

# Référencement

Internet

Juin 2013

Jérôme GUY

# Référencement internet

---

## Contenu

Définition.....	4
Tester si son site est référencé.....	4
Le triangle d'or .....	4
Quelques chiffres.....	5
Historique .....	5
Le navigateur .....	5
Les Robots .....	6
Fréquence de passage de googlebot.....	6
Les sanctions de Google .....	6
Concepts d'un site SEO-Friendly.....	6
Le langage HTML .....	6
Les balises de titre .....	6
Les balises strong et em .....	6
Les liens (anchor text) .....	6
Les images .....	7
Ratio contenu/HTML .....	7
Ergonomie du site .....	7
Code source HTML de base actuellement.....	7
Règle d'or des 600 px .....	8
De bonnes performances .....	8
Utiliser les sprites CSS.....	9
JavaScript.....	9
Requêtes SQL.....	9
Communiquez avec GoogleBot .....	10
La balise title.....	10
La balise meta description.....	10
La balise meta keywords .....	10
La balise meta Robots .....	10
Fichier robots.txt .....	11
URL .....	11

Nom de domaine .....	11
Noms des fichiers .....	11
Erreur 404.....	11
En résumé.....	12
Google Juice.....	12
Les mots clefs .....	12
Suivi des mots clefs .....	13
Aide du choix des mots clefs Google trends .....	13
La longue traine.....	13
En résumé.....	14
Les mauvaises habitudes.....	14

## Définition

Search Engine Optimisation : **SEO**.

Le référencement désigne l'ensemble des techniques permettant d'être présent dans les résultats des moteurs de recherche et annuaires

## Tester si son site est référencé

Site:monsite.com

## Le triangle d'or

Eye tracking

The image shows a Google search results page for the query "appareil photo numérique". A heatmap overlay indicates user attention, with the highest concentration (red) on the first three search results, forming a triangle. The search results include various links related to digital cameras, such as "Appareil Photo Numérique | Acheter Appareils Photo pas cher | Malligo", "Appareil photo numérique - Comparer et acheter votre appareil...", "Guide pour choisir appareil photo numérique février 2009", "Photo et Vidéo Numérique - appareil photo / numérique - argentique...", "APPAREIL PHOTO NUMÉRIQUE - achat, comparaison et test, prix moins cher", "Acheter Appareil photo numérique au meilleur prix, comparer et...", "Appareils photo numériques - Photo Caméscope / Appareil photo...", "Appareils Photo numérique - Test, avis, comparatif et forum", "Appareil photo numérique - Achat appareil photo numérique discount...", and "Appareil photo numérique - conseils d'achat et d'utilisation". The page footer shows "Recherches apparentées à appareil photo numérique" and the URL "www.miratech.eu".

**SERP** : Search Engine Results Page

Le triangle d'or est positionné sur les 3 premiers résultats de la recherche.

Etre indexé par google c'est bien, mieux vaut être bien référencé.

## Quelques chiffres

**85 %** des visites viennent des moteurs de recherche

**10 %** par accès direct

**5%** lien entrant

## Historique

### *1993-1998 : avant google*

Altavista était leader, quelques mots clefs dans les balises méta suffisaient

### *1998-maintenant : google, le référencement devient complexe*

En 1995, deux étudiants de l'université de Stanford, Larry Page et Sergeï Brin créés un algorithme révolutionnaire. La pertinence d'un site est fonction du nombre de liens pointant vers ce site et de leur qualité. Google est né en 1998.

Aujourd'hui Google retient plus de 200 critères de pertinence : qualité du code HTML, présence sur réseaux sociaux...

Des ingénieurs travaillent constamment pour l'améliorer.

90% en France, Bing 3%, Yahoo 2%

65 % dans le monde, Yahoo 15 %, Bing 15 %

Les autres moteurs utilisent maintenant les mêmes critères que Google.

Google ne diffuse pas ses critères.

Google fournit des outils liés au référencement.

