

Les Balises META et le référencement

1. Utilité

1.1 Présentation

Véritable "sac à malice" du concepteur de sites Web, ces mystérieuses balises META feront de votre site une gentille œuvre d'amateur ou lui donneront la signature d'un site de "pro".

Conformément à la philosophie du Html, les éléments relatifs au document sont dissociés du corps du document lui-même et repris dans la zone d'en-tête (c-à-d entre les balises <HEAD> et </HEAD>). Les balises META apportent une série d'informations relatives à la teneur de votre page Web et sont donc toujours comprises entre les balises <HEAD>.

Les informations ainsi apportées intéressent :

- principalement le référencement de votre site et les moteurs de recherche.
- un peu vos lecteurs (avertis bien entendu).
- et accessoirement le browser ou le navigateur.

1.2. Les moteurs de recherche

Quand on apprend que plus de la moitié des utilisateurs du Web passent par des moteurs de recherche, il importe de soigner le référencement de son site pour y apparaître de façon correcte et si possible en rang utile.

Au fil de vos "joggings" sur le Web, vous ne pouvez pas ne pas avoir rencontré des sites référencés de cette façon sibylline et peu informative :

No title

<FRAMESET><FRAME ...><FRAME ...></FRAMESET><NOFRAMES>Ce site comporte des frames</NOFRAMES>

ou si vous souhaitez une variante :

Untitled

<SCRIPT LANGUAGE="JAVASCRIPT"><!--function animate(form){if (!var). etc.//--></SCRIPT>

1.3. Les lecteurs

Ne vous est-il jamais arrivé, surfeur pressé ou économe, d'avoir enregistré un page jugée intéressante et d'avoir, par la suite, voulu désespérément retrouver l'adresse de la page d'accueil du site ou l'adresse e-mail de l'auteur ?

De judicieuses balises META peuvent, entre autres choses, offrir ce petit plus à vos lecteurs.

1.4. Le navigateur

Rares mais au combien spectaculaires, certaines balises META sont directement interprétées par le navigateur pour, par exemple, charger automatiquement un page Html après x secondes, effectuer une transition style PowerPoint ou forcer le rechargement d'une page sur le réseau bien que déjà présente dans le cache.

Mais on entre ici dans le déplaisant domaine du "Cela fonctionne sous le browser Machin version x mais pas sous le browser Chose version y"... Sans commentaires !

2. Les indispensables

2.1. <META NAME="keywords" CONTENT="mot-clé1,mot-clé2,mot-clé3,...">

S'il n'en fallait qu'une, ce serait assurément celle-là ! Tous [ou presque] les moteurs de recherche l'utilisent.

Quelques commentaires s'imposent cependant :

- Les mots-clés doivent être séparés par une virgule. L'auteur préconise de ne pas mettre d'espaces (bien que cela se rencontre chez des webmestres confirmés) afin d'éviter toute confusion lors des recherches par mot-clé.

- On parle dans la littérature d'une limitation à 1000 mots-clés. Plus qu'il n'en faut.
- On conseille aussi de mettre tous les mots clés en minuscules pour éviter les problèmes selon que le moteur de recherche tient compte de la "casse", c'est-à-dire de la différence entre majuscules et minuscules (en anglais « case sensitive » ou non)
-
- Tout l'art consiste à trouver les bons mots-clés relatifs au contenu de votre site. Mettez-vous à la place de vos lecteurs potentiels. Quels mots, quels synonymes, quelles alternatives peuvent être utilisés pour décrire votre site?
- De grâce, éviter de mettre des mots-clés "bidon" qui sont bien entendu très attirants mais sans rapport avec votre site.
- La tentation est grande de répéter un certain nombre de fois un même mot-clé pour espérer un meilleur classement. Par exemple, <META NAME="keywords" CONTENT="html,html,html,html,html">. Désolé pour vous mais ce truc a vécu et est maintenant pénalisé [spam] par les moteurs de recherche.

2.2. <META NAME="description" CONTENT="une brève description de la page">

Cette description sera pertinente, attirante et brève. En effet, selon les moteurs de recherche seuls les 150 à 240 premiers mots seront repris. Bavards s'abstenir.

2.3. <TITLE> ... </TITLE>

Les moteurs de recherche tiennent fortement compte des titres des documents. Il faut impérativement en mettre sur toutes les pages d'un site. Même pour les pages qui n'apparaissent pas directement comme les sous-pages d'une page de frames !

On dit également que le fait de reprendre un ou des mot(s)-clé(s) dans le titre de toutes les pages d'un site est très favorable pour un meilleur classement.

