

{ASCII}

8 bits
(1 byte)

8 bits
(1 byte)

16 bits
(2 bytes)

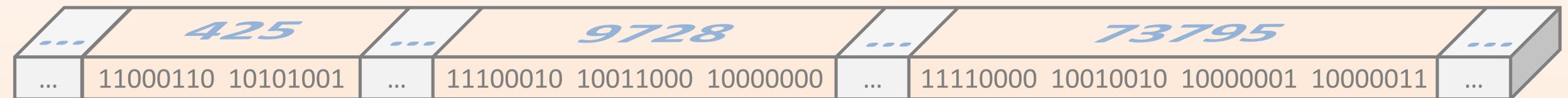
16 bits
(2 bytes)



16 bits
(2 bytes)

24 bits
(3 bytes)

32 bits
(4 bytes)



32 bits
(4 bytes)



{UTF-8}

ASCII and UTF-8. It shows the back compatibility of UTF-8. On the vertical axis, the first half of the figure shows the structure of ASCII, which encodes for symbols using 8-bit sequences (1 byte). A schematic of UTF-8 is unrivaled in the second half of the figure. The UTF-8 relationship with ASCII is preserved for encoding positions starting from 0 to 127. However, starting from position 128 up to 255, ASCII and UTF-8 use different encodings. Namely, ASCII uses 1 byte for this range, whereas UTF-8 uses 2 bytes. Outside the ASCII range, UTF-8 uses 2 bytes up to 4 bytes to encode new arrivals in the symbol set. UTF-8 may stop at 32 bit (4 bytes) representations, as all symbols with meaning in all human history, does not exceed 4.3 billion, as 4 bytes can encode.