

RFC 2277 : IETF Policy on Character Sets and Languages

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 22 janvier 2007

Date de publication du RFC : Janvier 1998

<https://www.bortzmeyer.org/2277.html>

L'internationalisation est un sujet chaud aujourd'hui mais il y a des années que l'IETF s'était penché sur la question. Ce RFC présente les recommandations de l'IETF pour les futurs protocoles, développés après la sortie du RFC.

Notre RFC commence par poser des principes simples : d'abord, les humains parlent des langues différentes et il n'est pas prévu de changer cela. (Dans certaines réunions, on entend parfois des opposants à la diversité linguistique marmonner que ce serait plus simple si tout le monde parlait la même langue et déplorer que la « dictature du politiquement correct » les empêche de le dire; de fait, on n'entend jamais ce genre d'objections publiquement.)

Les protocoles réseaux doivent donc s'adapter à cette situation. Ces protocoles transportent du texte à deux endroits : comme commandes, pour tous les protocoles où les commandes ne sont pas en binaire (c'est le cas de SMTP, avec EHLO, MAIL FROM, etc ou de HTTP avec GET et PUT). Ces commandes ne sont normalement jamais vues des utilisateurs, uniquement des implémenteurs et n'ont donc pas besoin d'être internationalisées. De même, des mots-clés de certains formats comme From: ou Subject: dans le format du courrier (RFC 5322¹) n'ont pas besoin d'être présentés tels quels à l'utilisateur et le protocole n'a donc pas à les traduire.

Mais les protocoles transportent aussi du texte conçu pour être montré aux humains : contenu de la page Web pour HTTP, du courrier pour SMTP, etc. Et, là, le protocole doit respecter le texte, et ne pas imposer qu'il soit écrit en anglais, donc notamment ne pas le restreindre au jeu de caractères ASCII.

Dans tous les cas, le protocole doit permettre (via les étiquettes de langue, normalisées dans le RFC 5646) d'indiquer quelle est la langue d'un texte. Et les protocoles synchrones doivent permettre de

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5322.txt>

négocier la langue utilisée. Dans les cas où cela n'a pas été possible, notre RFC impose, dans sa section 4.5, l'anglais, seul cas où cette langue est privilégiée.

On notera que ces excellents principes laissent de nombreux cas indéterminés. Par exemple, un nom de domaine doit-il être considéré plutôt comme un élément de protocole, qui n'a donc pas besoin d'être internationalisé (c'est plus ou moins l'idée derrière le RFC 4185), ou bien comme un nom prévu pour être connu des utilisateurs et pour lequel l'internationalisation est donc nécessaire (ce qui est le point de vue du RFC 3490)?

Enfin, il y a loin de la coupe aux lèvres pour les anciens protocoles : l'auteur de notre RFC, Harald Tveit Alvestrand, un pilier de l'internationalisation à l'IETF (et l'auteur du premier RFC sur les étiquettes de langue, le RFC 1766) est très actif dans le groupe de travail IETF EAI/IMA <<http://www.ietf.org/html.charters/eai-charter.html>> qui n'a pas encore fini d'internationaliser les adresses de courrier électronique...