

Leerzeichen zwischen den hexadezimalen oder den binären Zeichenfolgen müssen nicht beachtet werden. Sie dienen der besseren Lesbarkeit.

Wenn nicht anderes steht, kann davon ausgegangen werden, dass bei Multibyte-Codierungen, die Bytes in der Big Endian-Reihenfolge angeordnet sind.

### Aufgabe 1

Stelle den Text „Wow!“ hexadezimal in den folgenden Codierungen dar:

- (a) UTF-8
- (b) UTF-16
- (c) UTF-32

### Aufgabe 2

Handelt es sich bei der hexadezimalen Folge

ef b7 b2

grundsätzlich um ein korrekte UTF-8-Codierung?

### Aufgabe 3

Wie interpretiert ein Programm, das UTF-16-Daten erwartet, den ASCII-Text

Forget it!

### Aufgabe 4

Gib die Binärcodierung des Unicode-Zeichens U+2701 in den folgenden Formaten an:

- (a) UTF-16
- (b) UTF-32
- (c) UTF-8

### Aufgabe 5

Sind die Zeichen, deren Binärdarstellung gegeben ist, im rechts angegebenen Format?

- (a) 11010101 10101111 ASCII
- (b) 01010101 10101111 UTF-16
- (c) 01010101 01110101 UTF-8
- (d) 11010101 01111001 UTF-8

## Aufgabe 6

Die folgenden Hex-Ziffern codieren *ein* Unicode-Symbol in UTF-8:

F0 93 81 A8

- (a) Welche hexadezimale Nummer hat das Symbol?
- (b) Warum reichen 3 Bytes für die UTF-8-Codierung nicht aus?

*Nebenbei:* Es handelt sich bei dem Symbol um die ägyptische Hieroglyphe



zur Darstellung der Zahl 1 000 000.