

# Version 10 d'Unicode

Stéphane Bortzmeyer  
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 26 septembre 2017

<https://www.bortzmeyer.org/unicode-10-0.html>

---

Le 20 juin (oui, je suis en retard), la nouvelle version d'Unicode <<http://blog.unicode.org/2017/06/announcing-unicode-standard-version-100.html>> est sortie, la 10.0. Une description officielle des principaux changements est disponible <<http://www.unicode.org/versions/Unicode10.0.0/>> 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
-----
136755
```

Combien caractères sont arrivés avec la version 10?

```
ucd=> SELECT version,count(version) FROM Characters GROUP BY version ORDER BY version::float;
...
8.0      | 7716
9.0      | 7500
10.0     | 8518
```

8 518 nouveaux. Lesquels?

```

ucd=> SELECT To_U(codepoint) AS Codepoint, name FROM Characters WHERE version='10.0';
codepoint | name
-----+-----
...
U+20BF    | BITCOIN SIGN
...
U+11A00   | ZANABAZAR SQUARE LETTER A
...
U+11D01   | MASARAM GONDI LETTER AA
...
U+1F6F8   | FLYING SAUCER
...
U+1F96A   | SANDWICH

```

Le premier cité va évidemment faire plaisir aux possesseurs de bitcoins. Mais on trouve aussi quatre écritures entièrement nouvelles comme le zanabazar ou le masaram gondi cités en exemple plus haut. Et il y a bien sûr l'habituel lot d'emojis pour faire rire les réseaux sociaux. Tiens, d'ailleurs, combien de caractères 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
-----
6978

```

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
-----+-----
Currency_Symbol | 54
Other_Symbol | 5855
Modifier_Symbol | 121
Math_Symbol | 948
(4 rows)

```

Si vous avez les bonnes polices de caractères, voici les caractères pris en exemple plus haut : [Caractère Unicode non montré<sup>1</sup>] <<https://r12a.github.io/uniview/?char=20BF>>, [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=11A00>>, [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=11D01>>, [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F6F8>>, [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F96A>>... (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  $\LaTeX$