

Version 15 d'Unicode

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 15 septembre 2022

<https://www.bortzmeyer.org/unicode-15-0.html>

Le mardi 13 septembre est sortie la version 15 d'Unicode <<https://home.unicode.org/announcing-the-unicode-15-0/>>. Une description officielle des principaux changements est disponible <<https://www.unicode.org/versions/Unicode15.0.0/>> 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
-----
149251
```

Combien de caractères sont arrivés avec la version 15?

```
ucd=> SELECT version,count(version) FROM Characters GROUP BY version ORDER BY version::float;
...
12.0 | 554
12.1 | 1
13.0 | 5930
14.0 | 838
15.0 | 4489
```

4489 nouveaux caractères, c'est pas mal (la version 14 <<https://www.bortzmeyer.org/unicode-14-0.html>> était plus calme). Quels sont ces nouveaux caractères?

```

ucd=> SELECT To_U(codepoint) AS Code_point, name FROM Characters WHERE version='15.0' ORDER BY Codepoint;
code_point | name
-----+-----
...
U+1D2C0    | KAKTOVIK NUMERAL ZERO
...
U+1E4D4    | NAG MUNDARI LETTER ONG
...
U+1F776    | LUNAR ECLIPSE
...
U+1F77C    | MAKEMAKE
U+1F77D    | GONGGONG
...
U+1FACF    | DONKEY

```

Cette version amène en effet des écritures nouvelles comme le Nag Mundari, ou les chiffres de Kaktovik.

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 la lettre ETT du Nag Mundari [Caractère Unicode non montré ¹] <<https://r12a.github.io/uniview/?char=1E4E9>> et le chiffre de Kaktovik 2 [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1D2C2>> Il y a également de nouveaux symboles, notamment liés à l'astronomie <<http://blog.unicode.org/2022/05/out-of-this-world-new-astronomy-symbols.html>>, comme l'occultation [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F775>> ou la petite planète Orcus [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F77F>>. En plus anecdotique, on a le sans-fil [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F6DC>>, les maracas [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1FA87>> ou l'élan [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1FACE>>. Ce dernier a d'ailleurs eu droit au point de code, U+1FACE, une utilisation amusante de l'hexadécimal <<https://www.bortzmeyer.org/ipv6-rigolo.html>>.

Tiens, d'ailleurs, combien de caractères Unicode sont des symboles (il n'y a pas que les emojis parmi eux, mais Unicode n'a pas de catégorie « emoji ») :

```

ucd=> SELECT count(*) FROM Characters WHERE category IN ('Sm', 'Sc', 'Sk', 'So');
count
-----
7770

```

Ou, en plus détaillé, et avec les noms longs des catégories :

```

ucd=> SELECT description, count(category) FROM Characters, Categories WHERE Categories.name = Characters.category;
description | count
-----+-----
Modifier_Symbol | 125
Other_Symbol | 6634
Math_Symbol | 948
Currency_Symbol | 63

```

1. Car trop difficile à faire afficher par \LaTeX