

Version 12 d'Unicode

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 8 octobre 2019

<https://www.bortzmeyer.org/unicode-12-0.html>

En mars dernier est sortie la version 12 d'Unicode <<http://blog.unicode.org/2019/03/announcing-unicode.html>>. Une description officielle des principaux changements est disponible <<https://www.unicode.org/versions/Unicode12.0.0/>> 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
-----
137994
```

Combien de caractères sont arrivés avec la version 12?

```
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
```

554 nouveaux, cette version 12 est très modérée. Quels sont ces nouveaux caractères?

```

ucd=> SELECT To_U(codepoint) AS Codepoint, name FROM Characters WHERE version='12.0';
codepoint | name
-----+-----
...
U+1FA70   | BALLET SHOES
...
U+2E4F    | CORNISH VERSE DIVIDER
...
U+A7C3    | LATIN SMALL LETTER ANGLICANA W
...
U+10FE0   | ELYMAIC LETTER ALEPH
U+10FE1   | ELYMAIC LETTER BETH
U+10FE2   | ELYMAIC LETTER GIMEL
...
U+1E100   | NYIAKENG PUACHUE HMONG LETTER MA
U+1E101   | NYIAKENG PUACHUE HMONG LETTER TSA
U+1E102   | NYIAKENG PUACHUE HMONG LETTER NTA
...
U+1F6D5   | HINDU TEMPLE
U+1F6FA   | AUTO RICKSHAW
...
U+1F97B   | SARI
...
U+1F9A7   | ORANGUTAN
...
U+1F9BB   | EAR WITH HEARING AID
U+1F9BC   | MOTORIZED WHEELCHAIR
U+1F9BD   | MANUAL WHEELCHAIR
...
U+1FA30   | WHITE CHESS KNIGHT ROTATED TWO HUNDRED TWENTY-FIVE DEGREES
U+1FA31   | BLACK CHESS KNIGHT ROTATED TWO HUNDRED TWENTY-FIVE DEGREES

```

Parmi les emojis de cette version, beaucoup concernent l'Inde, comme le sari ou le rickshaw. Beaucoup de caractères liés au handicap ont été créés, ainsi, mais c'est plus anecdotique, que beaucoup de caractères pour le jeu d'échecs. On trouve aussi des caractères étonnants comme U+2E4F, qui ne sert apparemment qu'à la poésie cornouaillaise. Et il y a bien sûr de nouvelles écritures comme l'élymaïque ou le nyiakeng puachue hmong. Même l'alphabet latin voit arriver de nouveaux caractères comme le U+A7C3.

Au fait, l'unique caractère de la version 12.1, c'était quoi ?

```

ucd=> SELECT To_U(codepoint) AS Codepoint, name FROM Characters WHERE version='12.1';
codepoint | name
-----+-----
U+32FF    | SQUARE ERA NAME REIWA

```

Une version d'Unicode uniquement pour introduire <http://blog.unicode.org/2019/05/unicode-12-1-en.html> un caractère japonais permettant de noter l'ère Reiwa... (Merci à John Shaft pour avoir repéré celui-là.)

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
```

Si vous avez les bonnes polices de caractères, voici les caractères pris en exemple plus haut : [Caractère Unicode non montré¹]<<https://r12a.github.io/uniview/?char=1FA70>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=2E4F>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=A7C3>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10FE0>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10FE1>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=10FE2>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1E100>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1E101>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1E102>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F6D5>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F6FA>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F97B>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9A7>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9BB>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9BC>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1F9BD>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1FA30>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=1FA31>>, [Caractère Unicode non montré]<<https://r12a.github.io/uniview/?char=32FF>> ... (Si vous n'avez pas les bonnes polices, chaque lettre est un lien vers Uniview <<https://r12a.github.io/uniview/>>.)

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