l'affaire. Un moteur comme Google indéxant plus de 4 milliards de page !

Mot clé :

La question est de savoir si le (les) mots clés que vous avez prévus sont bien ceux tapés par un internaute, cible potentielle de votre site web.



ANNUAIRES & MOTEURS

D'apparence similaire, ce sont tous les deux des sites Web dotés d'une fenêtre de saisie destinée à recueillir l'objet de votre requête. Les deux outils se distinguent cependant par leur façon de recenser et d'organiser l'imposante somme d'informations disponible sur le Web.

Les Annuaire(s):

Un annuaire est un outil de recherche qui recense un certain nombre de sites au travers de fiches descriptives comprenant, en règle générale :

- le titre,
- l'adresse (l'URL)
- et un bref descriptif d'une longueur allant le plus souvent de 15 à 25 mots au maximum.

Chaque site est inscrit dans une ou plusieurs **catégorie(s)** - on parle également de **rubrique(s)** -. Ces catégories thématiques (finances, sports, tourisme, etc.) vont du plus général au plus précis.

Ces outils peuvent ainsi être considérés comme les pages jaunes du Web.

Par exemple, la rubrique principale "Santé" de Yahoo donne naissance à une trentaine de sous-catégories ("Assurances", "Hôpitaux", "Magazines", etc.) qui elles-mêmes se divisent en plusieurs autres parties et ainsi de suite.

Lorsqu'un **mot-clé est saisi** dans le formulaire proposé, **l'annuaire effectue une recherche sur les occurrences de ce terme dans ses fiches descriptives** de site, et **non pas dans le contenu des pages** du site en question. Il s'agit là de la différence la plus notable avec les moteurs de recherche.

"**Référencer un site**" correspond à la démarche que doit réaliser un concepteur de site pour inscrire un site dans un annuaire (nous verrons plus tard dans un moteur). Le référencement est donc un acte volontaire.

Comprendre le référencement d'un site dans un annuaire permet de comprendre comment rechercher dans cet annuaire.

Les annuaires référencent des sites et plus particulièrement la page d'accueil (Home Page qui souvent s'appelle **index.htm**). Le classement est généralement organisé en catégories et sous catégories. C'est le webmaster qui référence son site en choisissant la catégorie dans laquelle il désire le voir apparaître. Sur le site de l'annuaire, une intervention humaine permettra de vérifier ce choix ainsi que (très rapidement) le contenu.



Les Moteur de Recherche

Le moteur de recherche fonctionne sur un système radicalement différent de celui de l'annuaire. Des **robots** logiciels (appelés crawlers ou spiders) **scrutent le Web**, vont de page en page (en fait de lien en lien) et sauvegardent au fur et à mesure de leurs pérégrinations le contenu texte des pages rencontrées, **constituant ainsi un "index"**, c'est-à-dire une collection plus ou moins grande de pages Web. Pour exemple, AltaVista en stocke 275 millions, HotBot 150 millions, WebCrawler 2 millions

Le **robot** logiciel **repass**e selon des **délais plus ou moins fréquents** sur les pages qu'il a indexées au préalable, pour en sauvegarder une version plus récente. On dit alors qu'il "rafraîchit sa base (ou son index)".

Lorsque l'internaute saisit un mot clé dans le formulaire proposé, le moteur va en rechercher les occurrences dans son index, c'est-à-dire dans le contenu texte des pages Web sauvegardées au préalable. Une fois le "lot" de pages contenant le terme demandé identifié, le moteur classe les pages par ordre de pertinence, selon un ordre et un algorithme (basé sur certains critères de tri) qui lui est spécifique.

Les robots indexeurs agissent donc exactement comme des surfeurs du Web : ils parcourent les pages en notant précieusement chacun des mots contenus et en visitant tour à tour chacune des URLS présentes

Lorsque les robots indexent un site, ils utilisent en priorité les balises META (voir ce chapitre) pour construire le résumé qui apparaît lors d'une recherche sur ce site. Les META tags sont des balises HTML que l'on insère dans les documents et qui indiquent aux robots comment la page doit être indexée (prise en compte du titre, des mots-clés, de la description...).

L'indexation est réalisée en plein texte, ce qui signifie que les robots considèrent tous les mots d'une page web. Autrement dit, lors d'une recherche sur un mot-clé, les robots répondent en fonction de la note qu'ils accordent à chacun des mots du site : cette note tient compte non seulement des balises META, mais aussi de la position des mots dans la page, de leur fréquence, de leur mise en forme (taille, gras, italique...). Mais les titres des pages, les texte alternatifs des images, sont aussi pris en compte par les robots

Attention, parfois les robots n'aiment pas les "ruses" telles que les écritures de même couleur que le fond (de manière à ce qu'une page contienne un nombre élevé d'occurrence d'un mot clé.... sans que cela nuise à la visibilité du site) et considère ceci comme du spamming ! (sorte de publicité abusive) Du coup la page n'est plus du tout référencée ...

Le moteur de recherche effectue donc ses **recherches sur des pages Web**, alors que l'annuaire vous proposera des **sites Web**. Là est toute la différence qui explique qu'il est absolument impossible de comparer les résultats fournis par les deux outils.



FONCTIONNEMENT D'UN ROBOT

Google comme exemple

Nous allons travailler sur google, car celui-ci occupe une place capitale dans les moteurs de recherche utilisés mondialement.

Il existe en plus de 100 langues et nationalités différentes, et beaucoup de ses mécanismes illustrent aussi le fonctionnement des autres moteurs de recherche.

Chaque fois que cela sera nécessaire, lorsque nous présenterons une fonctionnalité, si cette fonctionnalité est particulière à google (comme par exemple la notion de webrank), cela sera précisé.

Autrement dit, tous les mécanismes génériques présentés ici permettent un référencement pour tous les moteurs

Collecte – Analyse Indexation - Recherche

La **collecte (crawl)** c'est le parcours du web de page en page, soit en suivant des liens existant sur le web, soit en acceptant des soumissions volontaires sur le site du moteur.

Ce type de parcours est effectué par un robot

Ces pages collectées sont stockées, copiées dans un espace disque et sont prêtes pour la phase d'analyse.

NB : Il est important donc de voir que l'analyse ne se fait pas « en ligne » mais sur une copie de la page rapatriée localement. Ce point sera étudié plus tard dans le chapitre "les pages vues par un robot" (page n° 18.)

L'**analyse** ou **indexation** peut donc se faire de manière différée de la collecte et consiste à analyser en détails chaque page.

Il s'agit de dresser une liste des mots clés trouvés dans la page, **pondérés par une recette maison**. Chaque moteur a sa recette, jalousement gardée secrète, car c'est la qualité de cette recette qui fait que l'index de la base de donnée ainsi construit est plus ou moins pertinent !

On parle alors d'algorithme d'indexation...

La **recherche** proprement dite, n'est qu'une consultation dans l'index des mots clés tapés par l'internaute



Collecte sur google

La **collecte (crawl)** chez google est effectuée par le robot **googlebot**, ou plutôt de centaines de robots travaillant en simultanée.

Ces pages copiées dans un espace disque correspondent aux « **pages en cache** »

La collecte se faisait historiquement par deux robots différents sur Google

des robots appelés **Full Crawl**: actifs qu'une partie du mois, ils faisaient une analyse complète de toutes les pages

des robots appelés **Fresh Crawl**: actifs tout le temps mais uniquement sur les pages récentes

Depuis 2003, la manière de collecter les pages dans Google est mixte (**Full+Fresh**) plus sophistiquée car et c'est cela qu'il faut retenir la fréquence des visites sur un site dépend des paramètres du site.

N.B: Si googlebot se rends compte qu'une page est souvent mise à jour, il viendra souvent l'indexer, alors qu'une page peu modifiée verra sa fréquence de visite s'allonger notablement.

Indexation sur google

Lorsque l'indexation se recalcule, il est manifeste que le temps de recalcul dure plusieurs journées, au cours desquelles les positions des index peuvent être très changeantes. On nomme cela la **Google Dance**

Chez google l'algorithme d'indexation fait appel à plus de 100 critères, tantôt internes à la page, ce qui est classique, mais aussi externes, ce qui est une spécificité Google !

Pour google, une page n'est pas seulement décrite par les mots qui la composent, mais aussi par les opinions que les autres auteurs de sites internet ont !

- ✓ Les critères internes sont plus facilement modifiables, car on en a le contrôle complet : balises HTML, densités des mots clés, niveaux et mises en forme
- ✓ Les critères externes sont plus difficilement manipulables, car on n'en a pas directement le contrôle : on parlera de **PageRank** et des texte des liens entrants dits **backlinks**



BACKLINKS

Définition :

On parle aussi de **BL** pour les branchés...

Il s'agit tout simplement des liens entrant sur votre site, qu'ils soient internes à votre site (facile à créer), ou externes (plus difficiles à obtenir...).

Par extension, la page HTML complète est souvent aussi appelée backlinks.