- **Google Analytics** : statistiques sur les visites
- **Google Webmaster tools** : aides pour webmaster
- **Google Adsense** et **Google Adwords** : référencement payant

En résumé : Le référencement c'est l'optimisation de son site pour obtenir une meilleure place dans les moteurs de recherche.

## Le navigateur

Comme pour HTML, il faut utiliser plusieurs navigateurs à jour.

Parts de marché en 2013 :

- **Internet Explorer** : 35 %
- **Chrome** : 25%
- **FireFox** : 20 %
- **Safari** : 15 %

Extensions FireFox : Web Developer, Firebug, WebRank Toolbar, KGen

Extensions Chrome : Page Rank Status, SEO Site Tools

## Les Robots

GoogleBot, programme informatique qui parcourt les sites pour les indexer.

Aujourd'hui plus de 2 milliards de sites.

Fonctionnement :

- Scanne le code HTML du site et l'envoie à Google
- Explore les liens `<a> </a>`
- Scanne la nouvelle page et recommence

Le robot s'intéresse au contenu pas à la mise en forme, seule la source HTML l'intéresse.

## Fréquence de passage de googlebot

Au démarrage du site, le robot passe fréquemment pour vérifier si le site évolue. Si c'est le cas, il repassera souvent, c'est positif pour le référencement. Si le site n'évolue pas, il passera moins souvent voir très rarement, ce n'est pas bon pour le référencement du site.

Dès les premiers jours du site quand il a été soumis à google, il faut faire vivre le site.

Le référencement se construit sur la durée, et rien n'est jamais acquis.

## Les sanctions de Google

Si le site respecte les consignes de Google et propose du contenu pertinent et de qualité, alors il sera bien noté

Si au contraire le site ne respecte pas les consignes de Google et plagie d'autres sites alors, il aura des pénalités et son référencement sera mauvais. Google peut même supprimer le site de son index.

## Concepts d'un site SEO-Friendly

Ne jamais proposer à Google un site en développement, sinon la barre est difficile à redresser.

## Le langage HTML

Ecrire du code HTML propre et valide W3C. Cf W3c validator. Il faut des solides bases HTML et CSS.

Importance d'utiliser les bonnes balises.

`<strong>` n'a pas la même signification que `<span>`

`<strong>Formation Référencement</strong>` : ce texte est important

## Les balises de titre

`<h1>` à `<h6>` : primordiales pour le référencement

Il faut absolument une balise `<h1>` sur une page : c'est le titre de la page. Il peut y avoir plusieurs h1 par sections si cela est justifié.

## Les balises strong et em

Elles sont importantes, elles signifient "texte important". Ne pas en abuser, à utiliser à bon escient.

## Les liens (anchor text)

`<a href="lien">texte</a>`

Le texte est important, il donne des informations précieuses sur la nature du lien à google. Fait gagner des points.

## Les images

```

```

Ne pas renseigner la balise alt entraîne des pénalités.

## Ratio contenu/HTML

Il faut au minimum 15% du code HTML qui soit du contenu. Plus le ratio est élevé mieux c'est.

Il faut du code HTML propre, pas trop complexe, et conforme aux derniers standards en cours.

Produire si possible du code qui passe la validation W3C.

## Ergonomie du site

- Un **header**, avec un logo sur toutes les pages affiché par la feuille de style.
- Le **centre de la page** : découpé en 2 parties : un menu et le contenu
- Un **footer** : on y trouve souvent 'contact' et 'mentions légales'.

Les internautes sont habitués à cette mise en page sobre mais terriblement efficace. C'est de cette manière que google indexera au mieux un site.

