

# Un exemple de nom de domaine inhabituel (IDN et émojis)

Stéphane Bortzmeyer  
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 12 mars 2021

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

---

C'est vendredi, bientôt le week-end, détendons-nous avec un nom de domaine rigolo, [Caractère Unicode non montré <sup>1</sup> ].ws. Oui, vous avez déjà vu des noms de domaines avec des émojis. Mais celui-ci a une particularité.

Un petit retour en arrière s'impose. D'abord, contrairement à une légende répandue, le DNS, le protocole qui résout les noms de domaine en informations concrètes, ce protocole n'est pas limité à l'ASCII, encore moins au sous-ensemble LDH ("*Letters-Digits-Hyphen*") d'ASCII. On peut mettre n'importe quel octet dans un nom de domaine. Toutefois, le faire peut entraîner des surprises, parce qu'il n'existe pas d'encodage standard (si le DNS était créé aujourd'hui, UTF-8 serait l'encodage standard mais, à l'époque, ce n'était pas aussi clair) et surtout parce qu'il n'y a pas de règles de canonicalisation (depuis, le RFC 5198<sup>2</sup> fournirait certainement une solution mais, là encore, c'est trop tard). Et, surtout, pas mal d'applications (mais, répétons-le, pas le DNS lui-même) seraient surprises de recevoir des noms qui ne sont pas en ASCII. C'est pour ces raisons (que j'ai détaillées dans un autre article <<https://www.bortzmeyer.org/pourquoi-idn-et-pas-un-dns-unicode.html>>) que, si on veut de l'Unicode dans les noms de domaine (que l'on nommera alors IDN), on doit utiliser la norme décrite dans les RFC 5890 et suivants. Cette norme impose notamment :

- Une restriction des caractères autorisés, décrite dans le RFC 5892, en gros les lettres (toutes les lettres, pas uniquement celles d'ASCII), les chiffres et le tiret. Les symboles comme &, \$ ou les émojis ne sont pas autorisés.
- Un encodage standard, Punycode, normalisé dans le RFC 3492. C'est ainsi que `gêmeaux.bortzmeyer.org` se retrouve encodé en `xn--gmeaux-bva.bortzmeyer.org` et [Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ].[Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ][Caractère Unicode non montré ] est encodé `xn--rlcuo9h.xn--wlcmbh10gb5a8kc.xn--xkc2a13hye2a`.

---

1. Car trop difficile à faire afficher par L<sup>A</sup>T<sub>E</sub>X

2. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5198.txt>