

Version 16 d'Unicode

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 14 septembre 2024

<https://www.bortzmeyer.org/unicode-16-0.html>

Le 10 septembre est sortie la version 16 d'Unicode <<https://blog.unicode.org/2024/09/announcing-unicode-standard-version-160.html>>. Une description officielle des principaux changements est disponible <<https://www.unicode.org/versions/Unicode16.0.0/>> mais voici ceux qui m'ont intéressé particulièrement. (Il n'y a pas de changement radical.)

Pour explorer plus facilement la grande base Unicode, j'utilise un programme qui la convertit en SQL <<https://www.bortzmeyer.org/unicode-to-sql.html>> et permet ensuite de faire des analyses variées. Faisons quelques requêtes SQL :

```
ucd=> SELECT count(*) AS Total FROM Characters;
total
-----
155063
```

Combien de caractères sont arrivés avec la version 16?

```
ucd=> SELECT version,count(version) FROM Characters GROUP BY version ORDER BY version::float;
...
13.0 | 5930
14.0 | 838
15.0 | 4489
15.1 | 627
16.0 | 5185
```

5 185 nouveaux caractères, c'est pas mal, la plus grosse addition depuis la version 13.0. Quels sont ces nouveaux caractères?

```

ucd=> SELECT To_U(codepoint) AS Code_point, name FROM Characters WHERE version='16.0' ORDER BY Codepoint;
code_point | name
-----+-----
U+A7CB     | LATIN CAPITAL LETTER RAMS HORN
U+A7CC     | LATIN CAPITAL LETTER S WITH DIAGONAL STROKE
...
U+10D50    | GARAY CAPITAL LETTER A
U+10D51    | GARAY CAPITAL LETTER CA
U+10D52    | GARAY CAPITAL LETTER MA
...
U+11BC2    | SUNUWAR LETTER EKO
U+11BC3    | SUNUWAR LETTER IMAR
U+11BC4    | SUNUWAR LETTER REU
...
U+1346E    | EGYPTIAN HIEROGLYPH-1346E
U+1346F    | EGYPTIAN HIEROGLYPH-1346F
U+13470    | EGYPTIAN HIEROGLYPH-13470
...
U+1CC15    | LOGIC GATE OR
U+1CC16    | LOGIC GATE AND
...
U+1CC4A    | FLYING SAUCER WITH BEAMS
U+1CC4B    | FLYING SAUCER WITHOUT BEAMS
...
U+1CC4E    | ALIEN SQUID OPEN TENTACLES
U+1CC4F    | ALIEN SQUID CLOSED TENTACLES
...
U+1CC60    | LEFT-POINTING ATOMIC BOMB
U+1CC61    | UP-POINTING ATOMIC BOMB
...
U+1CE20    | LARGE TYPE PIECE CROSSBAR WITH LOWER STEM
U+1CE21    | LARGE TYPE PIECE UPPER HALF VERTEX OF M
U+1CE22    | LARGE TYPE PIECE DIAGONAL LOWER LEFT
...

```

Cette version amène en effet plusieurs nouvelles écritures. C'est le cas du Sunuwar, par exemple. On voit aussi arriver le Garay. Cette dernière écriture ne semble d'ailleurs pas avoir tellement eu de succès, le nom de domaine du site officiel, `garay-ecriture.com` a disparu en 2012. Parmi les nouveautés, beaucoup d'hieroglyphes égyptiens, et l'habituelle arrivée d'emojis comme les différentes variantes d'extra-terrestres, avec ou sans tentacules (bon, d'accord, ceux-là ne sont pas officiellement des emojis, ils ont été enregistrés via une autre voie mais ils seront sans doute utilisés comme des emojis).

Revenu du passé, le jeu de caractères des machines Sharp MZ est désormais inclus. Cela apporte notamment les symboles des circuits électroniques.

Si vous avez les bonnes polices de caractères, vous allez pouvoir voir quelques exemples (sinon, le lien mène vers Uniview <<https://r12a.github.io/uniview/>>). Voici par exemple le grand S barré [Caractère Unicode non montré ¹] <<https://r12a.github.io/uniview/?char=A7CC>>, la première lettre de l'alphabet garay [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=10D4A>>, celle du sunuwar [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=11BC0>>, un des nouveaux et nombreux hieroglyphes [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=13460>>, une diode [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1CC0E>> et un extra-terrestre [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1CC4C>>.

Sur le même sujet, je vous recommande l'article d'Ysabeau sur LinuxFr <<https://linuxfr.org/news/unicode-en-version-16-0-0-le-plein-de-hieroglyphes-egyptiens-et-de-symboles-info>>

1. Car trop difficile à faire afficher par \LaTeX