## Code source HTML de base actuellement

```
<!DOCTYPE html>
<!-- La section <head>. Nous aurons l'occasion d'en parler plus tard dans le cours ! -->
<head>
<title>Le titre de ma page</title>
<meta http-equiv="Content-Type" content="text/html"; charset="utf-8" />
<link rel="stylesheet" type="text/css" href="style.css" />
<link rel="shortcut icon" href="images/favicon.ico" />
</head>
<!-- Mise en page classique avec le header, le corps, et le footer ! Les balises structurantes de HTML5
sont employées. -->
<body>
<div id="site">
<header>L'entête de mon site ! Et bientôt, il y aura une bannière ou un logo ici !</header>
<nav id="menu">
<ul>
<li>Élément de menu 1</li>
<li>Élément de menu 2</li>
<li>Élément de menu 3</li>
<li>Élément de menu 4</li>
</ul>
</nav>
<section>
<article>
<h1>Le titre de mon article ;)</h1>
<p>Salut, voici le contenu de ma page web  </p>
</article>
</section>
```

```
<footer>
<p>Le contenu de mon pied de page, mention légale, etc. </p>
</footer>
</div>
</body>
</html>
```

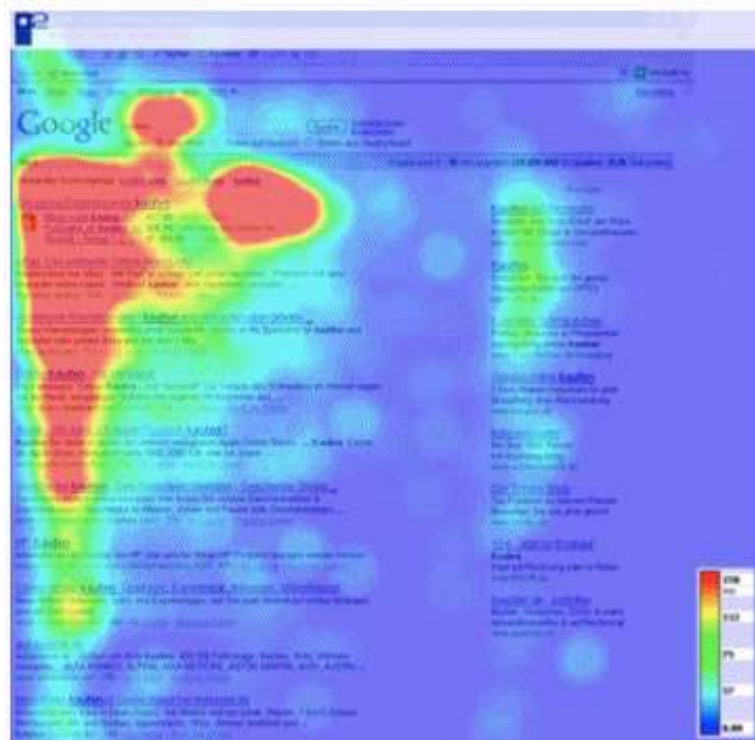
Cf le site W3Schools

A tester les frameworks CSS 960 Grid System et Bootstrap twitter

Plus un visiteur passe de temps sur un site mieux il sera noté. Google est capable de déterminer le temps que passe un visiteur sur un site. Il faut donc un site ergonomique.

## Règle d'or des 600 px

Equivalent de la une d'un journal, c'est l'accroche qui donne envie d'acheter et de lire l'article complet.



Il est impératif de bien travailler le début d'un article pour qu'il soit accrocheur.

D'un point de vue typographique et ergonomique, il faut un gros titre, une introduction de quelques lignes et ensuite l'article.

Pour Google, c'est identique, le début de l'article est primordial, bien plus que la fin de l'article.

Une expression placée tôt dans la page a bien plus d'importance pour le référencement.

## De bonnes performances

Google préfère un site qui s'affiche rapidement qu'un site original et design, il se met à la place de l'utilisateur.

L'outil **google pagespeed**: analyse et donne des conseils pour optimiser les performances de son site.



## Utiliser les sprites CSS

Pour limiter les requêtes http, on compile toutes les images en une seule.



L'image est ensuite positionnée à l'endroit voulu par du CSS.

