

RFC 7997 : The Use of Non-ASCII Characters in RFCs

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 16 décembre 2016

Date de publication du RFC : Décembre 2016

<https://www.bortzmeyer.org/7997.html>

Les RFC sont forcément écrits en anglais, qui restera la langue officielle (cf. RFC 7322¹). L'anglais peut s'écrire avec uniquement les caractères ASCII (avec quelques exceptions : "resume" <<https://en.wiktionary.org/wiki/resume>> et "résumé" <<https://en.wiktionary.org/wiki/r%C3%A9sum%C3%A9>> ne sont pas le même mot). Mais on pourra désormais inclure des caractères non-ASCII, par exemple pour le nom des auteurs (chic, je pourrais écrire correctement mon prénom dans les RFC). Cette possibilité permettra aussi aux exemples de protocoles Internet utilisant Unicode (la grande majorité) d'être plus lisibles.

Cette nouvelle possibilité fait partie de celles qu'offre le nouveau format des RFC, décrit dans le RFC 7990. Il n'y a quand même pas d'autorisation générale d'inclure n'importe quel caractère Unicode dans les RFC, à n'importe quel endroit. Le "RFC Editor" pourra toujours refuser tel ou tel caractère, par exemple parce qu'il n'existe pas de police permettant de l'afficher. Et le « non-ASCII » n'est autorisé que dans certains cas, décrits plus loin. La grande majorité du texte sera donc du pur ASCII (RFC 20).

L'encodage de ces caractères sera bien sûr UTF-8.

Il ne suffit pas de proclamer « on a droit à Unicode ». Il faut aussi adapter les outils. Par exemple, notre RFC impose (section 2) que les outils de recherche dans les RFC gèrent correctement la recherche Unicode. (C'est pour traiter le cas des outils imparfaits que le RFC demande aussi que les noms d'auteurs en Unicode soient accompagnés d'une version en ASCII.) Et que le RFC soit affichable correctement sur un bon nombre de plate-formes (d'où la possibilité de rejeter les caractères les plus rares).

Ce problème du repli (vers une version en ASCII pur) est souvent cité dans le RFC. Ainsi, lorsqu'on veut mentionner un caractère Unicode (mettons le thorn islandais), le RFC permet désormais

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc7322.txt>