

RFC 6082 : Deprecating Unicode Language Tag Characters: RFC 2482 is Historic

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 8 novembre 2010

Date de publication du RFC : Novembre 2010

<https://www.bortzmeyer.org/6082.html>

Parfois, il y a des mauvaises idées. Parfois, elles sont même inscrites dans un RFC et sont alors utilisées. Lorsqu'une idée est reconnue comme mauvaise, il faut alors savoir l'abandonner. C'est ce que fait ce RFC 6082¹ qui abandonne officiellement l'idée du RFC 2482 qui fournissait un mécanisme d'indication de la langue utilisée, dans un texte brut en Unicode.

Le RFC 2482 réservait plusieurs caractères Unicode entre U+E0000 et U+E007F à des fins d'**étiquetage**, pour indiquer la langue utilisée pour écrire un texte. Par exemple, U+E0001 U+E006A U+E0061 (caractère "*language tag*", puis caractère "*tag letter j*" puis enfin "*tag letter a*") indiquait un texte en japonais (ja étant le japonais pour les étiquettes de langue <<http://www.langtag.net/>>). Il n'a pas été tellement utilisé en pratique et, aujourd'hui, a été remplacé par des mécanismes d'étiquetage globaux (comme l'en-tête Content-Language: du RFC 3282) pour le texte brut ou bien par des mécanismes plus fins pour les documents structurés (par exemple l'attribut xml:lang pour les textes en XML). Le consortium Unicode <<http://www.unicode.org/>> a officiellement abandonné ce mécanisme (section 16.9 de la norme Unicode 5.2 <<http://www.unicode.org/versions/Unicode5.2.0/>>) et notre RFC 6082 a été fait pour éviter que des implémenteurs tombent sur le RFC 2482 et ne le mettent en œuvre en le croyant toujours d'actualité.

Fin du parcours, donc, pour le RFC 2482, qui entre ainsi dans la catégorie « Intérêt historique uniquement » (catégorie décrite dans la section 4.2.4 du RFC 2026).

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc6082.txt>