

Unicode

Codierung	Beschreibung
UTF-8	Die am häufigsten verwendete Codierung. (HTML, Mail) Die ersten 128 Zeichen entsprechen ASCII Variable Länge 1-4 Byte Platzsparend beim lateinischen Alphabet
UTF-16	Die älteste Abbildungsvariante von Unicode Es werden 2 Byte (16 Bit) zur Codierung verwendet (teilweise auch 4 Byte) Teilweise kürzer als UTF-8 (bei nicht lateinischen Alphabeten) Es gibt 2 Varianten (Little und Big Endian)
UTF-32	Fixe Länge von 32 Bit (4 Byte) Braucht immer am meisten Speicherplatz Einfach zu gebrauchen, da fixe Länge Es gibt 2 Varianten (Little und Big Endian)

	Markierung	Ebene	Zeichen	Unicode	In Windows (Dezimal)
	U+	00 - 10	0000 - FFFF		Alt+
Bsp: A	U+	00	0041	U+0041	65
Bsp: ∞	U+	00	221E	U+221E	8734
Bsp: ☿	U+	1	F640	U+1F640	128576

Der gesamte vom Unicode-Standard beschriebene Bereich umfasst 1.114.112

UTF-8

Unicode Zeichen	UTF-8 Codierung	Verwendungszweck
0000 0000 - 0000 007F ₁₆ (7F ₁₆ =127 ₁₀)	0xxx xxxx ₂	UTF-8 gleich wie ASCII Zeichensatz 0-127
0000 0041 ₁₆ 100 0001 ₂	0100 0001 ₂	Bsp: A (7 Bit)
0000 0080 - 0000 07FF ₁₆ (80 ₁₆ =128 ₁₀ ; 7FF ₁₆ =2047 ₁₀ =111 1111 1111 ₂)	110x xxxx 10xx xxxx ₂	Unicode Zeichen wird von rechts her aufgefüllt.
0000 03A9 ₁₆ 1110101001 ₂	1100 1110 1010 1001 ₂ CEA9 ₁₆	Bsp: Ω (Alt+937)
0000 0800 - 0000 FFFF ₁₆ (800 ₁₆ =2048 ₁₀ ; 7FF ₁₆ =65535 ₁₀ =1111 1111 1111 1111 ₂)	1110 xxxx 10xx xxxx 10xx xxxx ₂	Unicode Zeichen wird von rechts her aufgefüllt
0000 221E ₁₆ 10 0010 0001 1110 ₂	1110 0010 1000 1000 1001 1110 ₂ E2889E ₁₆	Bsp: ∞(Alt+937)
0001 0000 - 0010 FFFF ₁₆ (10000 ₁₆ =65535 ₁₀ ; 10FFFF ₁₆ = 1'114'111 ₁₀)	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx ₂	Unicode Zeichen wird von rechts her aufgefüllt
0001 F640 ₁₆ 1111101100100000 ₂	1111 0000 1001 1111 1001 1001 1000 0000 ₂ F09F9980 ₁₆	Bsp: ☿ (Alt+128576)

UTF-16

Unicode Zeichen	UTF-16 Codierung (Bits)	
0000 0000 - 0000 FFFF ₁₆ (7FF ₁₆ =65535 ₁₀ =1111 1111 1111 1111 ₂)	xxxx xxxx xxxxx xxxx	Unicode Ebene 0 -> 2 Byte
0000 0041 ₁₆ 0100 0001 ₂	0041 ₁₆ 0000 0000 0100 0001 ₂	Bsp: A (Ebene 0)
0000 20AC ₁₆ 00100000 10101100 ₂	20AC ₁₆ 00100000 10101100 ₂	Bsp: € (Ebene 0)
0001 0000 - 0010 FFFF ₁₆ (10000 ₁₆ =65535 ₁₀ ; 10FFFF ₁₆ = 1'114'111 ₁₀)	1101 10xx xxxx xxxx 1101 11xx xxxx xxxx	Unicode Ebene 1 Berechnung: Unicode - 10000 ₁₆ Aufteilung auf 2 Byte
Unicode: 0001 F36B ₁₆ - 0001 0000 ₁₆ F36B ₁₆ 0000 1111 00 11 0110 1011 ₂	1101 1000 0011 1100 ₂ 1101 1111 0110 1011 ₂ entspricht: D83C DF6B _{16BE} entspricht: 3CD8 6BDF _{16LE}	Bsp: ☿ Schokolade (Alt+127851)
Unicode: 0002 4B62 ₁₆ - 0001 0000 ₁₆ 1 4B62 ₁₆ 0001 0100 10 11 0110 0010 ₂	1101 1000 0101 0010 ₂ 1101 1111 0110 0010 ₂ entspricht:D852DF62 ₁₆ entspricht: 52D8 62DF ₁₆	Bsp: 鞆 (Alt+150370)

UTF-32

Unicode Zeichen	UTF-32 Codierung (Bits)	
0000 0041 ₁₆ 0100 0001 ₂	0000 0041 ₁₆ 0000 0000 0000 0000 0000 0000 0100 0001 ₂	Bsp: A (Ebene 0)
Unicode: 0001 F36B ₁₆ 1 1111 0011 0110 1011 ₂	0000 0000 0000 0001 1111 0011 0110 1011 ₂	Bsp: ☿ Schokolade