

## Lettres accentuées et référencement (2ème partie)

[Retour au sommaire de la lettre](#)

*Nous avons déjà consacré un article, en juillet dernier, au codage des lettres accentuées en vue d'un référencement. Voici quelques indications complémentaires à ce sujet pour faire suite à la première partie de ce dossier et aux questions qu'elle a générées de la part de nos lecteurs.*

Pour faire suite à notre article intitulé "Référencement et lettres accentuées" (lettre R&R de juillet-août 2007), nous avons reçu plusieurs questions de lecteurs nous demandant plus de précisions à ce sujet. Nous allons donc essayer ci-dessous de faire un point plus précis concernant ce domaine si complexe du codage des lettres accentuées pour le référencement.

Tout d'abord, il est important de savoir que les deux codages les plus fréquemment rencontrés sur le Web francophone sont l'ISO-8859-1 - ou l'ISO-8859-15 qui rajoute notamment le signe € à la norme précédente - (pour les pages affichant un jeu de caractère occidental), et UTF-8 (pour des jeux de caractères plus internationaux comme le russe, le chinois, le grec, etc.). Les indications de codage des lettres accentuées et caractères diacritiques (<http://fr.wikipedia.org/wiki/Diacritique>) peuvent alors venir de 3 sources différentes :

- Les données : lorsque le langage PHP, par exemple, récupère des données depuis la base de données Mysql, il précise l'encodage de connexion et l'encodage de réception des données. Par défaut, Mysql stocke les données en latin1 (ISO-8859-1). Ainsi, si l'on désire un texte encodé en UTF-8, il faudra indiquer une configuration spéciale de ses tables et de sa connexion avec les modules écrits en PHP. Attention : les fonctions de modifications de chaîne de caractères pour PHP sont également nativement en ISO et donc incompatibles avec l'UTF-8, il faut donc décoder le tout avant d'utiliser une fonction type `str_replace`.

- Le serveur Internet : Apache propose également un paramètre nommé `AddDefaultCharset` qui prend comme valeur l'encodage par défaut. Si `AddDefaultCharset` est positionné comme de l'UTF-8, le navigateur Internet va forcer la page à être en UTF-8 quel que soit le meta dans la page (voir ci-après)...

- La page Internet : *via* la balise meta adéquate (exemple : `<meta http-equiv="Content-Type" content="text/html; charset=utf-8">`) on peut spécifier l'encodage de la page. La norme veut que quand le navigateur rencontre la balise, il recharge la page et la réinterprète avec le bon *charset*. En XHTML, on peut également rencontrer la balise `<xml version="1.0" encoding="utf-8"?>`. A noter que l'encodage par défaut du XML est l'UTF-8.

Si vous respectez ces 3 étapes au niveau du *charset* et gardez une cohérence entre elles, vous n'êtes pas obligé d'encoder (&acute; pour "é", &ccedil; pour "ç", etc.) les accents dans vos pages HTML, les navigateurs et robots sachant exactement comment lire la page et ses lettres accentuées. Si vous n'êtes pas sûr de l'une des 3 étapes, il vaudra alors mieux encoder les accents en HTML pour que les moteurs de recherche les comprennent sans problèmes.

Nous espérons que ces quelques informations complémentaires vous auront permis de mieux appréhender ce domaine si abscons parfois et que vos lettres accentuées seront ainsi comprises sans aucun problèmes par les moteurs.

*Olivier Andrieu*

*Merci à Vincent Vermersch (Brioude Internet) pour son aide à la rédaction de cet article.*

**Réagissez à cet article sur le blog des abonnés d'Abondance :**  
<http://abonnes.abondance.com/blogpro/2007/10/octobre-2007-lettres-accentues-et.html>