

RFC 4690 : Review and Recommendations for Internationalized Domain Names (IDNs)

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 3 octobre 2006

Date de publication du RFC : Septembre 2006

<https://www.bortzmeyer.org/4690.html>

Les IDN ont toujours déclenché des passions et suscité des oppositions vigoureuses, notamment de la part de ceux qui regrettent que tout le monde n'écrive pas en anglais. Autant dire que ce RFC, pourtant écrit de manière très prudente et donc les aspérités ont été gommées, ne va pas passer inaperçu. Il dresse un bilan plutôt négatif des IDN et exprime de nombreuses craintes.

Après un rappel des normes existantes (les IDN ont été normalisés à l'origine dans le RFC 3490¹) et du vocabulaire du domaine (beaucoup de gens, à commencer par l'ICANN confondent encore langue et écriture), ainsi que la littérature existante (l'ICANN a produit des documents sans intérêt, uniquement pour tenter de faire croire qu'elle avait un rôle dans l'internationalisation de l'Internet), le RFC consacre sa section 2 à lister plusieurs problèmes.

L'équivalence de deux noms dépend du langage : est-ce que `cafe.fr` et `café.fr` sont le même domaine ? D'autres peuples d'Europe accordent une plus grosse importance aux accents que les français.

Certaines langues peuvent s'exprimer en plusieurs écritures, par exemple l'azéri. Le même mot en différentes écritures sera considéré comme différent par le DNS.

La canonicalisation des caractères par IDN peut ne pas correspondre aux attentes des utilisateurs (`stra[Caractère Unicode non montré2]e.de` est canonicalisé en `strasse.de`).

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

La forme imprimée de l'IDN peut ne pas être évidente à lire. C'est le problème connu à l'IETF sous le nom de "problème du côté des bus". Une publicité comprenant un URL, chose banale aujourd'hui, pourra t-elle être lue si elle se trouve sur le côté d'un bus qui passe, sachant que de nombreux caractères se ressemblent ?

De même, certains caractères peuvent être facilement confondus (comme `.py` - le Paraguay et `. [Caractère Unicode non montré] [Caractère Unicode non montré]` - la Russie, écrit en russe avec les caractères cyrilliques). Cela peut mener à des problèmes de sécurité, par exemple en cas d'hameçonnage.

Une longue section 3 est consacrée à un problème urgent : la migration vers de nouvelles versions d'Unicode. IDN est limité à la version 3.2, mais la 5.0 est déjà sortie. Un manque de stabilité de certaines normes Unicode fait qu'IDN se limite donc à une vieille version d'Unicode, sans moyen simple de mettre à jour.

De ces questions, notre RFC tire les conclusions suivantes :

- Nécessité de revoir la norme IDN (une révision partielle a eu lieu depuis dans le RFC 5891), notamment en passant d'un modèle "Tous les caractères sont autorisés sauf" à un modèle "Sont autorisés les caractères suivants :",
- Étudier les solutions non-DNS (par exemple avec un couche supplémentaire dans l'interface utilisateur pour traduire `.com` en chinois),
- Restrictions sur l'usage de multiple écritures dans un seul nom.

D'autres conclusions sont explicitement dirigées plutôt vers l'ICANN (notons que le RFC fait comme si l'ICANN était responsable des politiques des registres) :

- Politiques d'enregistrement plus strictes chez les registres,
- Politique d'introduction des IDN dans la racine du DNS (ce qu'on nomme souvent les IDN.IDN par exemple `café.ré` pour l'île de la Réunion).

Le RFC appelle aussi à reconsidérer le rôle (limité) du DNS et à considérer que, peut-être, les efforts d'internationalisation pourraient porter sur des solutions non-DNS, par exemple des moteurs de recherche.

Si ce RFC est très prudent dans ses recommandations (tout est présenté sous forme de question, il n'y a pratiquement pas de directive concrète qui pourrait fâcher), il appartient quand même à la vaste catégorie des textes s'inquiétant, en général assez hypocritement, des problèmes liés à IDN. La plupart de ces textes sont émis par des organismes qui ne veulent pas de l'internationalisation de l'Internet (on peut lire un bon exemple <http://www.iccwbo.org/policy/ebitt/iccibec/index.html>). En fait, la plupart de ces textes, tout comme notre RFC, oublie deux choses :

- La complexité ne vient pas d'IDN ou d'Unicode, elle vient des langues humaines. Cells-ci sont très complexes, point. On peut le regretter et se dire que cela serait plus simple si tout le monde parlait anglais ou lojban mais ce n'est pas le cas. L'IETF et les autres SDO doivent faire avec.
- La plupart de ces problèmes sont marginaux. Ils ne se produisent que dans des cas rares, avec des écritures peu courantes. Retarder les IDN à cause de problèmes aussi artificiels serait retarder une évolution de l'Internet vers plus de prise en compte des autres cultures.

Ce RFC, comme beaucoup d'autres textes et discours analogues est, quoique écrit dans le style IETF, de la même eau : du FUD.