

RFC 3743 : Joint Engineering Team (JET) Guidelines for Internationalized Domain Names (IDN) Registration and Administration for Chinese, Japanese, and Korean

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 3 janvier 2008

Date de publication du RFC : Avril 2004

<https://www.bortzmeyer.org/3743.html>

Une fois définie une norme pour les noms de domaines internationalisés, il reste à régler quelques problèmes liés à leur enregistrement. Par exemple, le chinois s'écrit couramment avec deux écritures, la traditionnelle et la simplifiée. Faut-il permettre l'enregistrement de deux noms de domaines identiques, ne se différenciant que par cette écriture ? Notre RFC propose un cadre pour spécifier de telles règles et crée la notion de **lot** ("*package*"), un ensemble de domaines qu'IDN considérerait différents mais que le registre préfère traiter comme équivalents, et ne pouvant donc faire l'objet que d'un seul enregistrement.

IDN, normalisé dans le RFC 3490¹, permet d'exprimer des noms de domaine en Unicode. Une canonicalisation est effectuée par le mécanisme nameprep (RFC 3491) et certaines chaînes de caractères Unicode (comme "*Stral*[Caractère Unicode non montré² *le*" et "*strasse*") sont donc identiques pour nameprep. Il ne peut donc y avoir qu'un seul nom de domaine pour ces deux chaînes.

Mais d'autres chaînes Unicode sont considérées comme distinctes par nameprep, et donc peuvent être enregistrées séparément, alors qu'un locuteur de telle ou telle langue pourrait les voir comme identiques. C'est le cas de deux mots en chinois en écriture traditionnelle ou simplifiée, par exemple. De même, en français, certains pourraient considérer que *congres* et *congrès* doivent être considérés comme identiques, car ne se différenciant que par un accent.

Notre RFC, qui n'a pas le statut de norme, n'essaie pas de décrire une telle politique pour toutes les langues du monde. Il fournit simplement un cadre permettant de décrire de telles politiques, en mettant

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3490.txt>

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

l'accès sur les questions spécifiques des langues asiatiques, souvent désignées par le sigle CJC. Il a été co-écrit par les registres de noms de domaine de Chine, de Taïwan, de Corée et du Japon, regroupés dans une structure ad-hoc, le JET. Le RFC 4290 étendra ensuite cette méthode aux autres écritures du monde.

Avant de voir la technique, il est bon de se rappeler que l'IETF n'a aucune légitimité pour décider des politiques d'enregistrement suivies dans `.tw` ou `.kr`. Il ne peut s'agir ici que de recommandations, telles qu'exprimées dans la note de l'IESG qui précède le RFC, qui incite les autres registres à suivre le même chemin, en rappelant (à regret?) que l'IESG ne peut pas les y obliger.

En quoi consiste le mécanisme de ce RFC? Simplement en la définition d'une table des **variantes** de chaque caractère (sections 2.1.11 et 3.2.1). Toute chaîne peut donc être **canonicalisée** en étant réduite à la **variante préférée** (section 2.1.13). Les chaînes qui ont la même forme canonique font partie du même **lot** ("*package*", section 2.1.18, nommé "*bundle*" dans le RFC 4290) et ne peuvent donc pas faire l'objet d'enregistrements distincts. Seule la variante préférée est enregistrée. Les autres sont réservées.

La section 3 du RFC détaille ce mécanisme et traite de certains cas particuliers, comme la suppression d'un nom de domaine (section 3.3).

Appliqué au français, avec une table de variantes telle que celle proposée pour `.fr` <<https://www.bortzmeyer.org/4290.html>>, cette méthode ferait de `congres` et `congrès`, mais aussi de `côngrès` et `congrês` les membres d'un même lot.

Notre RFC ne répond pas à la question posée plus haut sur l'équivalence des écritures chinoises traditionnelles et simplifiées. Comme expliqué en section 4, cette tâche est laissé à des tables de variantes, dont certaines sont déposées sur le registre (non obligatoire) <<https://www.iana.org/assignments/idn/registered.htm>> de l'IANA.

La section 5 du RFC décrit une proposition de syntaxe pour les tables de variantes, qui sera reprise dans le RFC 4290.