Unicode …

UTF-8

1-byte: (6 effective bits) 10xxxxxx

2-byte: (11 effective bits) 110xxxxx 10xxxxxx

3-byte: (16 effective bits) 1110xxxx 10xxxxxx 10xxxxxx

4-byte: (21 effective bits) 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

5-byte: (26 effective bits) 111110xx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx

6-byte: (31 effective bits) 1111110x 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx