

RFC 5335 : Internationalized Email Headers

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 Septembre 2008

Date de publication du RFC : Septembre 2008

<http://www.bortzmeyer.org/5335.html>

Dans l'ensemble des RFC qui décrivent les adresses de courrier électronique internationalisées (c'est-à-dire pouvant utiliser tout le répertoire Unicode), celui-ci se consacre au nouveau format des en-têtes, mettant donc à jour le fameux RFC 2822¹ (en toute rigueur, il n'est mis à jour que par ceux qui participent à l'expérience, cf. section 1.2 du RFC).

Désormais, on peut, si on participe à l'« expérience » (tel est son statut officiel) EAI (*"Email Address Internationalization"*), avoir dans un message un en-tête comme :

From: Stéphane Bortzmeyer <stephane@sources.org>

Oui, avec du vrai UTF-8, y compris dans l'adresse proprement dite, pas seulement dans les commentaires, et encore, à condition de les surencoder selon le RFC 2047 comme c'était le cas avant.

L'éditeur du RFC travaille à TWNIC <<http://www.twnic.tw/>>, le registre chinois de .tw. Les chinois ont, fort logiquement, été particulièrement actifs dans tout le processus EAI. Celui-ci visait à terminer l'internationalisation du courrier électronique, qui avait passé une étape décisive en novembre 1996, avec la sortie du RFC 2045 sur MIME. Mais cette norme MIME n'internationalisait que le corps des messages. EAI permet désormais d'internationaliser également les en-têtes comme From: ou Received: (mais pas, par exemple, Date: ou Message-ID:, qui resteront en ASCII pur, cf. section 4.3; Date:, par exemple, est censé être analysé par le MUA et présenté ensuite à l'utilisateur dans sa langue).

EAI comprend plusieurs RFC, par exemple le RFC 4952 définit le cadre général, le RFC 5336, l'utilisation des adresses Unicode dans SMTP (avec l'extension UTF8SMTP), des futurs RFC qui traiteront le cas de POP ou IMAP et notre RFC 5335, qui se concentre sur le format des messages.

Concrètement, que change donc ce RFC ? La section 4 détaille ce qui est modifié :

1. Pour voir le RFC de numéro NNN, <http://www.ietf.org/rfc/rfcNNN.txt>, par exemple <http://www.ietf.org/rfc/rfc2822.txt>

- La grammaire ABNF qui décrit les en-têtes est changée (sections 4.1 et 4.3) pour accepter des caractères UTF-8, de préférence normalisés en NFC (cf. RFC 5198). Déjà très complexe, cette grammaire devient ainsi réellement difficile.
- La syntaxe des adresses (telles qu'on les trouve dans des en-têtes comme `From:` ou `To:`) est modifiée pour accepter UTF-8. Pour les sites qui utilisent une adresse électronique comme identificateur, comme le fait `Amazon.com`, ce sera un intéressant défi !
- Un nouveau type MIME est créé, `message/global`, pour permettre d'identifier un message EAI (section 4.6).

Parmi les conséquences pratiques de ce changement, il faut noter que les programmeurs qui écrivent des logiciels traitant le courrier doivent désormais gérer Unicode, même s'ils se contentent d'analyser les en-têtes... L'époque du courrier traité avec des scripts shell est bien passée.

La section 5 liste également quelques conséquences pour la sécurité comme le fait qu'un mécanisme d'authentification devra être prêt à gérer plusieurs adresses pour la même personne (car, pendant un certain temps, plusieurs utilisateurs auront à la fois une adresse en Unicode et une en ASCII pur).