Texte des liens :

Le texte du lien est capital pour le positionnement. Il faut bien penser aux mots clés à utiliser pour construire les liens.

Ainsi voici des liens pointant vers le même site mais de qualité très inégale.

Allez sur le site de `<A href="www.xxx.com">formation word</A>`

Favorise la recherche sur les mots clés formation word

Pour une formation word `<A href="www.xxx.com">www.xxx.com</A>`

Favorise la recherche sur les mots clés de l'URL, reste à espérer que cette URL soit pertinente

Pour une formation word cliquez `<A href="www.xxx.com">ici</A>`

Ne Favorise rien, c'est catastrophique !

Liens absolus - relatifs et autres:

Il est tout à fait possible de créer pour des liens internes à un site, deux types de liens :

- ✓ soit des liens en relatifs (juste le nom de la page)
- ✓ soit liens en absolu (URL complète)

Privilégiez les liens en relatifs pour toutes vos pages, mais pour la page index, inscrivez un lien absolu (du genre `http://www.xxx.com`), en effet index ne parle pas du tout en mots clés à un moteur...



Autres liens interdits !:

Aucun autre type de liens n'est susceptible de fonctionner, liste déroulante, menu javascript, redirection, requête de base de donnée...

TOUT CES LIENS SONT SANS EFFET POUR LES MOTEURS DE RECHERCHE !

Certain types de liens comme des liens sur des images, sont nettement moins bons car on perd le sens du mot clé dans le lien..

Comment pouvoir faire alors si vous voulez utiliser ces techniques ?

Et bien il va falloir ruser et prévoir une solution alternative, destinée uniquement aux moteurs de recherche, ces techniques consistent à :

- Doubler systématiquement les liens images en lien texte
- Prévoir une zone "plan du site" qui en fait sert souvent à avoir une navigation texte complète dans le site...
- Dans le cas de javascript aucun lien ne sera suivi, il faut donc recourir à une des techniques précédentes

Liens et frames:

Les **frames** ou **cadres** sont une technique de découpe de page en plusieurs fenêtres. Beaucoup utilisées un temps, cette technique est tombée en désuétude car elle constitue un obstacle majeur pour tous les moteurs de recherche.

De plus souvent, dans le cas d'indexation de la première page on obtient une page seule, sans les autres morceaux (autres frames) ce qui rends le résultat peu lisible...

Si cependant vous souhaitez persister, il faut utiliser alors la partie balise

<NOFRAMES>

dans laquelle on va mettre des liens textes avec des mots clés importants...

</NOFRAMES>

Liens et flash:

L'indexation du texte présent dans une incrustation flash est plus qu'aléatoire... mais le fichier HTML qui lance l'animation lui est pris en compte... il comporte alors une section

<NOEMBED>

dans laquelle on va mettre des liens textes avec des mots clés importants...

</NOEMBED>



PAGERANK

Définition

Le **Pagerank** s'agit d'une grandeur chiffrée représentant la popularité d'une page sur internet.

Cette grandeur ne tient absolument pas compte de la constitution physique de la page, mais uniquement de l'opinion, au travers des liens, que les autres sites web peuvent avoir sur votre page.

Pour comprendre le fonctionnement de l'indice pagerank, on peut parler de vote, avec deux idées de base :

1 lien = 1 vote

1 lien depuis un site avec un fort pagerank = 1 vote plus important...

Il est inutile de rentrer dans le détails de comment l'indice est calculé, ce qui serai d'ailleurs quasi impossible dans l'absolu car il s'agit d'une partie importante de la recette de l'algorithme d'indexation google, et donc top-secret !

N.B: Hotbot, Lycos utilisent aussi un critère similaire....

Evaluer son Pagerank – googletoolbar:

Il devient alors intéressant de pouvoir évaluer le pagerank d'une page, afin de mieux comprendre son degré de pertinence dans google

Le pagerank n'est pas donné en standard dans l'interrogation de google (à la différence de la page en cache, ou autres informations)

Pour connaître le pagerank d'une page il est nécessaire d'installer une barre d'outils spécifique, nommée **googletoolbar**.

Nom	Taille
 Google Updater.exe	1 228 Ko

Téléchargeable chez google

Optimiser votre ordinateur



Pack

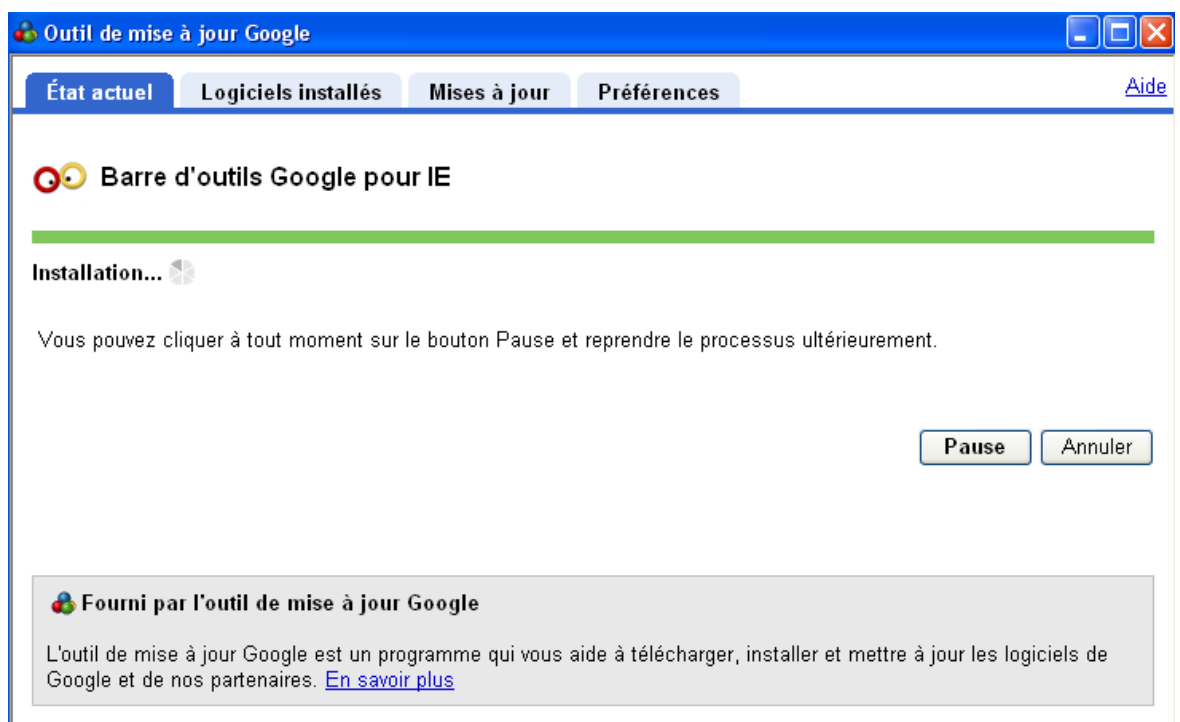
Une suite gratuite de logiciels essentiels



<http://toolbar.google.fr>

genre "google toolbar macintosh"

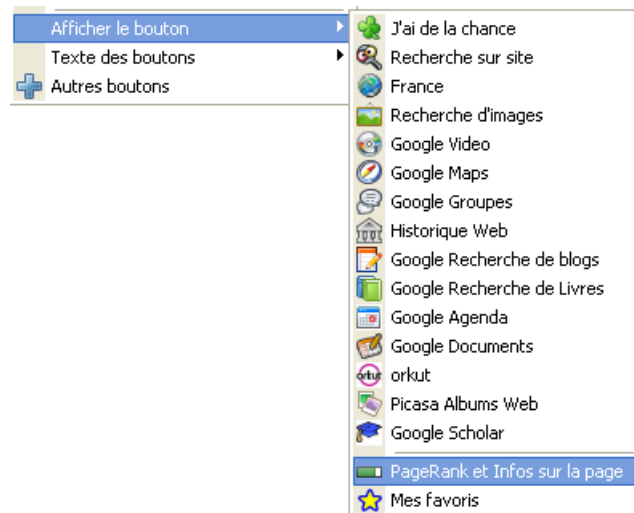
Cette google tool bar s'installe très facilement



et la googletoolbar est installée



- Pour avoir le pagerank il faut activer les fonctions avancées :



le Pagerank de la googletoolbar:

Cette barre d'outils contient plusieurs fonctions parmi lesquelles la possibilité de bloquer des pop-up, et autres...

Nous utiliserons cette barre uniquement à des fins d'information sur le **pagerank** d'une page, et par conséquent seule cette information sera exploitée



la valeur du pagerank débute à **0** et culmine à **10** selon une échelle logarithmique. Plus le pagerank est élevé, plus la page est "populaire" !

N.B: il s'agit ici d'un pagerank affiché, légèrement différent de celui réel...

Mais pour un meilleurs positionnement il vaut mieux avoir un backlink avec des mots clés judicieux, depuis une page ayant un fort pagerank, qu'avoir un pagerank élevé tout court....

