

Version 17 d'Unicode

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 9 septembre 2025

<https://www.bortzmeyer.org/unicode-17-0.html>

Le 9 septembre 2025 est sortie la version 17 d'Unicode <<https://blog.unicode.org/2025/09/unicode-170-release-announcement.html>>. Une description officielle des principaux changements est disponible <<https://www.unicode.org/versions/Unicode17.0.0/>> 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
-----
159866
```

Combien de caractères sont arrivés avec la version 17?

```
ucd=> SELECT version,count(version) FROM Characters GROUP BY version ORDER BY version::float;
...
14.0 | 838
15.0 | 4489
15.1 | 627
16.0 | 5185
17.0 | 4803
```

4 803 nouveaux caractères, c'est pas mal, mais moins que la précédente version. Quels sont ces nouveaux caractères?

```

ucd=> SELECT To_U(codepoint) AS Code_point, name FROM Characters WHERE version='17.0' ORDER BY Codepoint;
code_point | name
-----+-----
U+FDCC     | ARABIC LIGATURE SALLALLAHU ALAYHI WA-ALAA AALIHEE WA-SALLAM
...
U+10940    | SIDETIC LETTER N01
U+10941    | SIDETIC LETTER N02
U+10942    | SIDETIC LETTER N03
...
U+11DB0    | TOLONG SIKI LETTER I
U+11DB1    | TOLONG SIKI LETTER E
U+11DB2    | TOLONG SIKI LETTER U
U+11DB3    | TOLONG SIKI LETTER O
...
U+16EA0    | BERIA ERFE CAPITAL LETTER ARKAB
U+16EA1    | BERIA ERFE CAPITAL LETTER BASIGNA
U+16EA2    | BERIA ERFE CAPITAL LETTER DARBAI
...
U+1CEC8    | EUNOMIA
U+1CEC9    | PSYCHE
U+1CECA    | THETIS
U+1CECB    | MELPOMENE
...
U+1F6D8    | LANDSLIDE
...
U+1FAC8    | HAIRY CREATURE
...

```

Cette version amène en effet quatre nouvelles écritures. Si l'écriture sidétique n'est connue que par une douzaine d'inscriptions, le zaghawa (Beria Erfe), lui, est toujours en usage. On voit aussi arriver le Tolong Siki et le Lai Tay (Tai Yo). Vous noterez aussi les symboles astronomiques pour les plus gros astéroïdes, comme Victoria.

Il y a d'autres nouveautés que l'addition de caractères. Par exemple les propriétés pour la coupure de mots de la cédille, malgré l'ancienneté de ce caractère, ont été changées pour se conformer aux règles du saanich. Ou bien une nouvelle classe pour les tirets a été créée, "*Unambiguous Hyphen*", pour des caractères comme le [Caractère Unicode non montré ¹] (regardez bien, c'est le tiret Unicode, U+2010, pas le tiret ASCII, U+002D, qui peut être un tiret ou bien le signe moins).

En plus futile, il y a l'habituelle arrivée <<https://blog.unicode.org/2025/07/say-hello-to-new-emoj.html>> d'emojis comme le trombone ou le coffre au trésor. Par contre, le trognon de pomme a finalement été retiré et n'apparaîtra pas avant Unicode 18.

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 troisième lettre de l'alphabet sidétique [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=10942>>, le caractère Zaghawa Kobo [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=16EA9>>, l'astéroïde Psyche [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1C1C9>>, le yéti [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1FAC8>> et le glissement de terrain [Caractère Unicode non montré] <<https://r12a.github.io/uniview/?char=1F6D8>>.

1. Car trop difficile à faire afficher par \LaTeX