Lektion 06

Kodierung von Buchstaben

ASCII und UTF-8 · Skript, Kapitel 2

Wie kodiert man Buchstaben?

Computer speichern nur Nullen und Einsen. Wie könnte man damit **Buchstaben** repräsentieren?

ASCII

American Standard Code for Information Interchange (ASCII) – eine der ersten Text-Kodierungen

- 1 Byte pro Zeichen – z. B. `0100 0001` $\rightarrow$ A

ASCII

ASCII – eine der ersten Text-Kodierungen

- 1 Byte pro Zeichen – z. B. `0100 0001` $\rightarrow$ A
- 1 Bit war für Fehlerkontrolle reserviert (Paritätsbit)

Frage: Wie viele Zeichen kann man mit einem Byte abbilden? Und mit nur 7 nutzbaren Bits (ASCII)?

ASCII

ASCII – eine der ersten Text-Kodierungen

- 1 Byte pro Zeichen – z. B. `0100 0001` $\rightarrow$ A
- 1 Bit war für Fehlerkontrolle reserviert (Paritätsbit)

Frage: Wie viele Zeichen kann man mit einem Byte abbilden? Und mit nur 7 nutzbaren Bits (ASCII)?

$$2^8 = 256, \quad 2^7 = 128$$

ASCII-Tabelle (Auszug)

Dezimal	Hexadezimal	Binär	Zeichen
32	20	00100000	Space
48	30	00110000	0
49	31	00110001	1
...	...	...	...
65	41	01000001	A
66	42	01000010	B
67	43	01000011	C
...	...	...	...
97	61	01100001	a
98	62	01100010	b

Vollständige Tabelle: Skript, Kapitel 3.

Entschlüsseln Sie!

01101000 01100101 01101100 01101100 01101111

Nutzen Sie die ASCII-Tabelle im Skript: welches Wort ist das?

Entschlüsseln Sie!

`01101000 01100101 01101100 01101100 01101111`

Nutzen Sie die ASCII-Tabelle im Skript: welches Wort ist das?

`hello`

Die Grenzen von ASCII

128 Zeichen reichen bei weitem nicht für Umlaute, Akzente, Emojis, Sprachen wie Chinesisch, Arabisch, ...

→ Unicode Transformation Format — 8-bit (UTF-8): bis zu **4 Bytes** pro Zeichen

Die Grenzen von ASCII

128 Zeichen reichen bei weitem nicht für Umlaute, Akzente, Emojis, Sprachen wie Chinesisch, Arabisch, ...

$\rightarrow$ UTF-8: bis zu **4 Bytes** pro Zeichen

Frage: Wie viele Zeichen kann man mit 4 Bytes maximal abbilden?

Die Grenzen von ASCII

128 Zeichen reichen bei weitem nicht für Umlaute, Akzente, Emojis, Sprachen wie Chinesisch, Arabisch, ...

$\rightarrow$ UTF-8: bis zu **4 Bytes** pro Zeichen

Frage: Wie viele Zeichen kann man mit 4 Bytes maximal abbilden?

$2^{32} = 4\,294\,967\,296$ (etwas weniger nutzbar, da Bits zur Längenangabe gebraucht werden)

Auftrag

Kapitel 3: ASCII & UTF-8

Aufgabe 3.3

Aufgabe 3.4

Aufgabe 3.5