## JavaScript

```
<script src="fichier.js"> </script>
```

L'idéal est de le charger juste avant la fermeture de la balise </body>. Quand le contenu de la page est chargé, le javascript est appelé et ne ralentit pas le chargement de la page.

## Requêtes SQL

Un site avec une BDD est 5 fois plus lent qu'un site utilisant uniquement des fichiers.

Il faut limiter les requêtes SQL et utiliser des index sur les tables.

**Minify** est un utilitaire qui nettoie et optimise le code du site.

### En résumé

- Penser à l'utilisateur, pas au GoogleBot
- Code propre en HTML5
- Balise de titre et de mise en valeur du texte
- Texte en début de page plus important que celui en fin de page
- Le site doit se charger vite pour plaire.

## Communiquez avec GoogleBot

Lui dire quelle page indexer, ne pas indexer...

A placer dans l'entête <head> ...</head>

### La balise title

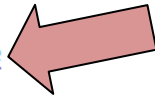
Balise **<title>** très importante, elle génère le résultat de la recherche google. Elle doit être unique sur chaque page du site. Il faut énormément travailler cette balise.

[Wikipédia, l'encyclopédie libre](#)

[fr.wikipedia.org/](http://fr.wikipedia.org/)

Articles à nature encyclopédique sur tous les sujets, créés et édités bénévolement par des contributeurs anonymes. Wiki basé sur le logiciel libre Mediawiki.

[Portail:Accueil](#) - [Français](#) - [Accueil](#) - [Portail:Sciences](#)



Pas plus de 60 caractères, précis et direct. Placer quelques mots clefs dans le title. Ne jamais omettre de renseigner le title.

### La balise meta description

**<meta name="description" content="La description de mon site"/>**

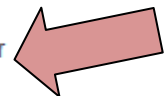
200 caractères maximum

[Wikipédia, l'encyclopédie libre](#)

[fr.wikipedia.org/](http://fr.wikipedia.org/)

Articles à nature encyclopédique sur tous les sujets, créés et édités bénévolement par des contributeurs anonymes. Wiki basé sur le logiciel libre Mediawiki.

[Portail:Accueil](#) - [Français](#) - [Accueil](#) - [Portail:Sciences](#)



Si pas de description google en trouve une dans la page.

### La balise meta keywords

**<meta name="keywords" content="Mes mots clefs"/>**

Plus aussi importante aujourd'hui que du temps d'Altavista. Les mots clefs sont séparés par des virgules.

### La balise meta Robots

**<meta name="Robots" content="follow, index"/>**

Si on ne dit rien, il indexe la page et suit les liens.

*No follow, no index.*

*All* : tout suivre, tout indexer

*None* : ne rien suivre, ne rien indexer

Pour un lien spécifique, attribut *rel*

**< a href="mapage.html" rel="no follow">ma page</a>**

## Fichier robots.txt

A la racine du site uniquement. "robots.txt" en minuscule

- User-agent : spécifie avec quel robot on communique

*User-agent : Googlebot/2.1 (+http://www.google.com/bot.html)*

*User-agent : \** => tous les robots

- allow, disallow

*Allow : /index.php*

*Disallow : /admin*

Il est bien vu de créer un fichier robots.txt qui autorise toutes les pages pour tous les robots :

*User-agent : \**

*Disallow :*

## En résumé

- Title et meta description très importants
- Meta keywords plus indispensable aujourd'hui
- Meta robots et robots.txt pour donner des instructions aux robots

## URL

La dernière mise à jour de Panda rend obsolète l'URL pour le référencement, mais reste indispensable pour l'utilisateur

## Nom de domaine

Le nom de domaine a une place primordiale. Bien choisir son nom de domaine est important. Il doit être en adéquation avec le thème et les mots clefs du site.

Doit avoir entre 10 et 15 caractères, 20 max.

## Noms des fichiers

Si beaucoup de pages, il faut structurer le site avec des dossiers et sous-dossiers, nommés proprement.