2.4. Et c'est tout pour les "indispensables"

Cela peut paraître léger mais si vous tenez compte de ces trois éléments, 90% d'un bon référencement seront ainsi réalisés. Les autres conseils font partie des "trucs et astuces" dont on ne peut assurer l'efficacité. En effet, dès qu'on tente de pousser plus avant une étude sur le réel fonctionnement d'un robot de recherche par rapport à un autre, on entre dans une zone de brouillard plus ou moins épais...

Nous reviendrons sur ces trucs et astuces du référencement plus loin dans ce chapitre.

3. Les utiles

3.1. <META NAME="author" CONTENT="nom de l'auteur">

Cette balise est d'une utilité discutable car rares sont les moteurs de recherche en tiennent compte. A ma connaissance, seul Nomade fournit l'URL du site après avoir introduit le nom de l'auteur.

Il n'empêche qu'il est d'une légitime fierté de signer son œuvre. N'est-il pas ?

3.2. <META NAME="Copyright" CONTENT="Copyright © date nom">

Il ne m'importe pas de discuter ici de la valeur juridique et pratique d'une mention de copyright sur le Web. D'autres sites s'en chargent.

Il me plaît cependant de noter que dans les e-mails reçus, il n'est pas rare que des étudiants ou des formateurs me demandent de pouvoir reproduire tout ou partie de mon site. Une règle élémentaire de la Netiquette est de respecter le copyright. "Le gratuit n'est pas forcément sans valeur".

3.3. <META NAME="Distribution" CONTENT="Global ou Local">

Cette balise indique la destination de l'information de la page. Soit qu'elle est "Global" et donc destinée à être largement diffusée soit qu'elle est "Local" et donc à diffusion restreinte.

3.4. <META NAME="Generator" CONTENT="nom de l'éditeur Html utilisé">

Aucune influence, ni pour les moteurs de recherche, ni pour le navigateur utilisé. Cette information intéresse cependant au plus au point les responsables marketing des sociétés proposant des éditeurs Html pour calculer la part de marché de leur produit. Alors, pourquoi pas si cela les amuse...

3.5. <META NAME="Rating" CONTENT="Destination de votre audience">

Permet de définir le contenu de votre site. Les appréciations sont General ou Mature ou Restricted ou 14 years pour respectivement tout public, adulte, accès restreint ou 14 ans.

Cet indice d'audience n'a aucune influence directe.

3.6. <META NAME="Robots" CONTENT="instructions pour les robots">

Par cette balise vous pouvez indiquer aux robots de recherche automatique si vous souhaitez que votre site soit ou ne soit pas indexé par eux.

Les instructions sont :

- All (défaut) permet aux robots d'indexer vos pages et de suivre les liens hypertextes d'une page à l'autre.
- None dira aux robots de ne pas indexer vos pages et de ne pas suivre les liens.
- Index indique que vos pages peuvent être indexées par les robots.
- NoIndex pour que les robots ne procèdent à aucune indexation.
- Follow donne la permission aux robots de suivre les liens hypertextes des pages
- NoFollow pour le contraire.

Exemples :

<META NAME="Robots" CONTENT="None">

<META NAME="Robots" CONTENT="Index, Nofollow">

Cette balise est reconnue par tous les robots de recherche automatique.

3.7. <META HTTP-EQUIV="Content-language" CONTENT="fr">

Cette balise déclare la langue utilisée dans le document Html. De plus en plus utile, maintenant que des moteurs de recherche anglo-saxons (et non des moindres comme Altavista et Hotbot) ont inclus la langue dans leurs critères de recherche.

3.8. <META HTTP-EQUIV="Reply-to" CONTENT="votre adresse e-mail">

Cette balise permet au lecteur averti de connaître votre adresse e-mail si elle n'apparaît pas sur la page qu'il consulte (peut-être hors connexion, si elle a été préalablement enregistrée). Finalement, il est peut-être plus utile de donner son adresse électronique que son nom.

3.9. <META HTTP-EQUIV="Reply-to" CONTENT="URL de votre page d'accueil">

Variante de la balise précédente. Il est parfois utile de rappeler l'adresse de la page d'accueil de votre site pour ceux qui y sont entrés par une des pages de celui-ci.

4. Les éventuelles

A côté des balises META indispensables à un bon référencement auprès de moteurs de recherche, à côté des balises META qu'il est bon d'ajouter pour la bonne information du lecteur, il y a encore celles que vous utiliserez de temps en temps lors de l'élaboration de sites. Parcourons en quelques uns.

4.1. Refresh

Un grand classique du Html (2.0 ?) même si cette balise META n'est que rarement utilisée et alimente les trucs et astuces depuis des années. Cette balise META charge automatiquement la page spécifiée à l'attribut URL après un délai de x secondes.

<META HTTP-EQUIV="Refresh" CONTENT="x;URL="adresse">

Cette balise est fréquemment utilisée pour rediriger automatiquement un visiteur dans le cas où l'adresse de votre site a été modifiée.

