

Pourquoi les RFC sont-ils encore en texte brut ?

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 7 novembre 2008

<http://www.bortzmeyer.org/rfc-en-texte-brut.html>

Les RFC, documents qui incluent les normes de l'Internet, sont publiés en texte brut, sans aucune fioriture ou marquage. Pourquoi un tel conservatisme ?

Le RFC 2223¹, le RFC qui normalisait les RFC (il a été remplacé depuis par le RFC 7322), précise que le seul format faisant autorité est la version en texte seul, encodée en ASCII. Malgré de nombreuses tentatives, il n'a pas encore été possible de faire évoluer cette règle. Il y a des bonnes et des mauvaises raisons à cela. (Depuis la parution de cet article, le texte brut est en voie d'abandon, cf. le RFC 7990.)

Le premier RFC, le RFC 1, a été publié en avril 1969. À l'époque, même le jeu de caractères ASCII était une nouveauté. Bien sûr, ce RFC n'a plus d'utilité de nos jours. Mais plusieurs RFC très anciens restent des normes en vigueur. Même si on écarte le RFC 20 qui normalise le texte en ASCII, le plus ancien RFC encore en service semble être le RFC 768, qui normalise UDP, et qui date d'août 1980, il y a vingt-huit ans.

Cela nous donne la première explication : la **permanence**. Une norme, surtout celles utilisées par un réseau aussi essentiel que l'Internet, doit pouvoir être encore lisible des années après, longtemps après que les tendances en matière de format aient changé. Si le RFC 768 avait été écrit avec le traitement de textes à la mode chez les managers de l'époque, il ne serait plus lisible aujourd'hui. Le format « texte brut » (RFC 3676) est le seul qui aie traversé cette époque.

Mais il y a d'autres critères : l'**accessibilité** en est un. Tout le monde, quelle que soit la plate-forme informatique utilisée, peut lire et écrire du texte brut. Cela garantit la participation maximale. Si, au lieu de distribuer les RFC en texte brut, on utilisait le format privé d'un éditeur de logiciels états-unien particulier (au hasard, Microsoft Word puisque beaucoup de blaireaux envoient systématiquement leurs documents dans ce format, sans s'être demandé si leurs destinataires pouvaient le lire), les RFC cesseraient d'être universels. Le texte brut est le PGCD des formats.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc2223.txt>

Il présente également un intérêt pour le programmeur : écrire un programme qui cherche dans les RFC ne nécessite pas de lire des spécifications compliquées d'abord (comme c'est le cas avec l'ultra-difficile PDF, pour lequel il existe très peu de mises en œuvre distinctes). Un simple grep (ou outil équivalent) suffit.

Mais le texte brut en ASCII n'a-t-il pas également des limites ? Bien sûr. Il rend difficile de représenter les noms des auteurs, s'ils sont chinois, allemands ou arabes (par exemple Martin Dürst qui, dans le RFC 3987 doit expliquer comment écrire proprement son nom ou Bill de h[Caractère Unicode non montré²]ra dans le RFC 5023 ce dont son co-auteur s'est plaint <<http://www.tbray.org/ongoing/When/200x/2007/07/24/Atom-is-Finished>>). Il rend pénible les exemples Unicode comme dans les RFC 3490 ou RFC 5198. Il ne permet pas de représenter les formules mathématiques, même relativement simples, comme celles des RFC 5905 ou RFC 5348. Il ne permet pas d'inclure des images autres que l'art ASCII encore que ce point soit discutable : la plupart des images contenues dans les RFC sont des diagrammes assez formels, qui seraient mieux représentés dans un langage spécialisé comme Cosmogol <<http://www.cosmogol.fr/>>. L'ASCII ne permet même pas de mettre en évidence un mot en l'indiquant en gras ; pire, il semble que le "RFC editor" refuse les conventions ASCII courantes qui consistent à entourer le mot de * ou de `_`. Ainsi, distinguer une variable d'un littéral dans une spécification est souvent problématique (alors qu'un format comme Docbook le fait trivialement avec l'élément <code>). Avec l'ASCII pur, il faut deviner que, lorsqu'une norme parle de `http://example.com` que `http` est un littéral et `example.com` une variable, juste un exemple.

Face à ces limites, les tentatives de faire évoluer la règle du RFC 2223 n'ont pas manquées. Ainsi, le RFC 7749 propose un format XML. XML est une norme ouverte, présente depuis suffisamment d'années pour qu'on puisse avoir confiance dans sa stabilité. Si les outils d'édition sont un peu rudes, ce n'est normalement pas un problème pour les auteurs de RFC, qui sont tous des techniciens pointus. Hélas, ce format n'a jamais été adopté officiellement, en partie à cause de l'opposition virulente de Microsoft, en partie par simple conservatisme. Le "RFC editor" (une organisation séparée de l'IETF, sur laquelle l'IETF n'a pas de pouvoir) est en effet très conservateur. C'est dans son travail (qui est de rendre les normes accessibles à tous pendant très longtemps) mais la déformation professionnelle devient parfois excessive. (Depuis, la sortie du RFC 7990 a marqué le début de la fin pour le texte brut.)

À la rigueur, serait-il possible d'autoriser au moins le jeu de caractères Unicode (après tout, « texte brut » n'implique pas forcément ASCII) ? C'est ce qui a été proposé de nombreuses fois, par exemple dans l'"Internet-Draft" `draft-hoffman-utf8-rfcs`, actuellement en cours de discussion. Mais le "RFC editor" continue à s'y opposer. (Les partisans d'Unicode ont finalement eu gain de cause avec le RFC 7997.)

2. Car trop difficile à faire afficher par \LaTeX