Utiliser des tirets et pas des underscores, c'est plus lisible.

cours/referencement-web.html

Il est recommandé de retrouver dans l'url le H1 de la page.

## Erreur 404

Si une page ou un fichier est supprimé, google ne le verra pas immédiatement => erreur 404.

Cela entraîne de grosses pénalités. Il faut absolument éviter les erreurs 404.

Dans le fichier .htaccess, personnaliser la page 404 :

*ErrorDocument 404 http://www.site.com/erreur.html*

Créer des redirections dans le .htaccess pour éviter les erreurs 404 (redirection 301) :

*RedirectPermanent /ancienne-url.html http://www.site.com/nouvelle-url.html*

URL à éviter :

<http://www.site.com/page.php?id=1&ion=le-site&titre=ouverture-!>

<http://www.site.com/page.php?id=1>

<http://www.site.com/page.php?ref=KHDH0184DAAX578451F>

<http://www.youtube.com/watch?v=7Hk5uVv8JpM>

Pour pallier à ce problème il faut utiliser URL Rewriting

## En résumé

- Nom de domaine très important, possédant un mot clé en rapport avec la thématique, il apparait sur chaque page du site
- Créer une arborescence logique
- URL propres même si google n'en tient plus compte, c'est plus clair pour l'utilisateur
- Les erreurs 404 doivent être absolument évitées

## Google Juice

Autre nom *Link juice*.

Il faut une arborescence bien pensée. Tout le jus est envoyé sur l'accueil *index.html*

Le jus se déverse par les liens hypertextes. L'idéal est que chaque page d'un site déverse son jus vers une autre page du site, il ne faut pas de pages orphelines.

Il ne faut pas que la page d'accueil ait des liens vers toutes les pages du site.

Les erreurs 404 font perdre du jus. Il faut corriger les liens morts ou au pire personnaliser sa page 404.

Il faut boucher les trous avec `rel="nofollow"` par ex sur un forum, pas géré directement par le concepteur.

Chaque balise `<a>` d'une page divise le jus. Si attribut no follow le pourcentage est perdu, il n'est pas redistribué.

## En résumé

- Il faut transmettre le jus aux différentes pages du site
- Jamais de page 404
- Chaque page du site doit être ciblée par un lien interne

## Les mots clefs

Un bon mot clef doit être beaucoup recherché dans Google.

Il doit être lié à la thématique du site.

Lister les mots clefs et les trier par importance.

Les incorporer dans :

- Les balises **title**
- Les liens hypertextes : `<a href="lien.html"> mots clefs </a>`
- Titre **H1**
- Titres **H2** à **H6**