4.2. Transition

De très jolis effets de transition style PowerPoint sont possibles avec simplement une ligne de code. Mais autant vous le dire tout de suite, cela ne fonctionne **que sous Explorer 4 et plus**.

Le code est :

```
<META HTTP-EQUIV="Page-Enter" content="revealTrans(Duration=1.0,Transition=23)">
<META HTTP-EQUIV="Page-Exit" content="revealTrans(Duration=1.0,Transition=23)">
```

Quelques explications :

- Page-Enter et Page-Exit signifie que l'effet de transition se produira à l'entrée de la page ou à la sortie de celle-ci.
- Duration détermine la durée de la transition en secondes. Elle est dans l'exemple de 1 seconde. A l'usage, cette durée n'est pas d'une précision absolue.
- Transition est un nombre de 1 à 23 pour l'effet de transition choisi. Le chiffre 23 donne une transition aléatoire (au hasard). Les autres transitions se répartissent de 1 à 22. Ainsi, 7 ouvre la page de droite à gauche, 17 a le même effet mais en diagonale, 22 découvre la page avec un effet de lignes horizontales aléatoires, etc.
- Précisons que si ces transitions ressemblent furieusement aux transitions de PowerPoint, elles fonctionneront très bien même si PowerPoint n'est pas installé sur la machine de votre visiteur.
- Et pour terminer, vous ne verrez les effets de cette transition que lorsque vous entrez dans la page à partir d'une autre page.

4.3. Expires

Cette balise META "dit" au navigateur la date à laquelle la page Html doit être considérée comme périmée. Avec Netscape, une requête pour un document dont la date est expirée initialisera une recherche sur le réseau au lieu de prendre la page éventuellement présente dans le cache. Ceci est très utile pour les pages fréquemment mises à jour.

Exemple :

```
<META HTTP-EQUIV="expires" CONTENT="Wed, 23 Feb 1999 10:49:02 GMT">
<META HTTP-EQUIV="expires" CONTENT="0">
```

Il faut noter que les robots de recherche peuvent retirer ces pages dites périmées de leur base de donnée.

4.4. Pragma

C'est une autre façon de contrôler le cache du navigateur. Avec ce META, vous pouvez demander au browser de ne pas tenir la page dans le cache.

```
<META HTTP-EQUIV="Pragma" CONTENT="no-cache">
```

On rapporte que cela fonctionne sous Netscape mais pas par Explorer...

5. Trucs et astuces de référencement

Alors que les balises META font partie du Html, il existe des tas de trucs et astuces - plus ou moins vérifiables ou vérifiés - qui sont sensés améliorer le rang de votre site parmi les centaines renvoyés lors d'une recherche par un mot-clé. Vérité ou rumeur... Je vous laisse seul juge.

5.1. Référez-vous

Cela peut paraître une évidence mais si vous attendez qu'un robot de recherche vienne visiter votre site, vous risquez de devoir attendre longtemps... Hotbot, le moteur de recherche le plus complet estime lui-même que sa base de donnée ne reprendrait que 30 à 40% des sites du World Wide Web.

Alors préparez votre liste des moteurs de recherche, prévoyez quelques heures dans votre emploi du temps et au travail ! Sur la page d'accueil de chaque moteur de recherche, vous trouverez "Ajouter un site", "Référencement" ou quelque chose du même acabit. En cliquant sur le lien, vous ouvrirez un formulaire, plus ou moins long, à remplir pour référencer directement votre site.

N'oubliez pas que certains moteurs de recherche style annuaire (comme Yahoo) se réservent de droit de reprendre ou de ne pas reprendre votre site, qu'il faut compter généralement 2 à 3 semaines avant d'apparaître dans la base de donnée et un bon mois avant que les internautes commencent à s'intéresser à votre site. Patience donc ...

5.2. De l'importance des mots-clés

Pour obtenir un bon classement, il faut non seulement définir ces mots-clés dans la balise <META NAME="keywords" CONTENT="mot-clé1,mot-clé2,mot-clé3,..."> mais il est aussi recommandé :

- que le ou les mots-clés soi(en)t repris dans le premier paragraphe de la page.
- que le ou les mots-clés soi(en)t repris dans le titre de la page (<TITLE>).
- que le ou les mots-clés revienne(nt) plus fréquemment que les autres mots de la page.

5.3. La page d'accueil en frames

Les frames ne sont que modérément appréciés par les moteurs de recherche qui ne les tiennent pas ou difficilement en compte (voir tableau comparatif).

