

Apostrophes et autres non-lettres dans les noms de domaine

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 20 juin 2012

<https://www.bortzmeyer.org/idn-symbol.html>

On me demande de temps en temps « Puis-je mettre une apostrophe dans un nom de domaine ? » La question a aussi des variantes comme « Et avec les IDN, c'est possible ? »

La réponse est complexe. En deux mots, non, vous n'y arriverez pas. Plus exactement, vous pouvez techniquement, c'est standard (RFC 2181¹, section 11) et ça marche (regardez le nom `dangerous' name.broken-on-purpose` il a un enregistrement TXT). Mais la quasi-totalité des registres ne vous permettront pas d'enregistrer un nom ayant une apostrophe. Pourquoi ? Cela vient d'une vieille règle, spécifiée dans le RFC 1123, section 2.1, qui limite les noms **de machines** (et pas de domaines) à « LDH », c'est-à-dire "*Letters Digits Hyphen*", excluant ainsi des non-lettres comme ', \$, &, ou %. Comme on enregistre en général un nom de domaine pour mettre des machines en dessous, la grande majorité des registres ont choisi la règle conservatrice. C'est ainsi que l'entreprise C&A ne peut pas déposer `c&a.fr` (ou `c&a.com` ou tout autre).

Et, inutile de dire, si vous tentez d'utiliser un nom contenant une apostrophe, vous allez planter plusieurs programmes mal écrits, et peut-être trouver une injection SQL <<https://www.bortzmeyer.org/sql-injection.html>> ou deux... (Si vous voulez planter les programmes de vos collègues, regardez l'enregistrement MX de `dangerousrecord.broken-on-purpose.generic-nic.net`...) Ce qui est techniquement légal n'est pas forcément utilisable en pratique.

Mais avec les IDN (RFC 5890), est-ce que les choses ne s'arrangent pas ? Non. La norme actuelle (voir le RFC 5892) limite les caractères autorisés aux lettres et chiffres. Les symboles restent interdits (interdits dans la norme technique et pas simplement par les politiques de registre). Pas de nom rigolo comme `I[Caractère Unicode non montré2]NY.us`. Les motivations pour cela sont diverses, allant de la crainte pour certains programmes, comme ceux mentionnés plus haut, à des vagues peurs sur des

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc2181.txt>

2. Car trop difficile à faire afficher par L^AT_EX

soi-disant « caractères dangereux » (l'ICANN avait fait fortement pression en ce sens). Les convertisseurs IDN <<http://mct.verisign-grs.com/conversiontool/convertServlet>> vous diront donc des choses comme `Error:: Disallowed/Unassigned 27` (U+27 étant le point de code de l'apostrophe).

C'est très gênant pour certaines langues. Par exemple, en hawaïen, le nom de l'archipel où on parle cette langue est « Hawai[Caractère Unicode non montré]i » et cette apostrophe Unicode (U+2018) n'est pas légale. Pas moyen, donc, d'écrire `hawaii.gov` correctement. Même problème pour le breton, la ville de Penmarc'h ne peut pas avoir `penmarc'h.fr`, et même le futur `.bzh` ne lui permettra pas.

Merci à Mathieu Crédou pour la discussion et pour avoir attiré mon attention sur le cas breton <<http://www.actualitte.com/actualite/patrimoine-education/scolarite-france/l-informati>
htm>. Jonathan L. <<https://twitter.com/TehHarry>> me fait remarquer à juste titre qu'il existe [Caractère Unicode non montré].ws et [Caractère Unicode non montré].net, par exemple. C'est exact. Cela peut être le résultat d'une politique de registre plus laxiste. Ou tout simplement une survivance de l'ancienne norme IDN (celle du RFC 3490), qui n'interdisait pas explicitement les symboles.