

2.3.A Codierung/UTF-8, UTF-16 und UTF-32 – Arbeitsblatt

Aufgabe 1: Umwandlung Codennummer nach UTF-8

In Code-Tabellen sind häufig die Codenummern als Hexadezimalzahlen zu den gesuchten Zeichen angegeben. Sie sollen nun diese Zeichen in UTF-8-Darstellung bringen; diese UTF-8-Darstellung wiederum soll für die Benutzung in C-Programmen in Oktalzahlen umgewandelt werden. Hier ein Beispiel für den Buchstaben „ß“:

- Für den Buchstaben ß wird die Codennummer $DF_{(16)}$ genannt.

- Umwandeln in eine Dualzahl:

$$DF_{(16)} = 11011111_{(2)}$$

- Umwandeln nach UTF-8:

$$11000011, 10011111_{(2)}$$

- In einem C-Programm schreibt man es entweder oktal:

$$303_{(8)}, 237_{(8)}$$

- Oder man schreibt es hexadezimal:

$$C3_{(16)}, 9F_{(16)}$$

- In einem Programmtext sieht das dann so aus:

1	<code>printf("Ohne_Flei\303\237_kein_Preis.\n"); // oktal</code>
2	<code>printf("Ohne_Flei\xC3\x9F_kein_Preis.\n"); // hexadezimal</code>

- a) Das Paragraph-Zeichen „§“ hat in Unicode die Codennummer $A7_{(16)}$.

1) Ermitteln Sie die Binärdarstellung der Codennummer!

2) Wandeln Sie den Code nach UTF-8!

3) Stellen Sie die beiden UTF-8-Bytes oktal dar!

- b) Der griechische Buchstabe „λ“ (in der Elektrotechnik das Symbol für die Wellenlänge) hat in Unicode die Codennummer $3BB_{(16)}$.

1) Ermitteln Sie die Binärdarstellung der Codennummer!

2) Wandeln Sie den Code nach UTF-8!

3) Stellen Sie die beiden UTF-8-Bytes oktal dar!

c) In manchen englischsprachigen Ländern wird als Einheit für den elektrischen Leitwert nicht das Siemens (abgekürzt S) verwendet, sondern das Mho. Es ist der Kehrwert von Ohm und wird abgekürzt mit dem Zeichen „ Ω “. Die zugehörige Unicode-Codenummer ist $2127_{(16)}$.

- 1) Ermitteln Sie die Binärdarstellung der Codenummer!
- 2) Wandeln Sie den Code nach UTF-8!
- 3) Stellen Sie die drei UTF-8-Bytes oktal dar!