- Balises <strong> <em>
- Texte du site

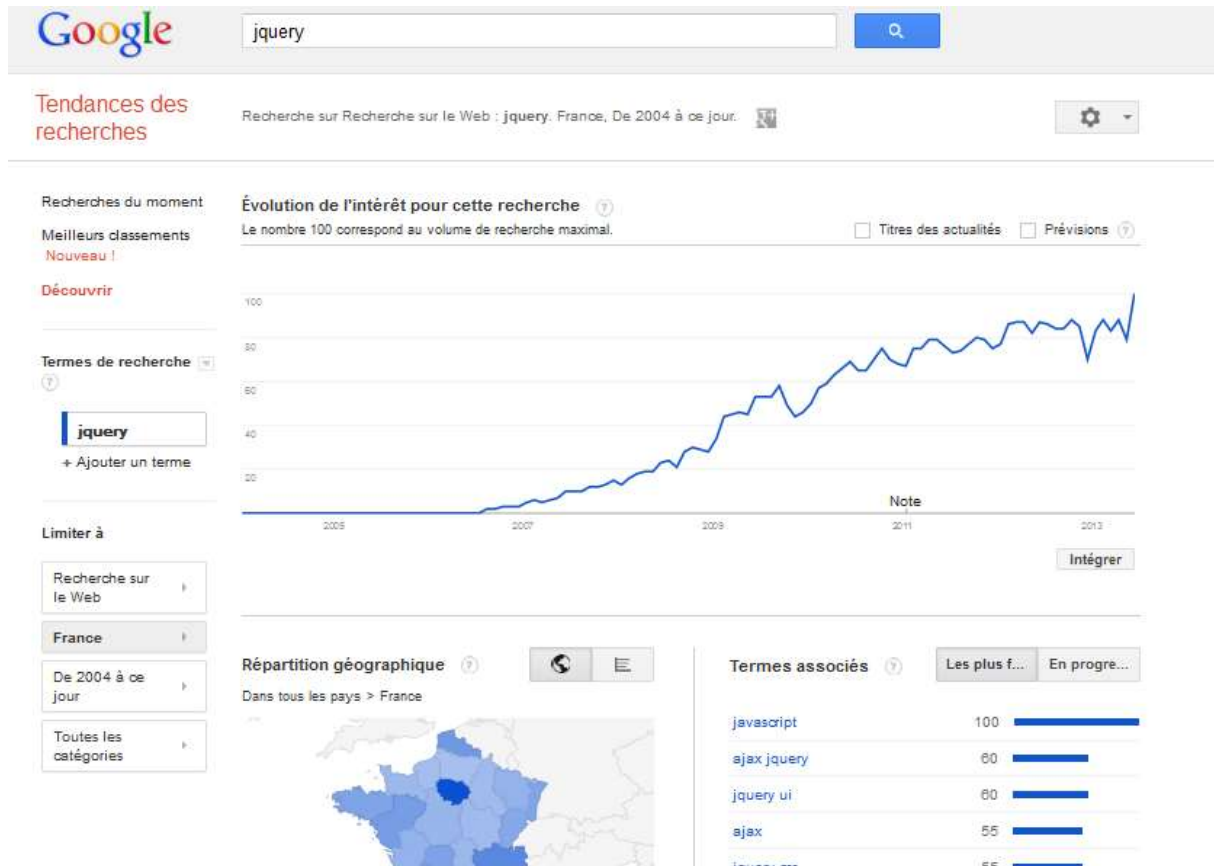
Dans les 600 premiers pixels. De manière naturelle.

## Suivi des mots clefs

Google Analytics ou Xiti. Code javascript à intégrer au site. Rapport sur les mots clefs.

## Aide du choix des mots clefs Google trends

En tapant un mot clef, il donne l'évolution dans le temps, et des mots clefs associés.



## La longue traîne

<http://www.seomix.fr/longue-traîne/>

Un journaliste a étudié le CA d'Amazon, 2 catégories de produits :

- Peu de produits très vendus, qui représentent 20% du CA
- Beaucoup de produits très peu vendus, mais qui représentent 80 % du CA

= > Tous ces produits qui se vendent peu rapportent bien plus que les best-sellers !



longue traîne en jaune

Dans le référencement, la longue traîne est une liste de mots qui va apporter beaucoup de trafic

Chaque mot clef rapporte peu de visites, mais mis bout à bout c'est énorme.

Les mots clefs best-sellers apportent des visites mais beaucoup de concurrence.  
Il faut donc beaucoup de contenu.

### En résumé

- Choisir les bons mots clefs est primordial, google trends aide à les choisir
- Intégrer les mots clefs naturellement dans le site
- Suivez l'efficacité des mots clefs
- Produire beaucoup de contenu pour la longue traine

L'avenir est le web sémantique généré grâce aux micro-données. Cf [schema.org](http://schema.org)

### Les mauvaises habitudes

- Plagier un site.
- Dupliquer du contenu sur plusieurs pages de son site

Pour dire au robot de ne pas tenir compte d'une page dupliquée :

`<link rel="canonical" href="a.html" />` `<!--dans l'en-tête de la page B qui duplique la page A-->`

Le robot ne doit pas suivre les pages au contenu pauvre, comme les pages contacts.

- Eviter le flash