

CS 208

F, 26 Sep 2025

char — 1 byte

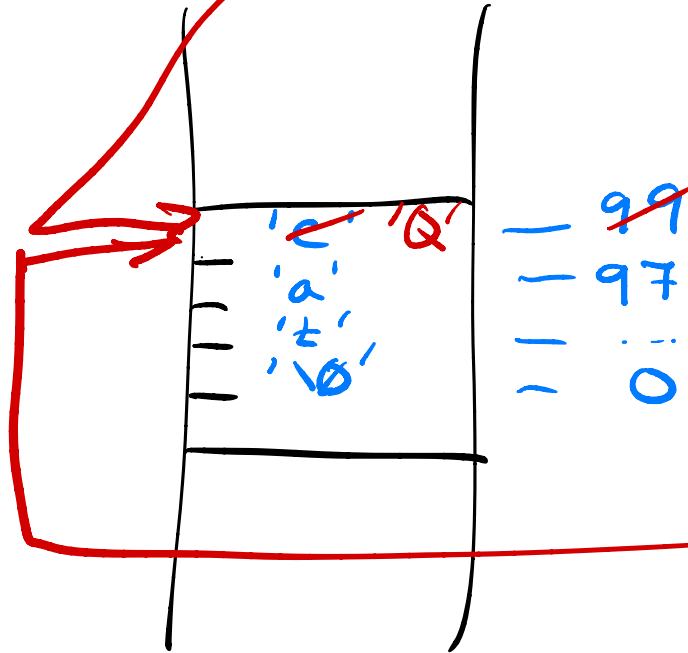
char * — the address
of a byte

OR (more commonly)

the address of an array
of chars/bytes

main

char a[5] = "cat";
to_upper(a);



to_upper(char *s)
s[0] = 'Q';

" — "

r/

null-terminated
string

' Z '

single char

int x = -17; 0xffffffffef



"little-endian"

Intel
AMD

Apple Silicon
(M1, M2, M3, M4)

big-endian

Motorola
(early Macs)

configurable

RISC

Gulliver's Travels, J. Swift

17(20)?
"little & big endian"

~1980 "On Holy Wars & a
Plea for Peace" ↗
Intel & Motorola ignored him

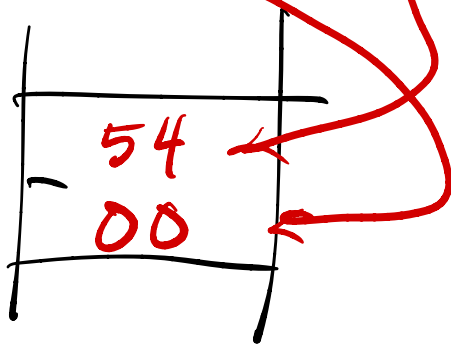
Character encodings

character $\propto$ $u+0251$

store in memory

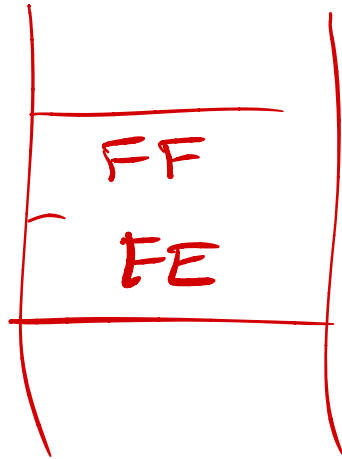
$T \sim u + 0054$

UTF16
LE

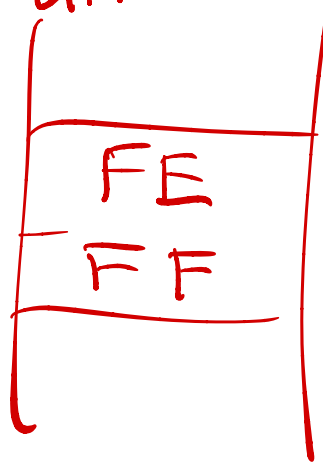


U+FEFF - "byte order mark"
BOM

UTF-16 LE



UTF-16 BE



Ken Thompson

U+0251

α

row 2 of
the Wikipedia
UTF-8
chart

0000|0010|0101|0001
xxx yy yy zzzz

Build the
UTF8
2-byte (2nd row of table)
encoding

C9
91

110xxx yy 10yy zzzz

11001001 10010001

C9

91

row 2

$u+0080 - u+07FF$

1000 0000 0111 1111 1111
8 bits $\rightarrow$ 11 bits

C9 1100

1100
 $\uparrow$ row 2

$$a = 97 = \underline{01} \textcircled{1} 0 \underline{0001} = 0x61$$

$$A = 65 = \underline{01} \textcircled{0} 0 \underline{0001} = 0x41$$