N.B: Le pagerank d'une page n' est pas un critère prépondérant pour le positionnement d'une page, c'est un ingrédient important, mais pas le plus important.

N.B: Par contre le pagerank est primordial pour accorder une importance aux backlinks...

Augmenter son pagerank ?:

En vrac voici plusieurs éléments permettant de faire évoluer le pagerank d'une page :

- ✓ Placer des liens depuis vos autres sites (si vous en avez)
- ✓ Inscrivez vos sites dans des annuaires généralistes, thématiques, régionaux... (voir en annexe...)
- ✓ Demandez à vos partenaires de placer des liens sur leur sites
- ✓ Créer du contenu attractifs pour que d'autres sites créés des liens vers vous
- ✓ On peut acheter des campagnes publicitaires, fournissant des liens...
- ✓ Soignez vos liens internes, si votre page index a un bon pagerank, les pages sur lesquelles elle pointe en bénéficieront !
- ✓

L'exemple du Google bombing:

A partir du moment où l'on prends une expression générant relativement peu de résultats dans Google, maximum 10000 à 15000, et uniquement dans ce cas, on peut essayer de jouer sur la faculté de pouvoir placer une page par son pagerank plutôt que par son contenu...

Le **Google Bombing** c'est en effet essayer de renvoyer sur une page existante par une expression très clairement peut en rapport avec la page en question...

Par exemple : "miserable-failure", "go to hell", "site le plus crétin" menaient à un moment donné vers une page parlant de Jacques Chirac, Georges Bush ou TF1 etc...

Les limites de cette action sont doubles :

- Encore une fois ces actions ne peuvent être entreprises que sur des requêtes spécifiques et ne contenant pas de termes génériques.
- Leur durée dans le temps est forcément limitée, car dès que le google bombing fonctionne, si de nouvelles pages relatant le fait apparaissent, elles vont devancer dans le positionnement de Google la page "visée", annulant de fait l'effet google bombing

Mais ce phénomène illustre bien l'importance que peut prendre le pagerank.



RECHERCHE GOOGLE DANS UNE PAGE

Plusieurs mots clés ET +:

Inutile de revenir sur la recherche de base dans un robot, mais spécifions quelques particularités pour la recherche avancée.

Par défaut la recherche s'effectue sur la contenu complet de la page.

Pour effectuer une recherche sur différents termes, dans google on sait que l'espace fonctionne comme un **ET** (majuscule):

Formation Word

équivalent à

Formation ET Word

Casse et accents:

Par défaut, Google ne tient pas compte de la casse (minuscule – majuscule) et des accents

Ansi toutes ces recherches sont en théorie identiques:

Réseau = RESEAU = Reseau

Pour effectuer une recherche en excluant les versions non accentuées ou dans une casse particulière, on peut faire précéder le mot clés d'un **+**

+Réseau (exclus reseau et Reseau)

+RESEAU (exclus reseau et Réseau)

Mots dits "stop word":

Par défaut, Google ne tient pas compte dans la requête des particules DU, LE, LES, DES ...

Si on veut inclure ces termes dans une requête on peut encore une fois jouer avec le signe +

+la formation réseau (exclus les formations réseaux)

Enlever un mot clé SAUF -:

Permet d'éviter les contresens ou affine la recherche.

labrador -chien

formation réseau -initiation



RECHERCHE GOOGLE SPECIFIQUES

Dans un titre **intitle** - **allintitle**:

Comme le titre est logiquement descriptif de la page (si son auteur a bien travaillé) l'opérateur **intitle** peut restreindre la recherche aux titres HTML uniquement :

Intitle:reseau

On peut effectuer une recherche plus complexe avec la balise **allintitle**:

Intitle:formation word

Intitle:+réseau windows

Rechercher dans un site:

Pour n'effectuer la recherche uniquement dans un site donné, l'opérateur **site:url** est utilisable :

reseau site:www.microsoft.com

Rechercher les backlinks:

Pour effectuer la recherche des backlinks d'une URL donnée, l'opérateur **link:url** est utilisable :

Link:www.microsoft.com

Identique à

Link:http://www.microsoft.com

Rechercher des types de fichiers:

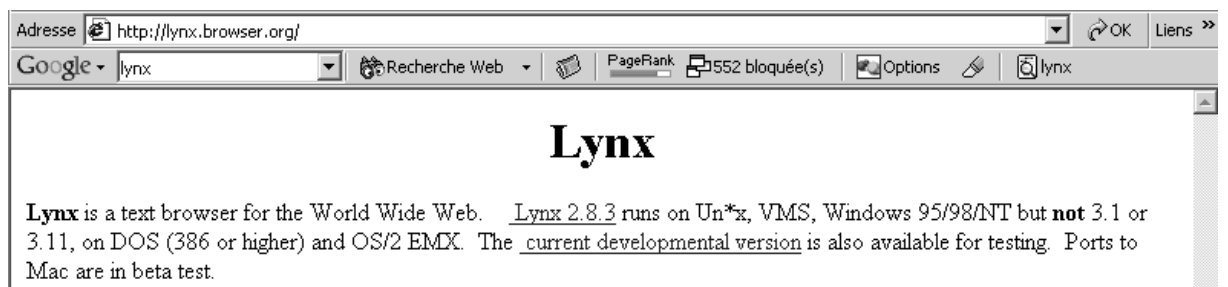


LES PAGES VUES PAR UN ROBOT

Principe

Il est important de vérifier comment le site est vu par un moteur de recherche.... On peut utiliser certains navigateurs spécifiques, permettant de "simuler" la vision texte qu'un robot peut avoir lors du parcours de votre site, tels que

Lynx



Mais le plus simple est de donner votre URL comme adresse de départ sur l'outil **spider simulator**...

<http://www.spider-simulator.com/fr>

A screenshot of the Spider-Simulator website. The browser address bar shows 'http://www.spider-simulator.com/fr/'. The page has a search bar with 'spider simulator' entered. Below the search bar, there are links for 'Autre outils gratuits', 'Déjà inscrit?', 'utilisateur', 'Mémoriser', and 'Mot de pas'. A navigation bar contains links: 'Recommander ce site', 'Partenariat', 'Liens vers ce site', 'Notre outil sur votre site', 'Help', and 'Bug Rep'. The main heading is 'Spider-Simulator' with a tagline 'Spider Simulator : explorez votre site avec l'oeil du spider'. A numbered instruction '1' says 'Saisissez l'adresse de votre site web et regardez-le avec "l'oeil d'un spider" ou d'un navigateur selon la liste proposée ci-dessous :'. Below this is a large text input field. Underneath the input field are radio buttons for 'Internet Explorer', 'Opera', 'YahooBot', 'Mozilla', 'GoogleBot', and 'MSH Bot'. A text prompt says 'Merci de saisir les caractères présents dans l'image :'. Below this is a CAPTCHA image showing the letters 'E S M Y' and a 'Rechercher' button.

On peut ainsi s'apercevoir des petites erreurs commises, voire des bizarreries d'interprétations...

En tout cas il est clairement acquis que entre voir et être référencé, il peut y avoir une certaine ... différence !

Ainsi sur un site entièrement réalisé en Flash...

<http://www.letoboggan.com>



L'analyse via **spider simulator** est assez intéressante

Et donne un résultat qui est sans appel !

En-tête de page

Images

URLs internes

URLs externes

Contenu textuel

Contenu mis en page

Minimal informations

URL: <http://www.letoboggan.com/>

Type: **text/html; charset=utf-8**

HTTP code: **200**

Temps de traitement: **1 s**

URL size: **2 kB**

[passer l'intro](#)

SOUMETTRE LES PAGES (A 'GOOGLE)

Google voit il votre site ? ?

Vous voulez savoir si google a vu votre site avec votre domaine <http://www.domaine.fr> .

Sur la page d'accueil de google on peut demander

A propos

Infos pour webmaster



Ce centre regroupe toutes les informations sur la façon dont Google explore et indexe les sites Web. Apprenez à créer un site que Google est en mesure d'explorer et d'indexer facilement. Vous trouverez également des outils conçus pour identifier les problèmes liés à l'exploration et pour consulter les statistiques relatives à l'indexation de votre site. Enfin, vous pourrez nous faire part de vos souhaits en matière d'exploration et d'indexation des sites.

dans lequel on trouve



[Assistant d'état du site](#)

Déterminez si Google procède actuellement à l'indexation de votre site.

Combien de pages de votre site sont référencées ?

Vous avez référencé un site avec votre domaine <http://www.domaine.fr> .

Comme un moteur n'indexe pas toutes les pages d'un site, on peut évaluer l'étendue de la récupération effectuée en utilisant l'opérateur **site:**

↑ **site:www.domaine.fr**

comme aucun mot clé n'est précisé, toutes les pages seront ramenées...

Qu'est-ce un plan du site :

Un plan du site est une page présentant sous forme d'arborescence de liens texte la navigation logique possible dans un site web....

