

UTF-16

Endian

Zur Erkennung wird die Byte Order Magnitude (BOM) verwendet.

Little Endian BOM: FFFE

Big Endian BOM: FFFF

BMP

Der Codepoint ist zwischen 0000_{16} und $FFFF_{16}$

Innerhalb des BMPs

wird der Codepoint 1:1 übersetzt

z.B.

U+0062 (b) = UTF16BE 0062

Außerhalb der BMPs

1. Codepoint in Binär umrechnen
2. 65536_{10} bzw. $0001\ 0000\ 0000\ 0000\ 0000_2$ abziehen
3. in 2 Zehn Bit Zahlen aufteilen
4. erste Hälfte $D800_{16}$ bzw. $1101\ 1000\ 00_2$ addieren
5. zweite Hälfte $DC00_{16}$ bzw. $11\ 0111\ 0000_2$ addieren
6. Die Hälften wieder zusammenfügen und in Hex umwandeln