

RFC 1345 : Character Mnemonics and Character Sets

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 4 octobre 2007

Date de publication du RFC : Juin 1992

<https://www.bortzmeyer.org/1345.html>

Un ancien RFC, qui représentait une des premières incursions de l'IETF dans le monde complexe des jeux de caractères.

Ce RFC tentait d'énumérer tous les jeux de caractères en usage à l'époque (la bibliographie cite le manuel de VMS, pour le jeu de caractères de ce système!). Les préoccupations d'internationalisation sont anciennes à l'IETF et ce RFC reflète une époque très reculée, où la version 1.0 d'Unicode venait tout juste d'être publié.

Ce RFC est donc essentiellement constitué d'un format pour décrire un jeu de caractères et d'une longue liste de jeux de caractères utilisés à l'époque, en utilisant les noms d'une version préliminaire d'ISO 10646. C'était donc un concurrent direct de la norme Unicode. Cette démarche n'a pas eu un succès fracassant et MIME, par exemple, se contentera d'attribuer un identificateur unique <<https://www.iana.org/assignments/character-sets>> à chaque jeu, sans essayer de le décrire complètement (mais en citant notre RFC). Comparons par exemple l'entrée de notre RFC pour US-ASCII :

```
&charset ANSI_X3.4-1968
&rem source: ECMA registry
&alias iso-ir-6
&alias ANSI_X3.4-1986
&alias ISO_646.irv:1991
&g0esc x2842 &g1esc x2942 &g2esc x2a42 &g3esc x2b42
&alias ASCII
&alias ISO646-US
&alias US-ASCII
&alias us
&alias IBM367
&alias cp367
&code 0
NU SH SX EX ET EQ AK BL BS HT LF VT FF CR SO SI
DL D1 D2 D3 D4 NK SY EB CN EM SB EC FS GS RS US
SP ! " Nb DO % & ' ( ) * + , - . / 0 1 2 3 4 5 6 7 8 9 : ; < = > ?
At A B C D E F G H I J K L M N O P Q R S T U V W X Y Z <( // )> '> _
'! a b c d e f g h i j k l m n o p q r s t u v w x y z (! !! !) '? DT
```

avec celle d'aujourd'hui, tirée du registre IANA, géré selon le RFC 2978¹ :

```
Name: ANSI_X3.4-1968 [RFC1345,KXS2]
MIBenum: 3
Source: ECMA registry
Alias: iso-ir-6
Alias: ANSI_X3.4-1986
Alias: ISO_646.irv:1991
Alias: ASCII
Alias: ISO646-US
Alias: US-ASCII (preferred MIME name)
Alias: us
Alias: IBM367
Alias: cp367
Alias: csASCII
```

On notera que, pour faire rentrer les jeux de caractères dans un RFC écrit en ASCII, ce RFC développe un codage se voulant « intuitif » (la section 2.3 dit *"The two characters are chosen so the graphical appearance in the reference set resembles as much as possible (within the possibilities available) the graphical appearance of the character."*). C'est ainsi que le é, le « e accent aigü » s'écrit e' et le è, l'« e accent grave » e!. Voici la seconde moitié de la table ISO 8859-1, ainsi exprimée :

```
A! A' A> A? A: AA AE C, E! E' E> E: I! I' I> I:
D- N? O! O' O> O? O: *X O/ U! U' U> U: Y' TH ss
a! a' a> a? a: aa ae c, e! e' e> e: i! i' i> i:
d- n? o! o' o> o? o: -: o/ u! u' u> u: y' th y:
```

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc2978.txt>