Longtemps le plan du site à été considéré par les "puristes" comme un palliatif au manque de lisibilité dans la charte graphique du site. On pouvait dire alors que un site bien fait se passait de "plan"....

Les plans sont apparus très vite cependant comme un moyen de sauver un site dans lequel la navigation ne se ferait que par image, liens javascripts, ou dans des sites sur lesquelles par exemple encore on travaille par technique



externes, comme FLASH ou similaires, peut référencables et utilisable par les moteurs de recherche.

MAIS même pour un site fort bien construit, la présence d'un plan du site permet au moteur de "se rendre compte de la pertinence des liens" et de voir la cohérence et la structure logique d'un site....

Sitemap de Google

Sur la page d'accueil de google on peut demander

A propos

Infos pour webmaster



Outils pour les webmasters (notamment Google Sitemaps)

Statistiques, diagnostics et gestion de l'exploration et de l'indexation par Google de votre site Web (avec accès aux fonctionnalités d'envoi de plans Sitemap et de création de rapports Sitemap).

Il faut absolument un compte google



Outils pour les webmasters

Les outils Google pour les webmasters vous permettent de découvrir gratuitement et simplement comment rendre votre site plus facilement explorable par Google. Grâce à nos outils, vous pouvez :

Découvrez le point de vue de Google sur votre site et diagnostiquez les problèmes potentiels. Découvrez comment Google explore et indexe votre site ainsi que les problèmes spécifiques que nous rencontrons en y accédant.

Découvrez les performances de votre site. Découvrez les requêtes qui génèrent du trafic vers votre site et la provenance des visiteurs.

Connectez-vous au service Google Outils pour webmasters à l'aide de votre

Compte Google

E-mail:

Mot de passe :

☒ Mémoriser mes informations sur cet ordinateur.

Tableau de bord

Sites

1 - 5 sur 5

Cliquez ici pour ajouter un site.

Ajouter un site

[?]

Gérer ▲

Sitemap

Vérifié ?

☐ www.cabare.biz

Ajouter

Vérifier

☐ www.cabare.net

1

Vérifier

☐ www.cabare.org

1

Vérifier

☐ www.cabare-formation.net

1

✓

Vue d'ensemble

Diagnostic

Statistiques

Liens

Plans Sitemap

Vue d'ensemble

Ajout d'un plan Sitemap

Outils

Plans Sitemap

www.cabare-formation.net ▼

Statistiques Sitemap :
Nombre total d'URL : 28
URL indexées : 25

Ajouter un plan Sitemap

1 - 1 of 1

Nom de fichier	Format	Type	Date d'envoi	Dernier téléchargement	État	Nombre d'URL fournies
☐ sitemap.xml	Sitemap	Web	20 oct. 2006	31 déc. 2007	OK	28

Détails

Supprimer la sélection

Envoyer de nouveau la sélection



Création d'un plan de site



Sur lequel

Please enter details for sitemap generation

Starting URL

Please enter the **full** http address for your site, only the links within the starting directory will be included. For instance, "domain.com" and "www.domain.com" are **not** the same.

Change frequency

None

Last modification

☐ None

☒ Use server's response

☐ Use this date/time:

Priority

None

We have automatic priority calculation implemented for online sitemap generator now!

Check your settings and click button below

Start

Maximum 500 pages will be indexed in sitemap

Pour obtenir

Initial website address

<http://www.cabare.net/>

Download un-compressed XML Sitemap

<http://www.xml-sitemaps.com/download/www.cabare.net/sitemap.xml> (44.89Kb)

Download compressed XML Sitemap

<http://www.xml-sitemaps.com/download/www.cabare.net/sitemap.xml.gz> (4.03Kb)

Download ROR Sitemap

<http://www.xml-sitemaps.com/download/www.cabare.net/ror.xml> (104.29Kb)

Download HTML Sitemap

<http://www.xml-sitemaps.com/download/www.cabare.net/sitemap.html> (59.64Kb)
[\[view\]](#)

Download Sitemap in Text Format

<http://www.xml-sitemaps.com/download/www.cabare.net/urllist.txt> (22.86Kb)

XML Sitemap details

Request date:

29 January 2010, 14:36

Processing time:

51.90s

Pages indexed:

231

Pages size:

3.22Mb



Do you like it?

We'll be happy to receive **your** [testimonial](#) for our service.



CHOIX DES MOTS CLES

Que tape l'internaute ?

On l'a déjà dit, mais il est important de le rappeler. Rien ne sert de référencer son site web comme il faut, avec toutes les difficultés que cela comporte, si ce mot clé n'est pas celui utilisé réellement par l'internaute lors d'une recherche.

Il existe un générateur de mots clés de la société Overture, a l'adresse suivante :

<http://inventory.fr.overture.com>

Les statistiques fournies concernent les moteurs affiliés, (Yahoo + MSN + AOL + Wanadoo-Voilà) uniquement pour le mois précédant.

Attention donc, si vous travailler avec des mots clés saisonniers, au moment ou vous faites votre étude...

Pour estimer la demande totale sur un mot clé en France, il suffit de multiplier environ par 4 le resultat (puisque google absent du regroupement, représente environ 75% du trafic sur les recherches par moterus en France)

Le Positionnement est il techniquement possible ?

La question peut se poser...

En effet si votre mot clé est trop généraliste, par exemple "voyage", il sera de toutes façon noyé dans la masse des retours...

Voici quelques fourchettes indicatives :

- Jusqu'à 5.000: Google Bombing possible !
- Jusqu'à 20.000: bon positionnement possible
- De 20.000 à 100.000: positionnement possible mais très délicat
- Au delà de 100.000: résultat très aléatoires



KSE - Rapport requête & résultats :

Pratiquement, il faut arriver à savoir si le jeu en vaut la chandelle, à savoir:

1. si les mots clés auxquels vous pensez, sont réellement utilisés sur le web
2. ensuite si les réponses obtenues actuellement ne sont pas trop nombreuses !

Sur l'URL suivante : <http://www.keyword-search-engine.com> on peut se livrer à l'exercice suivant:



Par rapport à une requête : ici "formation internet"

Un positionnement sur les moteurs de recherche est-il possible pour vos mots clés favoris ?

Pour le savoir, tapez un mot clé :

☒ Web français ☐ Web mondial

English version

on obtient un rapport permettant d'évaluer rapidement le bon critère choisi, par rapport au nombre de réponses retournées...



Résultat de l'audit de positionnement pour le mot clé formation internet :

Nombre de fois où la requête formation internet a été saisie le mois dernier sur le Web (estimation) :

8 716

[plus d'infos](#)



Notre diagnostic (intérêt du positionnement) : Le nombre de requêtes demandées sur le Web pour ce mot clé est important. Un positionnement sur ce terme doit être envisagé.

Nota : Ce diagnostic ne s'applique qu'à la partie "intérêt du positionnement"

Nombre de résultats identifiés par Google pour la requête formation internet :

9 000 000

[plus d'infos](#)



Notre diagnostic (faisabilité du positionnement) : Un positionnement sur ce terme sur un moteur de recherche sera extrêmement complexe, voire impossible. Recherchez des mots clés qui génèrent moins de résultats sur Google.

Nota : Ce diagnostic ne s'applique qu'à la partie "faisabilité du positionnement"

Intérêt et faisabilité du positionnement d'une page web pour la requête formation internet sur les moteurs de recherche :



(2/20)

[plus d'infos](#)



AdWord de Google – pour choisir les mots clés

Google a mis en place un système de référencement payant baptisé programme **Adword**, Sur l'adresse suivante

Sur <https://adwords.google.fr/select/Login>

La page de Adwords de google

ce n'est pas ce système qui nous intéresse, mais son outils de choix de mots clé...

Qui devrait amener à

<https://adwords.google.fr/select/KeywordToolExternal?defaultView=2>

Générateur de mots clés

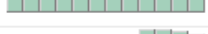
Utilisez le Générateur de mots clés pour obtenir de nouvelles idées de mots clés. Sélectionnez une des deux options proposées ci-dessous selon que vous souhaitez entrer des mots clés ou bien l'adresse URL de votre site.
[Conseils d'utilisation du Générateur de mots clés](#)

Important : Le 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. Les termes énumérés ne vous sont pas conseillés ni recommandés et Google ne peut garantir que ces mots clés amélioreront les performances de votre campagne. Nous nous réservons également le droit de désapprouver tout nouveau mot clé sélectionné. Vous êtes seul responsable des mots clés que vous sélectionnez et devez vous assurer que leur utilisation n'enfreint aucune législation ni ne porte atteinte aux droits de tiers, notamment au regard du droit des marques et de la concurrence déloyale. 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.icimarkes.com) et des sociétés (ex: www.euridile.com).

