

# UTF-8

## Umrechnung Unicode UTF-8

Unicode Codepoint	Gesetzte Unicode Bits	UTF-8 Bytes	UTF-8 Byte 1	UTF-8 Byte 2	UTF-8 Byte 3	UTF-8 Byte 4
U+0000 bis U+007F	7	1	0xxx xxxx	/	/	/
U+0080 bis U+07FF	11	2	110x xxxx	10xx xxxx	/	/
U+0800 bis U+FFFF	16	3	1110 xxxx	10xx xxxx	10xx xxxx	/
U+10000 bis U+10FFFF	21	4	1111 0xxx	10xx xxxx	10xx xxxx	10xx xxxx

Der erste Bereich ist gleich mit ASCII-Code

## Beispiel

### Beispiel für Codierung mit UTF-8

"€" - Codepoint U+20AC

1) Entsprechend der Tabelle werden 3 Bytes zur Codierung benötigt, da der Codepoint zwischen U+0800 und U+FFFF liegt.

2) 20AC => 0010 0000 1010 1100

3) 1110 0010 1000 0010 1010 1100

E 2 8 2 A C - UTF-8 Hex  
Aus Tabelle