

Référencement : faut-il coder les lettres accentuées dans les urls ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

La problématique du remplacement - ou non - des caractères accentués dans les urls de vos pages est importante. Faut-il supprimer purement et simplement ces caractères ? Les remplacer par leur équivalent non-accentué ? Les coder d'une façon ou d'une autre ? Quelle méthode est à privilégier pour obtenir le meilleur référencement possible ? Réponse de nos tests...

Dans le cadre d'une stratégie de référencement naturel, tous les points permettant de mieux optimiser son site pour les moteurs de recherche sont à prendre en compte. Parmi ceux-ci, le fait d'avoir des URLs "*Search Engine Friendly*" est un point important : proposer dans leur intitulé des mots clés cohérents par rapport au contenu de la page, séparés par des tirets, etc.

Mais il existe également un point crucial pour ces URLs, c'est le traitement des lettres accentuées et caractères diacritiques (<http://fr.wikipedia.org/wiki/Diacritique>) que sont le "é", le "à", le "ç", etc. dans ces urls : faut-il les remplacer par leur équivalent non-accentué ("é" devient "e", "ç" devient "c", etc.) ou les coder d'une façon ou d'une autre pour que les moteurs de recherche les prennent en compte ?

En règle générale, il existe en fait trois façons de régler ce problème :

Solution 1 : On "squeeze" les lettres accentuées dans les urls : "hôtel" devient "htel", "prélèvement" se mue en "prlvement", etc.

Solution 2 : On remplace chaque caractère accentué par son équivalent non accentué.

Exemples :

- L'article "Google Suggest se régionalise" donne l'url :

<http://actu.abondance.com/2009/04/google-suggest-se-regionalise.html>

- L'article "Google lève la voile sur ses datacenters" donne l'url :

<http://actu.abondance.com/2009/04/google-leve-le-voile-sur-ses.html>

Etc.

Cela peut se faire par une bonne gestion de la réécriture d'URL ou plus simplement par une gestion des urls côté serveur...

Solution 3 : On code les lettres accentuées, le plus souvent au travers d'un ensemble de 3 caractères "*quoted printable*" commençant par "%".

"é" devient %E9

"è" devient %E8

Exemples :

[http://www.oscaro.com/plan/m16-BMW+2000-3.2+Coup%E9+\(E9\)-224.htm](http://www.oscaro.com/plan/m16-BMW+2000-3.2+Coup%E9+(E9)-224.htm)

<http://rckiller.blog.radiofg.com/tag/logo%2Btecktonik%2Banim%25E9%2Bgif%2Bsaint%2Brc%2Btck%2Bkiller%2Bteam/>

Quelle solution choisir ?

La solution 1 est clairement à bannir puisque les mots accentués deviennent incompréhensibles aux moteurs de recherche.

Quelle solution, entre la 2 et la 3, faut-il alors choisir ? Pour en avoir le cœur net, nous avons créé deux pages de test sur le site Abondance, répondant aux adresses suivantes (contenant le mot "benebe" ou "bénébé" codé dans l'url et n'apparaissant nulle part ailleurs dans le code HTML) :

<http://www.abondance.com/benebe.html>

<http://www.abondance.com/b%E9n%E9b%E9.html>

Ces deux pages ont été indexées par Google (<http://www.google.fr/search?hl=fr&q=test+site%3Awww.abondance.com>). Rappelons ici que le mot "bénébé" (ou *benebe*) n'apparaît que dans l'URL de la page correspondante et pas son code HTML...

Regardons maintenant comment Google trouve ces pages :

- Sur la requête "benebe site:www.abondance.com" (<http://www.google.fr/search?hl=fr&q=benebe+site%3Awww.abondance.com>), Google trouve la page [benebe.html](#) mais pas la page [b%E9n%E9b%E9.html](#).

- Sur la requête "bénébé site:www.abondance.com" (<http://www.google.fr/search?hl=fr&q=b%C3%A9n%C3%A9b%C3%A9+site%3Awww.abondance.com>) Google trouve la page [benebe.html](#) mais pas la page [b%E9n%E9b%E9.html](#).

Ainsi, que ce soit sur la version accentuée ou non accentuée de la requête "bénébé", Google ne trouve jamais la version contenant les lettres accentuées en "%E9" pour la lettre "é". Ce codage n'est donc pas recommandé.

Remarquons ici que Google trouve bien la page [benebe.html](#) sur la requête *benebe*, ce qui prouve qu'il lit bien l'intitulé de l'URL comme champ de pertinence...

On trouve aussi parfois le codage "%25E9" pour la lettre "é" dans certaines URLs. Exemple (ici le mot "document divisé") :

http://mamrot.gouv.qc.ca/publications/recherche_publications/fiche.asp?noPublication=301&titre=Document+divis%E9&retour=%2Fpublications%2Frecherche_publications%2Ffiche.asp%3FnoPublication%3D301%26titre%3DDocument%2Bdivis%25E9%26retour%3D%252Fpublications%252Frecherche_publications%252Ffiche.asp%253FnoPublication%253D301%2526titre%253DGuide%252Bdivis%2525E9%252Ben%252Bcinq%252Bfichiers%252BPDF%252Bpour%252Bles%252Butilisateurs%252Bd%252526%25252339%25253BInternet%252B%2525E0%252Bbasse%252Bvitesse%2526retour%253D%25252Fpublications%25252Frecherche_publications%25252Fpublications.asp%25253Fmode%25253Dpub%252526noPage%25253D10

Notons que cette page contient dans son TITRE le mot "MAMROT". Donc, la requête "intitle:MAMROT inurl:divisé" devrait la trouver si Google comprend le codage "%25E9" comme un "é". Or cette requête renvoie zéro résultat...

On peut trouver d'autres façons de coder des lettres accentuées dans les URLs. Selon les tests que nous avons pu faire, Google, ne les trouve pas dès qu'on demande leur intitulé accentué dans la recherche...

On prendra comme conclusion de ces quelques tests qu'il est très important de remplacer les lettres accentuées par leur équivalent non accentué dans les URLs des pages. Toute autre traitement de ces caractères semble inefficace pour Google...

Olivier Andrieu
Editeur du site Abondance.com

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2009/04/referencement-faut-il-coder-les-lettres.html>