Les résultats sont adaptés au paramètre suivant : **Français, France** [Modifier](#)

Comment souhaitez-vous générer des idées de mots clés ?		Entrez un mot clé ou une expression par ligne :
☒ Expressions ou termes descriptifs (exemple : thé vert)		
☐ Contenu de site Web (exemple : www.exemple.fr/produit?id=74893)		
		☒ Utiliser des synonymes
		Entrez les caractères figurant dans l'image ci-dessous : 

Qui nous donne ensuite sur Formation Internet

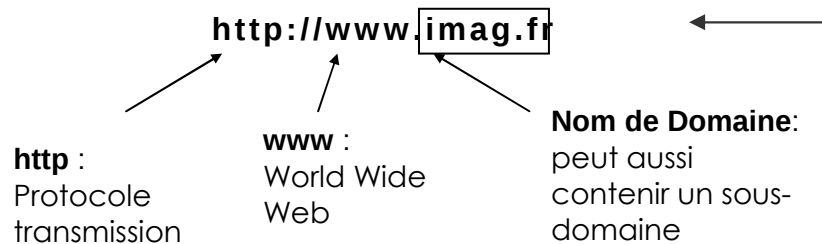
Mots clés	Volume de recherche mensuel global ?	Tendances des volumes de recherche (janv. - déc. 2009) ?	Volume le plus élevé observé en	Type de ciblage : ? Large
Mots clés en rapport avec le(s) terme(s) entré(s) - trié par pertinence ?				
formation professionnelle internet	91	Aucune donnée	Aucune donnée	Ajouter ⌵
formations internet	390		avr.	Ajouter ⌵
cours formation internet	4 400		sept.	Ajouter ⌵
formation recherche internet	110	Aucune donnée	Aucune donnée	Ajouter ⌵
formation pour internet	260	Aucune donnée	Aucune donnée	Ajouter ⌵
formation informatique internet	4 400		sept.	Ajouter ⌵
formation à internet	320	Aucune donnée	Aucune donnée	Ajouter ⌵
formation internet senior	36	Aucune donnée	Aucune donnée	Ajouter ⌵
formation site internet	3 600		nov.	Ajouter ⌵
formation anglais internet	170		sept.	Ajouter ⌵
formation marketing internet	390		nov.	Ajouter ⌵
formation internet	18 100		sept.	Ajouter ⌵
formation continue internet	170		nov.	Ajouter ⌵
formation à distance internet	73	Aucune donnée	Aucune donnée	Ajouter ⌵

...

URL – DOMAINE

Composition d'une URL

Uniform Resource Locator (Localisateur Universel de Ressources). Ce type d'adresse fait appel à l'alphabet pour rendre la mémorisation plus facile. A la différence d'une adresse IP, une URL se lit de droite à gauche



Un nom de domaine (imag.fr) se décompose en

- ▣ Un Top Level Domain (exemple : fr)
- ▣ Un nom d'organisation (appelé aussi nom de domaine) (ex : imag)

Les 6 Top Level Domain de base sont:

Clé	Contenu
.com	Entreprise commerciale
.edu	éducation
.gov	organismes gouvernementaux
.mil	organisations militaires
.net	intervenant d'internet
.org	instance gouvernementale

Cependant si ces domaines sont a priori internationaux, ils sont à forte dominante américaine.

De plus chaque pays possède son nom de domaine (à l'exception des USA qui utilisent les 6 domaines précédents).

Clé	Contenu
.au	Australie
.ca	Canada
.de	Allemagne
.fr	France
.uk	United Kingdom

L'internic se chargeant de l'attribution des adresses dans les domaines internationaux, c'est le NIC France qui se charge des attributions des noms de domaine en .fr <http://www.nic.fr>



Evolution :

Le succès d'Internet rend chaque jour le travail des organisations chargées d'attribuer une adresse de site de plus en plus difficile.

A la suite d'une réunion tenue à GENEVE fin Avril 97, le monde a été divisé en 7 régions, qui accueilleront chacune 4 organisations pour distribuer les adresses de sites, et on pourra faire jouer la concurrence entre elles (à terme ces organisations devraient se multiplier, même si on ne sait pas encore comment réellement ces organisations seront choisies)

N.B: Le texte de ce mémorandum est consultable à <http://www.itu.int>

Du même coup, les « org » « .net » « .fr » ou « .com » vont être épaulés par 7 nouvelles clés dont le détail suit:

Clé	Contenu
.firm	Site à vocation d'affaires et de relations inter-entreprises
.store	Site à vocation de commerce électronique
.web	Site d'organisations se contentant d'activités ayant trait au WEB
.arts	Site à vocation culturelle
.rec	Site spécialisé dans le divertissement
.nom	Site personnel
.info	Site spécialisé dans l'information « ON LINE »

De toutes nouvelles extensions de noms de domaines tel que .PRO, .INFO, .BIZ, .NAME, .MUSEUM, .AREO, et .COOP sont également disponibles!

Sous domaines - redirections - dossiers:

Il y a beaucoup de moyens de modifier une URL, les techniques sont diverses et les résultats pas toujours heureux... en tout cas au niveau référencement, certaines de ces techniques pourront vous jouer un sacré tour !

Tout dépend en effet de comment on a obtenu ces URL, car les moteurs éjectent quasi systématiquement les redirections d'adresses par HTML ou Javascript, et de manière générale par quelque technique que ce soit.

Ceci afin d'éviter la présence d'URL de « complaisance » ne correspondant pas à un site réel...



1. Nom de domaine propre ou « classique »

www.domaine.fr/index.html

exemple : www.microsoft.com

N.B : c'est le dépôt de nom de domaine classique

2. Nom de sous domaine principal DNS suivant les règles de **internet**

www.domaine.sousdomaine.fr/index.html

exemple : www.microsoft.france.com

exemple : www.microsoft.italie.com

N.B : ce dépôt de sous domaine ne peut se faire que avec l'accord du domaine principal, ici microsoft. La mise en œuvre ne peut se faire que si on gère la zone d'autorité en DNS au niveau informatique, ce qui limite l'accès aux entreprise dotées d'un certain savoir faire.

3. Nom de sous domaine avec redirection gérée en **interne**

sousdomaine.domaine/index.html

exemple : france.microsoft.com

exemple : italie.microsoft.com

N.B : ce dépôt de sous domaine n'en est pas vraiment un, il n'y a pas de dépôt auprès de l'afnic. La mise en œuvre peut se faire à travers les services de l'hébergeur, qui proposent des interfaces pour ses clients afin de créer de telles URL.

ATTENTION : selon la technique utilisée par l'hébergeur, il faut savoir que les URL ainsi créés peuvent ne pas être référencables. Il faut éviter les URL créées avec les techniques suivantes :

- ✓ Redirection META HTML
- ✓ Frames (à moins de travailler dans la zone NOFRAME)
- ✓ Redirection Javascript

Si l'URL est créée via un enregistrement DNS et via une entête différenciée au niveau du serveur Web, alors le référencement est possible.

4. Nom de sous domaine avec dossier et sous dossiers

www.domaine/dossier/index.html

www.domaine/dossier/sousdossier/index.html

exemple : www.microsoft.com/windows/index.html

exemple : www.microsoft.com/windows/server/index.html



N.B : c'est la création d'un dossier et de sous dossiers dans votre arborescence de site web qui va créer cette URL

URL et Référencement

Plus l'URL correspond à un document proche de la racine du site (pas de sous-sous-sous-répertoire), plus votre référencement sera efficace.

En d'autre termes, le principe de base est que l'importance des mots clés **décroît de gauche à droite** dans l'URL.

Mais ce principe est concurrencé par le fait que **l'ordre des mots** dans l'URL a aussi son importance.

Voici quelques exemples confondus :

<http://www.formation.com>

<http://www.formation.word.com> vrai sous domaine

<http://www.formation.word.com/initiation>

<http://initiation.formation.word.com>

<http://www.formation-word.com> non de domaine avec -

<http://www.formation-word.com/initiation> dossier

<http://initiation.formation-word.com> sous domaine « interne »

Voici encore quelques points :

- On préconise en général moins de 70 caractères pour l'adresse d'une page, certains moteurs n'indexant pas les documents dont l'url serait trop longue
- On peut utiliser des sous domaines, ou des URL différentes, mais pas avec des pages identiques... Comme certains moteurs (dont Google) limitent le nombre de page par URL (cela s'appelle le **Clustering** chez Google), les sous_domaines augmentent notre visibilité sur internet...(puisque l' peut avoir ainsi deux pages par sous domaine...)
- Proscrire les URL contenant des caractères autres que ceux utilisables pour les noms de dossier et de fichiers.



TITRE - TITLE

Nature du Title HTML

Syntaxe :

`<TITLE>Titre de votre site</TITLE>`

Compatibilité : Tous moteurs

Version : HTML 2.0

Dans le code HTML, placez le titre le plus haut possible (le plus proche possible du `<HEAD>`).

→ Le titre de vos pages est certainement le critère de pertinence le plus important pour les moteurs de recherche à l'heure actuelle. Soignez-le bien !

