

What is a character?

UTF-8, Unicode, and the Xchar wordset

M. Anton Ertl, TU Wien

Concepts

Concept	UTF-8	Forth
code unit	1 byte	char
code point	1-4 bytes	xchar or string
glyph/character	≥ 1 code point	string

Use Strings!

Usage inside Gforth (2021)

xchar		string		xchar ext		string	
3	x-size			2	+x/string	71	/string
1	xc!+	43	move	0	-trailing-garbage		
3	xc!+?	43	move	1	ekey>xchar		
5	xc,	9	mem,	6	x-width		
5	xc-size	114	nip	3	xc-width	6	x-width
5	xc@+	43	move	6	xchar-		
4	xchar+			0	xhold	4	holds
1	xemit	151	type	1	x\string-	71	/string
2	xkey						
1	xkey?						