Souvent, emporté par le désir d'en finir au plus tôt avec ce fichier délicat, on s'empresse d'écrire les balises de frame soit : <FRAMESET ROWS="30%,70%"> <FRAME SRC="A.htm"> <FRAMESET COLS="30%,70%"> <FRAME SRC="B.htm"> <FRAME SRC="C.htm"> </FRAMESET> </FRAMESET>.

On en oublie les règles les plus élémentaires du référencement.

On oublie généralement de mettre les balises META, plus que jamais indispensables pour les moteurs de recherche.

On néglige de mettre un titre, à priori inutile car il n'apparaît pas, aux fichiers liés dans le cadre (ici a.htm, b.htm, c.htm). On perd ainsi, des points précieux pour le classement du site parce qu'il y a des pages sans titre et sans référence aux mots-clés dans le titre.

Les moteurs qui ne suivent pas les liens dans les frames se rabattent sur le contenu des balises <NOFRAMES> ... </NOFRAMES>. Avouez que le classique "Désolé, ce site comporte des frames" ne leur apporte pas grand chose à se mettre sous la dent pour votre référencement ! On conseille de reprendre dans les balises <NOFRAMES> un petit texte descriptif du site (avec si possible un rappel des principaux mots-clés). Le "top" est d'y mettre aussi un lien vers les pages principales du site.

5.4. La page d'accueil avec une image

Il est parfois joli de concevoir une page d'accueil composée d'une seule image que le visiteur doit cliquer pour entrer dans le site. Cette situation est assez inconfortable pour le robot à la recherche de mots-clés car il n'a alors aucun texte (et donc de mots) pour faire son référencement...

Il importe dans ce cas extrême de reprendre le mot-clé dans l'attribut ALT de la balise de l'image.

5.5. La page d'accueil avec une image mapée

Pour rappel, les images mapées sont des images découpées en zones. Selon les zones cliquées par le lecteur, celui-ci est dirigé vers une page Html déterminée.

Il est assez risqué, pour obtenir un bon référencement, de prévoir une page d'accueil qui ne comporterait qu'une image mapée comme seul outil de navigation pour votre site. En effet, beaucoup de moteurs de recherche (et pas de moindres comme Hotbot, Excite, Lycos) ne poursuivent pas leur indexation dans les fichiers d'une image mapée. A réserver donc à d'autres usages à l'intérieur de votre site.

Cette caractéristique est reprise dans le tableau comparatif des moteurs de recherche plus loin dans notre étude.

5.6. Et encore

Sont ou seraient bénéfiques pour un bon référencement :

- le fait que votre site soit référencé par d'autres sites.
- le fait que votre site soit mis à jour régulièrement.

5.7. Désolé...

Désolé, mais les trucs suivants ne fonctionnent plus et seraient même pénalisés [spam] par les moteurs de recherche :

- texte invisible dans la même couleur que le fond de la page.
- texte repris en commentaire (balises <!-- ... -->).
- les mêmes mot-clé repris indéfiniment dans la balise <META NAME="keywords">.
- le texte dans l'élément de formulaire <INPUT type="hidden" ...> (éléments cachés).

6. Tableau comparatif

Je n'ai pas su résister à la tentation de reproduire ci-après, un excellent comparatif de la prise en compte des balises META et des particularités de référencement des cinq plus grands moteurs de recherche anglo-saxons. Ce comparatif est bien entendu largement inspiré d'un auteur américain dont voici les mentions de copyright :

Danny Sullivan
Search Engine Watch
<http://searchenginewatch.com/>
Copyright © 1996-98 Mecklermedia

Prise en compte	AltaVista	Hotbot	Excite	Infoseek	Lycos
Longueur du titre (mots)	78	115	70	70	60
Si le titre est absent...	No title	Url de la page	Untitled	Premiers mots de la page	Premiers mots de la page
Longueur de la description	150	249	170-240	395	135-200
Balises META retenues	mots-clés description	mots-clés description	description	mots-clés description	aucunes
Balise ROBOTS	Oui	Oui	Oui	Oui	Oui

Prise en compte	AltaVista	Hotbot	Excite	Infoseek	Lycos
Texte de ALT des images	Oui	Non	Non	Oui	Oui
Texte en commentaire	Non	Oui	Non	Non	Non
Texte invisible	Non [spam]	Non [spam]	Oui	Non [spam]	Non [spam]
Frames	Yes	No	No	No	No
Images mapées	Oui	Non	Non	Oui	Non
Liens dans d'autres sites	Non	Oui	Oui	Non	Oui
Date	Oui	Oui	Non	Oui	Non
Mises à jour	Oui	Oui	Non	Non	Oui

Les balises META
www.ccim.be/ccim328/htmlplus/idxmeta.htm
© copyright 1999

Van Lancker Luc
Rue des Brasseurs, 22
7700 Mouscron
Vanlancker.Luc@ccim.be