En règle générale, ne jamais dépasser les **100 caractères** (une lettre accentuée valant pour un caractère) ou une **dizaine de mots** pour cette zone

- Attention à ne pas répéter 2 fois le même mot dans le titre et en espaçant le plus possible ces occurrences .
- Ne proposez pas dans vos pages de titre entièrement en majuscules.
- Coder les lettres accentuées en HTML (´ = é).
- Chaque page de votre site doit proposer un titre différent et représentatif du contenu qui y est proposé.
- Evitez les répétitions de mots séparés par une virgule : proposez une vraie phrase.
- Ne commencez pas votre titre par un espace.
- Si les minuscules et les majuscules n'ont que peu d'importance (écrivez le titre de la manière la plus naturelle possible) le singuliers et le pluriels eux, comptent...
- Plus on navigue dans le site , plus on affine l'arborescence, et plus le titre peut être précis et contenir des termes techniques spécifiques.

NB: se rappeler de la maxime suivante : une page = un titre = une langue.

NB: Il vaut mieux plusieurs petites pages proposant un sujet précis, avec des titres précis, que plusieurs sujets traités dans une grande page au titre forcément plus généraliste.





BALISES META

Balise META Généralités

Les deux balises <META> ("description" et "keywords") doivent être placées après la balise <TITLE>...</TITLE> et avant la balise de fin d'en-tête (</HEAD>).

Un code HTML "propre" par rapport aux moteurs de recherche commencera donc toujours ainsi :

```
<HTML>
<HEAD>
<TITLE>Titre de la page</TITLE>
<META NAME="description" CONTENT="...">
<META NAME="keywords" CONTENT="..., ...">
```

Si d'autres d'autres balises <META> viennent s'ajouter à votre code HTML, positionnez-les après la balise <META NAME="keywords">.

N.B: les balises "Meta Keywords" ont aujourd'hui beaucoup moins d'importance dans le cadre d'un référencement de site web. Il faut cependant bien suivre les points suivants :

- Coder les lettres accentuées en HTML (é = é).
- Astuce : prenez comme base le titre de votre page et développez-le sur 200 caractères (au maximum). Comme pour le titre de vos pages, faites une vraie phrase, ne proposez pas une suite de mots séparés par une virgule.
- Essayez de proposer plusieurs orthographe (casse, accentuation, singuliers, pluriels, masculins, féminins, etc.) pour vos mots clés : référencement, référencement, REFERENCEMENT, etc.
- L'ordre des mots clés est important : les premiers ont un poids plus fort pour les moteurs de recherche que les derniers.
- Les balises "Meta Description" et "Meta Keywords" doivent normalement être différentes dans chaque page. Cependant, au vu de leur importance décroissante, vous pouvez découper votre site en grandes "Zones" (Produits, Présentation de la société, Actualité...) et, à l'intérieur de ces zones, dupliquer la même balise "Meta Description" sur chacune des pages.



Balise META Description

Syntaxe :

<META NAME="Description" CONTENT="Description de site...">

Compatibilité : Tous moteurs

Version : HTML 2.0

Une balise "**Meta Description**" doit comporter entre **150 et 200** caractères, espaces compris.

Phrase prise en compte pas le moteur de recherche comme description de votre site.

Les mots qui constituent cette phrase sont aussi considérés comme des mots clés par la plupart des moteurs.

Balise META Keywords

Syntaxe :

<META NAME="Keywords" CONTENT="Vos mots clés ici">

Compatibilité : Tous moteurs

Version : HTML 2.0

Une balise "**Meta Keywords**" ne doit pas comporter plus de **100 mots ou 1000 caractères**, la première limite atteinte étant la bonne

Indique aux moteurs de recherche les mots clés en rapport avec votre site.

Ne répétez pas plusieurs fois de suite le même mot clé.

Balise META http-equiv="Content-Language"

Syntaxe :

<META http-equiv="Content-Language" content="fr">

Compatibilité : Tous moteurs

Version : HTML 2.0

NB: se rappeler de la maxime suivante : une page = un titre = une langue.

Revisit After

Syntaxe :

<META NAME="Revisit-after" CONTENT="nombre">

Compatibilité : Certains moteurs

Version : HTML 2.0

Permet de spécifier au spider (robot du moteur) de réindexer votre site suivant l'intervalle en jours donnée. Evidemment cela reste indicatif.... Le seul à décider de la fréquence de ses visites, c'est le robot lui-même.

Exemple : **<META NAME="Revisit-after" CONTENT="15 day">**



Balise META robots

Syntaxe :

<META NAME="Robots" CONTENT="index, follow">

Compatibilité : Certains moteurs

Version : HTML 2.0

Permet d'indiquer aux moteurs la manière d'indexer le site. En ce sens elle est en double emploi avec le fichier que l'on peut poser à la racine du site nommé robots.txt (cf chapitre suivant)

Dans laquelle X sera au choix index - ou - noindex , et dans laquelle Y sera au choix follow - ou - nofollow

index : le robot indexe la page

noindex : le robot n'indexe pas la page

follow : le robot suit tous les liens qu'il trouvera sur la page.

nofollow : le robot ne suit aucun lien présent dans la page.

On peut donc combiner toutes les possibilités :

`<meta name="robots" content="index, follow">`

`<meta name="robots" content="index, nofollow">`

`<meta name="robots" content="noindex, follow">`

`<meta name="robots" content="noindex, nofollow">`

On peut remplacer noindex, nofollow par none

On peut remplacer index, follow par all

On peut dire que

<meta name="robots" content="index, follow">

équivalent à **<meta name="robots" content="all">**

<meta name="robots" content="noindex, nofollow">

équivalent à **<meta name="robots" content="none">**

FICHER ROBOTS.TXT

Le fichier robots.txt

Un dossier privé ou une page en construction, par exemple, ne doivent pas être référencé. Il faut alors empêcher certains spiders de les prendre en compte.

Cela se fait au moyen d'un fichier texte, appelé **robots.txt**, présent sur votre serveur. Dès que le spider d'un moteur arrive sur un site (par exemple, <http://www.monsite.com/>), il va rechercher le document présent à l'adresse <http://www.monsite.com/robots.txt> avant d'effectuer la moindre "aspiration de document".

- Si ce fichier existe, il le lit et suit les indications qui y sont inscrites.
- S'il ne le trouve pas, il commence son travail de lecture et de sauvegarde de la page HTML qu'il est venu visiter, considérant qu'a priori rien ne lui est interdit.

Ce fichier texte nommé impérativement

robots.txt

(non obligatoire) mais contient des indications à l'attention des moteurs de recherche. Une syntaxe classique serait la suivante :

User-agent: *

Disallow: /cgi-bin (n'indexe pas le contenu du dossier cgi/bin)

Disallow: /prive (n'indexe pas le contenu du dossier prive)

N.B: Il ne peut exister qu'un seul fichier robots.txt sur un site, et il doit se trouver au niveau de la racine, nommé en minuscule.

N.B: Il n'y a pas d'équivalent à Disallow (non autorisé) qui serait en anglais Allow. Tout ce qui n'est pas marqué Disallow est considéré comme autorisé

La structure d'un fichier robots.txt:

