

Au travail sur le .[Caractère Unicode non montré *][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré]

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 28 avril 2010. Dernière mise à jour le 21 juin 2010

<https://www.bortzmeyer.org/point-tounes.html>

Dans le cadre du programme d'enregistrement <<https://www.bortzmeyer.org/idn-racine.html>> d'IDN dans la racine du DNS, la Tunisie (via l'ATI) a demandé <<https://www.icann.org/en/topics/idn/fast-track/>> la création d'un .[Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré], équivalent en caractères arabes du .tn. Le travail se fait en partenariat avec l'AFNIC <<http://www.afnic.fr/actu/nouvelles/251/1-agence-tunisienne-d-internet-choisit-les-solutions-de-l-afn>> d'où mon implication.

Pourquoi des IDN en écriture arabe sont-ils nécessaires? D'une manière générale, tout humain a droit d'utiliser les caractères auquel il est habitué pour écrire des noms de domaine. Mais c'est encore plus important en arabe à cause de l'écriture de droite à gauche. Un FQDN mixte (caractères latins et arabes) serait très déroutant pour l'utilisateur, une partie allant de gauche à droite et l'autre en sens inverse. D'autant plus que le point, utilisé pour séparer les composants d'un nom de domaine, n'a pas de directionnalité dans l'algorithme bidi d'Unicode ("*Unicode Standard Annex #9*" <<http://unicode.org/reports/tr9/>>).

Maintenant, les tunisiens ont-ils vraiment « besoin » des domaines en arabe? On pourrait croire que non puisque la quasi-totalité des tunisiens qui savent lire et écrire maîtrisent autant l'alphabet latin que l'arabe. Mais il faut aussi penser aux arabes non tunisiens qui consulteront des sites Web en .[Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré]. Et, d'une manière générale, l'administration tunisienne évolue peu à peu vers une plus grande « arabisation », justifiant que l'adresse (l'URL) utilise l'écriture arabe, comme le contenu.

*Car trop difficile à faire afficher par L^AT_EX

Ce projet a déjà fait l'objet d'articles <<http://www.tekiano.com/net/web-2-0/2-7-2073/tunisie-le-tounes-bientot-disponible.html>>. Actuellement, la chaîne de caractères [Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré] (xn--pgbs0dh en Punycode, cf. RFC 3492¹) a été acceptée <<https://www.icann.org/en/topics/idn/fast-track/string-evaluation-completion-t-en.htm>> par l'ICANN, il reste :

- À obtenir qu'elle soit mise dans la racine, ce qui va prendre du temps. Cela dépend de l'ICANN et du gouvernement états-unien, qui a un contrôle exclusif sur le contenu de la racine.
- À enregistrer des noms en écriture arabe sous . [Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré]. Les caractères acceptés n'ont pas été annoncés mais il serait logique qu'ils soient ceux du RFC 5564, qui sont ceux de la langue arabe (je rappelle que langue et écriture sont deux choses différentes; à noter au passage qu'il y a très peu de berbérophones en Tunisie, ce qui simplifie la question).

Tiens, au fait, comment écrire le nom de ce domaine? . [Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré] ou bien [Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré][Caractère Unicode non montré]. ? Le problème est d'autant plus complexe que ce n'est pas réglé en caractères latins non plus. Le marketing écrit plutôt .FR et les techniciens FR. (le point représentant la racine, qui est à droite, et c'est en outre la syntaxe qu'accepte dig). Mais l'écriture de droite à gauche ajoute un nouveau niveau de perplexité.

Et les problèmes techniques? En fait, en utilisant des techniques récentes (langage de programmation Python, SGBD PostgreSQL) qui sont Unicode depuis longtemps, il n'y a pas grand'chose à faire. Unicode est compliqué conceptuellement mais, en pratique, le programmeur qui choisit les bons outils n'a guère de travail supplémentaire. Toutefois, tout n'est pas encore parfait et, comme on peut le voir sur cette copie d'écran (en ligne sur <https://www.bortzmeyer.org/files/gnome-terminal-requetes-sql.png>), si le terminal a bien accepté l'UTF-8, il a par contre affiché le texte arabe dans le mauvais sens... (Avertissement : il ne s'agit évidemment pas des noms de domaines réels, c'est une base de test.)

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3492.txt>