

IDNAbis

Stéphane Bortzmeyer
AFNIC
bortzmeyer@nic.fr

2 novembre 2010

Rappel : motivations pour IDN

Permettre d'exprimer des noms de domaine dans toutes les écritures du monde.

« L'usage de l'Internet doit être exactement aussi facile pour un chinois que pour un états-unien. » (Ted Hardie)

Rappel : fonctionnement d'IDN version 1

1. Canonicalisation (normalisation) du terme Unicode par nameprep
2. Encodage en Punycode (ACE) : xn--pgbs0dh.
3. Utilisation de l'ACE entre les serveurs de noms

Insatisfactions avec IDNA 1

1. Dépendance par rapport à Unicode 3.2 (cas du Tifinagh)
2. Étape de normalisation qui en faisait trop ou trop peu
3. Erreurs BIDI pour des écritures comme le Thaana
4. Réglements de compte
5. Désir de l'ICANN de retarder les IDN

En pratique, le hameçonnage dans le monde réel ne dépend quasiment jamais d'une confusion entre homographes.

Les messages de hameçonnage réels utilisent des adresses IP, des noms vaguement similaires (`secure-societegenerale.com` vs. `secure.societe-generale.com`) et ne cherchent pas à tromper M. Ali : celui-ci ne lit pas les URL...

Cinq nouveaux RFC

- ▶ RFC 5890 « Internationalized Domain Names for Applications (IDNA) : Definitions and Document Framework », donne les définitions des termes essentiels comme les nouveaux U-label (forme Unicode d'un nom légal) et A-label (forme Punycode d'un nom légal),
- ▶ RFC 5894 « Internationalized Domain Names for Applications (IDNA) : Background, Explanation, and Rationale », explique et justifie, le projet IDNAbis et ses concepts ; ce RFC n'est pas une norme, il n'est là que pour information,

Cinq nouveaux RFC, fin

- ▶ RFC 5892 « The Unicode code points and IDNA », spécifie l'algorithme utilisé pour déterminer si un caractère est légal en IDNAbis, illégale ou bien si sa légalité dépend du contexte,
- ▶ RFC 5893 « Right-to-left scripts for IDNA », expose les règles pour les noms de domaine dont une partie s'écrit de droite à gauche (par exemple en hébreu),
- ▶ RFC 3492, « Punycode : A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA) », est maintenu.

Les nouveautés en un transparent

- ▶ Le protocole est désormais indépendant de la version d'Unicode : tout changement dans Unicode est automatiquement disponible.
- ▶ Les caractères de ponctuation et les symboles sont désormais presque tous exclus.
- ▶ Il n'y a plus d'étape de normalisation standard.
- ▶ Le modèle de sélection des caractères autorisés est passé de « entièrement manuel, caractère par caractère » à « essentiellement algorithmique, fondé sur les propriétés Unicode ».

Désormais, un caractère est interdit s'il n'est pas explicitement autorisé.

Nouveautés : plus de canonicalisation standard

Le RFC 3491 (nameprep) n'est plus utilisé. Chaque application est désormais libre d'effectuer la correspondance entre ce qu'a tapé ou sélectionné l'utilisateur et l'IDN envoyé sur le réseau.

Cela permet à l'application de tenir compte du contexte, par exemple de la langue.

Il existe des canonicalisations possibles (mais non obligatoires) : RFC 5895, « Mapping Characters for Internationalized Domain Names in Applications » ou comme le TR #46 d'Unicode.

Beaucoup de noms légaux (mais peu utilisés en pratique) sont devenus illégaux. Nécessité d'une période de transition.

Des noms identiques sont devenus différentes (strasse.de et straÙe.de). DENIC fera un lever de soleil avec priorité à l'ancien titulaire `http:`

`//www.denic.de/en/denic-in-dialogue/news/2985.html`.

Premiers problèmes

La sortie de Unicode 6.0 le 11 octobre a été le premier test du nouvel algorithme.

Deux caractères voyagent des Interdits vers les Autorisés. OK.

Un caractère, U+19DA, le zéro du Tai Lue, voyage en sens inverse. Faut-il l'ajouter aux exceptions manuelles, pour qu'il reste autorisé ?

Donner la priorité à la stabilité ou bien à la fidélité à Unicode ?

RFC tout nouveaux, profonds changements en interne.

Mais, en pratique, peu de changements pour les registres. Même préfixe, la plupart des noms légaux restent légaux et les illégaux le restent.

Encore très peu d'implémentations.

<http://www.bortzmeyer.org/idnabis.html>