Par exemple

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /tempo/
Disallow: /perso/
Disallow: /entravaux/
Disallow: /abonnes/prix.html
```



- User-agent: * signifie que l'accès est accordé à tous les agents (tous les spiders), quels qu'ils soient.
- Le robot n'ira pas explorer les répertoires /cgi-bin/, /tempo/, /perso/ et /entravaux/ du serveur ni le fichier /abonnes/prix.html.
le répertoire /tempo/, par exemple, correspond à l'adresse <http://www.monsite.com/tempo/>.
chaque répertoire à exclure de l'aspiration du spider doit faire l'objet d'une ligne Disallow: spécifique.
- La commande Disallow: permet d'indiquer que "tout ce qui commence par" l'expression indiquée ne doit pas être indexé.

Disallow: /perso

ne permettra l'indexation ni de <http://www.monsite.com/perso/index.html>, ni de <http://www.monsite.com/perso.html>

Disallow: /perso/

n'indexera pas <http://www.monsite.com/perso/index.html>, mais ne s'appliquera pas à l'adresse <http://www.monsite.com/perso.html>

Disallow: /

permet d'exclure toutes les pages du serveur (aucune aspiration possible).

Disallow:

Permet de n'exclure aucune page du serveur (aucune contrainte).

- Le fichier robots.txt ne doit pas contenir de lignes vierges (blanches). L'étoile (*) n'est acceptée que dans le champ User-agent.
- Elle ne peut servir de joker (ou d'opérateur de troncature) comme dans l'exemple : Disallow: /entravaux/*.
- Il n'existe pas de champ correspondant à la permission, de type Allow:.
- Enfin, le champ de description (User-agent, Disallow) peut être indifféremment saisi en minuscules ou en majuscules.
- Les lignes qui commencent par un signe dièse (#), ou plus exactement tout ce qui se trouve à droite de ce signe sur une ligne, est considéré comme étant un commentaire.

Un fichier robots.txt à la racine :

On peut trouver une documentation de référence à l'adresse suivante

<http://www.robotstxt.org/wc/faq.html>



The Web Robots FAQ...

These frequently asked questions about Web robots.

Send suggestions and comments to [Martijn Koster](#). This information is in the public domain.

CORPS DE LA PAGE

Contenu idéal :

Mais la meilleure garantie d'un bon classement reste la présence de mots clés pertinents dans le texte visible, le plus près possible du haut de la page (le plus près possible en dessous de la balise **<BODY>** du code HTML).

- Cela explique pourquoi les pages trop graphiques (contenant trop peu de texte visible) sont le plus souvent mal classées sur les moteurs de recherche.
- Positionnez, dans la mesure du possible, vos mots clés importants vers le début de votre page et mettez-les en valeurs (gras, police de grande taille, lien hypertexte, etc.) par les balises HTML suivantes :
<H>xxxxx</H> et/ou avec des style au besoin
- Une page optimisée doit contenir 100 mots maximum, parmi lesquels seuls les 30 à 50 premiers seront réellement pris en compte par les moteurs ! **Par conséquent les 2-3 premières phrases sont capitales.**

Dans la densité des mots réside une bonne partie des secrets de la réussite d'un bon référencement ! Plus la densité d'un mot clé est importante, plus le mot clé sera réactif sur cette page.

- Mais un même mot ne doit pas être répété trop de fois non plus...
- Visez pour les 2 ou 3 mots clés importants de votre page, une densité optimale de 10 à 15 %.
- Ne répétez pas un mot avec la même casse plus de 10 fois dans une même page. Souvent les moteurs s'arrêtent bien avant .

Petits Plus ... :

Voici quelques éléments qui peuvent vous faire gagner quelques positions dans la course

- Vous pouvez mettre quelques mots clés dans un ou 2 commentaires HTML (balise : **<!--commentaire ici-->**), mais ne dépassez pas **50 caractères**. Cependant, en règle générale, les moteurs de recherche ne tiennent pas compte des commentaires pour leurs calculs de pertinence.
- Les options **ALT** des balises insertion image telles que la séquence **** sont en règle générale prises en compte par les moteurs de recherche, même si leur importance est assez faible.



Outiref :

Autre url à tester :

Analyse de l'URL : <http://www.festijeux.com> (voir code source)

Titre (40 caract.)	<u>Meta Description</u> (0 caract.)	<u>Meta Keywords</u> (0 caract.)	<u>URL</u> (24 caract.)	<u>Texte visible</u> (229 mots)	<u>Contenu</u>	<u>Popularité</u>	<u>Densité des mots</u>
-----------------------	--	-------------------------------------	----------------------------	------------------------------------	----------------	-------------------	-------------------------

Titre de la page :

festijeux animation spectacle jeu soirée

Analyse :

Le titre de votre page contient **40** caractère(s), et **5** mot(s).
Vous avez exploité **40 %** des possibilités.

★ ★ ★ ★ ★ ★ ★ ★ ★ ★

Vous avez encore un peu de marge. Il peut contenir entre 50 et 100 caractères (une lettre accentuée valant pour un caractère) pour cette zone.

Les conseils d'Outiref

 Attention à ne pas répéter 2 fois le même mot dans le titre.

 Le titre de vos pages est certainement le critère de pertinence le plus important pour les moteurs de recherche à l'heure actuelle. Soignez-le bien !

 Ne proposez pas dans vos pages de titre entièrement en majuscules.

 Vous pouvez également coder les lettres accentuées en HTML (é = é).

 Le titre de votre page doit contenir entre 7 et 10 mots, les plus représentatifs possibles du contenu de la page.

 Par exemple, évitez les titres du type "Bienvenue sur notre site web", pas assez représentatifs du contenu de vos pages.

CERTIFICATION DES PAGES W3C

Pourquoi certifier une page :

Sur google, il semblerait que un petit plus soit accordé lorsque une page est conforme aux spécifications HTML édictées par le W3C.

La présence des balises suivantes, permet de certifier la validation de la page :

```
<p>
  <a href="http://validator.w3.org/check?uri=referer"></a>
</p>
```

qui visuellement se traduira par l'apparition de cette image dans la page



Certification HTML XHTML:

Pour valider le code d'une page HTML il suffit d'aller sur le site du W3C à l'adresse suivante :

<http://validator.w3.org>

W3C QUALITY Assurance **MARKUP VALIDATION SERVICE v0.6.7**

HOME ABOUT... NEWS DOCS HELP & FAQ FEEDBACK LINK CHECKER

The W3C Markup Validation Service is a free service that checks Web documents in formats like HTML and XHTML for conformance to W3C Recommendations and other standards.

VALIDATE YOUR MARKUP

VALIDATE BY URL

Address:

Enter the URL of the page you want to check. Advanced options are available from the [Extended Interface](#).

VALIDATE BY FILE UPLOAD

Local File:

Select the file you want to upload and check. Advanced options are available from the [Extended File Upload Interface](#).

Il faut saisir
l'adresse de
la page à
valider...

puis
on obtient



Address:

Encoding: iso-8859-1

Doctype: HTML 4.01 Transitional

Errors: 12

REVALIDATE WITH OPTIONS

Revalidate : ☐ Show Source ☐ Outline
☐ Parse Tree ☐ ...no attributes
☐ Validate error pages ☐ Verbose Output

[Help on the options is available.](#)

NO DOCTYPE FOUND! FALLING BACK TO HTML 4.01 TRANSITIONAL

A DOCTYPE Declaration is mandatory for most current markup languages and without one it is impossible to reliably validate this document. I am falling back to "HTML 4.01 Transitional" and will attempt to validate the document anyway, but this is very likely to produce spurious error messages for most non-trivial documents.

So what should I do? [Tell me more...](#)

Une norme stricte permettant de mettre toutes les chances de notre côté sera constituée par le couple même si la norme iso-8859 peut faire l'affaire :

Après éventuelles corrections, on obtiendra alors la possibilité d'insérer nos

balises dans la page HTML...

Certification CSS:

il suffit d'aller sur le site du W3C à l'adresse suivante :

<http://jigsaw.w3.org/css-validator/>

W3C[®] Service de validation CSS

[ACCUEIL] DONNER SON AVIS TÉLÉCHARGER DOCUMENTATION CRÉDITS

Bienvenue sur le service de validation CSS du W3C, un service gratuit pour vérifier la conformité des feuilles de style en cascade (CSS) autonomes ou intégrées aux documents (X)HTML vis-à-vis des recommandations du W3C.

Remarque : Si vous souhaitez valider une feuille de style incorporée dans un document (X)HTML, vous devriez d'abord [vérifier la validité de votre document \(X\)HTML](#).

Validez votre feuille de style

Validation CSS : [par URI](#) | [par téléchargement du fichier vers le serveur](#) | [par saisie directe](#)

Validation par URI

Saisissez l'URI du document (HTML avec CSS, ou CSS seul) dont vous souhaitez la validation :

Adresse :

LES SITES DYNAMIQUES

Problèmes de référencement des URL :

Par définition, un site dynamique n'est pas constitué de pages HTML statiques, pré existantes , dans lesquelles les moteurs pourront faire leur collecte de mots clés...

Mais plutôt de pages générées à la demande de l'utilisateur, souvent en provenance d'une base de donnée.

La conséquence est que souvent les URL sont très complexes, avec de symboles quelques peut problématiques pour la plupart des moteurs (genre ?- & ...). Et lorsque les moteurs acceptent ces caractères, ils ne descendent pas loin dans l'arborescence des fichiers (c'est le cas de Google).

Problèmes des sessions PHP :

Si votre site utilise des sessions PHP, alors les deux techniques existantes pour gérer les session (par URL ou par cookies) posent problèmes.

En effet dans la technique par URL comme l'identifiant de session est fourni dans l'URL chaque fois de un moteur revient sur une URL, cette URL change... GOOGLE n'indexe pas du tout ce genre de page !

Par la technique via Cookies, comme les moteurs refusent systématiquement tout marquage, le site risque d'être mal indexé...

Des solutions complexes URL Rewriting - :

Plusieurs solutions existent, mais leur mise en œuvre n'est pas des plus simples.

Pour les URL complexe, il existe des mécanismes de réécriture d'URL, qui permettent via une URL plus simple de retrouver l'URL complexe permettant d'accéder à la page dans la base de donnée...

Genre

article-1-12-5.html

en

article.php?id=1&page=12&article=5

cette technique spécifique doit être possible auprès de votre hébergeur sous le nom **URL Rewriting**

Pour les sessions PHP, il faut distinguer si le visiteur est un robot, ou non, et activer ou non les sessions en conséquence...



ANNUAIRES IMPORTANTS

Google Directory – Open Directory Project :

Il existe un annuaire nommé Open Directory que l'on peut trouver l'adresse suivante

www.dmoz.org

De type open source, il est à la base de beaucoup d'annuaires qui peuvent reprendre librement sa base. Comme Google d'ailleurs.

Cet annuaire est issu uniquement de l'analyse des titres des pages, du résumé fourni en meta et de l'URL, le contenu des pages n'est pas analysé.

N.B: on peut chercher dans cet annuaire avec les opérateurs désormais classiques **inurl:** ou **allinurl:** (pour une recherche dans une adresse) et **intitle:** ou **allintitle:** (pour une recherche dans le titre)

Quelques Annuaires :

Principaux annuaires francophones (ordre alphabétique)

Nom	Adresse
C'est trouvé	www.ctrouve.com
Francité	www.francite.com
Hachette Net	www.hachette.net
Nomade	www.nomade.fr
Yahoo France	www.yahoo.fr

Principaux annuaires mondiaux (ordre alphabétique)

Nom	Adresse
Encyclopedia Britannica	www.britannica.com
Excite	www.excite.com
LookSmart	www.looksmart.com
Lycos	www-english.lycos.com
Yahoo	www.yahoo.com



GOOGLE ANALITICS

Principe:

Accédez à l'adresse <http://www.google.com/analytics/fr-FR/>.

- Entrez l'adresse e-mail et le mot de passe de votre compte Google (Gmail), puis cliquez sur Connexion.
- Cliquez sur Inscription.
- Entrez l'URL du site Web que vous possédez et veillez à sélectionner <http://> ou <https://> dans la liste déroulante.
- Entrez le pseudonyme du compte dans le champ Nom du compte, Continuer.
- Indiquez vos coordonnées Continuer.
- Lisez les conditions d'utilisation de Google Analytics. Si vous les acceptez, cochez la case Oui, puis cliquez sur Créer un compte pour continuer.

La page Instructions relatives au suivi qui apparaît contient le code de suivi que vous devrez coller dans chaque page de votre site

En image, cela donne

Paramètres Google Analytics > Création d'un profil de site Web

Création d'un profil de site Web

Sélectionnez un type de profil de site Web

Indiquez si vous souhaitez ajouter un profil supplémentaire à un domaine existant ou créer un profil afin d'effectuer le suivi d'un nouveau domaine.

☒ Ajouter un profil pour un nouveau domaine

OU

☐ Ajouter un profil pour un domaine existant

Ajouter un profil pour un nouveau domaine

Veuillez indiquer l'URL du site dont vous souhaitez réaliser le suivi.

<http://>

Exemples : www.monsite.fr

Remarque : Vous pouvez inclure d'autres profils après l'ajout de celui-ci.

Pays ou territoire du fuseau horaire :

Fuseau horaire : (GMT+01:00) Paris



Code de suivi

Informations relatives à l'état du suivi	
Nom du profil :	www.festijeux.com
URL du site Web :	http://www.festijeux.com
ID de site Web (?) :	UA-637262-3
État du suivi :	⚠ Suivi non configuré (Dernière vérification : 19 févr. 2010 23:07:40) Le code de suivi Google Analytics n'a pas été détecté sur la page d'accueil de votre site Web. Pour que Google Analytics fonctionne, vous devez ajouter le code de suivi à toutes les pages de votre site. Vous pouvez l'ajouter vous-même ou par l'intermédiaire de votre administrateur Web.

Instructions relatives à la mise en œuvre du suivi

Standard **Avancée** **Personnalisé**

[Nouveau ! Essayez](#)

1 Sur quoi votre suivi porte-t-il ?

☒ Un seul domaine (paramètre par défaut)

Domaine : www.festijeux.com

☐ Un seul domaine associé à plusieurs sous-domaines

☐ Plusieurs extensions de domaine

☐ Je souhaite effectuer le suivi de mes campagnes AdWords.

Enregistrer et terminer

2 Collez ce code sur votre site.

Copiez le code suivant, puis collez-le dans chaque page dont vous souhaitez effectuer le suivi, en le plaçant juste avant la balise </body>. [En savoir plus](#)

