


CS 208

F, 16 Jan 2026

Recap

characters : A

codepoints : U + 0041

encodings: save codepoints
as actual sequences of bytes

UTF-16LE encoding

U+0041 $\rightarrow$ 41 00
byte byte

UTF-16BE

U+0041 $\rightarrow$ 00 41

$$u + 03BI \sim \alpha$$

— Look at chart, decide which row applies

$u + 0080 - u + 07FF \leftarrow 03BI$ is here.

$110xxx yy \quad 10yy zzzz$

— $03BI \rightarrow 0000 \underbrace{0011}_{x's} \underbrace{1011}_{y's} \underbrace{0001}_{z's}$

$110xxx yy$
 $= 110 \underbrace{011}_{x} \underbrace{10}_{y}$

$10yy zzzz$
 $10 \underbrace{11}_{y} \underbrace{0001}_{z}$

11001110

CE

10110001

BI

U+266D

- pick chart row 3rd (0800 - FFFF)
- expand code point in binary
- move bits based on chart

0010 0110 0110 1101
wwwv xxxx yyzz zzzz

1110 0010 → E2

1001 1001 99

1010 1101 AD

B

0 1 0 0 0 0 1 0

b

0 1 0 0 0 1 0