

Version 13 d'Unicode

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 mars 2020

<https://www.bortzmeyer.org/unicode-13-0.html>

Ce mercredi 11 mars est sortie la version 13 d'Unicode <<https://home.unicode.org/announcing-the-unicode>>. Une description officielle des principaux changements est disponible <<https://www.unicode.org/versions/Unicode13.0.0/>> 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
-----
143924
```

Combien de caractères sont arrivés avec la version 13?

```
ucd=> SELECT version,count(version) FROM Characters GROUP BY version ORDER BY version::float;
...
10.0 | 8518
11.0 | 684
12.0 | 554
12.1 | 1
13.0 | 5930
```

5930 nouveaux, le rythme de création repart, après une version 12 <<https://www.bortzmeyer.org/unicode-12-0.html>> qui était très modérée. Quels sont ces nouveaux caractères?

```

ucd=> SELECT To_U(codepoint) AS Codepoint, name FROM Characters WHERE version='13.0';
codepoint | name
-----+-----
...
U+1FAD5   | FONDUE
...
U+10E80   | YEZIDI LETTER ELIF
U+10E81   | YEZIDI LETTER BE
U+10E82   | YEZIDI LETTER PE
...
U+1F9A3   | MAMMOTH
U+1F9A4   | DODO
U+1F9AB   | BEAVER
...
U+1FBA0   | BOX DRAWINGS LIGHT DIAGONAL UPPER CENTRE TO MIDDLE LEFT
U+1FBA1   | BOX DRAWINGS LIGHT DIAGONAL UPPER CENTRE TO MIDDLE RIGHT
U+1FBA2   | BOX DRAWINGS LIGHT DIAGONAL MIDDLE LEFT TO LOWER CENTRE

```

On trouve donc le lot habituel <http://blog.unicode.org/2020/01/unicode-emoji-130-now-final.html> de nouveaux émojis, comme la fondue ou le mammouth. Parmi ces nouveaux émojis, notons le drapeau trans. Comme souvent dans Unicode, pour limiter l'explosion du nombre d'émojis, le drapeau trans n'est pas en un seul caractère mais s'obtient par une combinaison, U+1F3F3 U+FE0F U+200D U+26A7 U+FE0F, où U+1F3F3 est le drapeau blanc et U+26A7 le signe transgenre. (Notons que tout le monde n'est pas d'accord sur cet ajout permanent d'émojis pour chaque groupe <http://www.unicode.org/L2/L2019/19084-trans-flag.pdf> de plus en plus spécifique.) Toujours dans les émojis, on notera l'arrivée des symboles liés aux communs comme U+1F16D, le symbole des "Creative Commons". Il avait été refusé pendant longtemps comme symbole politique, alors que le copyright a son U+A9 depuis la version 1 d'Unicode...

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
-----
7564

```

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 | 123
Other_Symbol | 6431
Math_Symbol | 948
Currency_Symbol | 62

```

Pour conclure sur les émojis, je vais citer Benjamin Bayart : « Alors, sans vouloir manquer de respects aux amateurices d'émojis, franchement, savoir si avec une série de caractères combinants on peut représenter une emoji "Courgette au curry qui fait du tambour avec un poireau bleu et un chapeau pointu", j'en comprend le côté amusant, mais c'est pas ça, le sens d'Unicode. Et ça m'exaspère qu'on laisse toujours le côté fantastique de cet outil [Unicode] dans l'ombre, pour montrer son côté le plus complètement inutile. »

Il y a surtout dans cette version 13 des nouvelles écritures comme le yezidi (qui renait, peut-être à cause de l'intérêt porté à leur lutte contre Daech), et plein de caractères pour faire de l'« art Unicode » en dessinant avec des caractères. Par contre, l'écriture maya n'est pas encore intégrée <<http://blog.unicode.org/2020/01/unicode-maya-hieroglyph-project.html>>.

Si vous avez les bonnes polices de caractères, voici les caractères pris en exemple plus haut : [Caractère Unicode non montré ¹]<<https://r12a.github.io/uniview/?char=1FAD5>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10E80>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10E81>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10E82>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9A3>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9A4>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9AB>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1FBA0>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1FBA1>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1FBA2>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=26A7>> (ce dernier étant dans Unicode depuis la version 4), [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F16D>>... (Si vous n'avez pas les bonnes polices, chaque lettre est un lien vers Uniview <<https://r12a.github.io/uniview/>>.) Pour le drapeau trans, voici la combinaison qui est censée l'afficher, si vous avez logiciel et police récents : [[Caractère Unicode non montré]][Caractère Unicode non montré]][Caractère Unicode non montré]][Caractère Unicode non montré]][Caractère Unicode non montré]].

1. Car trop difficile à faire afficher par L^AT_EX