```
<script type="text/javascript">
var gaJsHost = (("https:" == document.location.protocol) ?
document.write(unescape("%3Cscript src='" + gaJsHost + "goo
</script>
<script type="text/javascript">
try {
var pageTracker = _gat._getTracker("UA-637262-3");
pageTracker._trackPageview();
} catch(err) {}</script>
```

Insérer le code de suivi:

Google Analytics ne peut pas effectuer le suivi de vos pages si elles ne contiennent pas le code de suivi Google Analytics. Vous devez donc insérer ce code dans chacune des pages de votre site manuellement

- Notez que chaque profil est associé à un code de suivi spécifique.
- Dans la colonne **Paramètres** de ce profil, cliquez sur **Modifier**.
- Dans la partie supérieure droite de la zone "Informations relatives au profil du site Web principal", cliquez sur **Vérifier l'état**.
- Copiez ensuite le code de suivi apparaissant dans la zone de texte de la section "Instructions relatives à la mise en œuvre du suivi" et collez-le dans vos pages.
-

N.B: Lorsque vous ajoutez le code de suivi aux pages Web, assurez-vous de le coller dans la section du code HTML, juste avant la balise </body>.



N.B: Sites pilotés par base de données : insérez le code de suivi dans votre page index.php ou dans une page équivalente default.php...

Version de Analytics:

Si le site concerné utilise déjà Google Analytics, le code source contient un extrait de code JavaScript tel que :

ancienne version

```
<script type="text/javascript" src="http://www.google-analytics.com/urchin.js"></script>
```

```
<script type="text/javascript">
_uacct = "UA-xxxxxx-x"
urchinTracker();
</script>
```

nouvelle version

```
<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-xxxxxx-x");
pageTracker._trackPageview();
</script>
```

Résultat de Analytics:

assez classique et complet....

Tableau de bord

20 janv. 2010 - 19 févr. 2010



Fréquentation du site



2 036 Visites



3 843 Pages vues



1,89 Pages par visite



72,05 % Taux de rebond



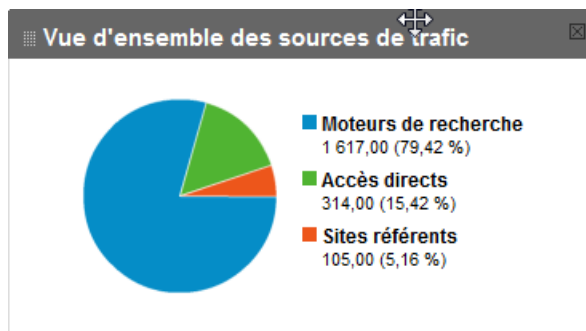
00:01:23 Temps moyen passé sur le site



83,45 % Nouvelles visites (en %)

et parfois surprenant...





voire très intéressant

Tableau de bord

- Alertes Beta
- Visiteurs
- Sources de trafic**
 - Vue d'ensemble
 - Accès directs
 - Sites référents
 - Moteurs de recherche**
 - Toutes les sources de trafic
 - AdWords
 - Mots clés
 - Campagnes
 - Versions d'annonce
- Contenu
- Objectifs

Fréquentation du site

Ensemble d'objectifs 1

Type d'affichage :

Visites

1 617

Total du site (en %) : 79,42 %

Pages par visite

1,91

Moyenne du site : 1,89 (1,18 %)

Temps moyen passé sur le site

00:01:24

Moyenne du site : 00:01:23 (1,48 %)

Nouvelles visites (en %)

88,81 %

Moyenne du site : 83,45 % (6,42 %)

Taux de rebond

71,55 %

Moyenne du site : 72,05 % (-0,70 %)

	Source	Aucun	Visites	Pages par visite	Temps moyen passé sur le site	Nouvelles visites (en %)	Taux de rebond
1.	google		1 536	1,90	00:01:23	88,35 %	71,81 %
2.	yahoo		42	2,55	00:01:55	97,62 %	66,67 %
3.	bing		18	1,44	00:00:21	94,44 %	72,22 %
4.	search		14	1,64	00:02:45	100,00 %	64,29 %
5.	voila		4	1,50	00:00:18	100,00 %	75,00 %
6.	aol		2	3,50	00:01:14	100,00 %	0,00 %
7.	ask		1	3,00	00:00:00	100,00 %	100,00 %

Et aussi

Visiteurs

- Vue d'ensemble
- Analyse comparative
- Synthèse géographique
- Nouveaux vs connus
- Langues
- Visiteurs - Tendances
- Fidélité des visiteurs
- Capacités du navigateur
 - Navigateurs**
 - Systèmes d'exploitation
 - Navigateurs et systèmes d'exploitation
 - Couleurs d'écran
 - Résolutions d'écran
 - Versions Flash
 - Prise en charge de Java

Navigateur **Aucun** **Visites** **Visites** **Contribution de "Navigateur" au total : Visites**

Navigateur	Visites	Pourcentage
1. Internet Explorer	1 130	55,50 %
2. Firefox	706	34,68 %
3. Chrome	100	4,91 %
4. Safari	45	2,21 %
5. Opera	23	1,13 %
6. Mozilla	18	0,88 %
7. Konqueror	11	0,54 %
8. Opera Mini	2	0,10 %
9. LG	1	0,05 %

