

Aufgabe 1:

Der Satz “Schöne Weihnachtszeit” ergibt in verschiedenen Encodings die folgenden Octaldumps:

1.)

```
0000000 177 377 000 123 000 143 000 150 000 366 000 156 000 145 000 040
0000020 000 127 000 145 000 151 000 150 000 156 000 141 000 143 000 150
0000040 000 164 000 163 000 172 000 145 000 151 000 164 000 041 000 012
0000060
```

UTF-16, BOM: 177377

2.)

```
0000000 061 523 141 550 067 266 020 145 062 527 064 151 060 556 064 143
0000020 071 564 062 572 072 151 005 041
0000030
```

UTF-8

3.)

```
0000000 000 000 000 123 000 000 000 143 000 000 000 150 000 000 000 366
0000020 000 000 000 156 000 000 000 145 000 000 000 040 000 000 000 127
0000040 000 000 000 145 000 000 000 151 000 000 000 150 000 000 000 156
0000060 000 000 000 141 000 000 000 143 000 000 000 150 000 000 000 164
0000100 000 000 000 163 000 000 000 172 000 000 000 145 000 000 000 151
0000120 000 000 000 164 000 000 000 041 000 000 000 012
0000140
```

UCS-4

4.)

```
0000000 061 523 173 150 062 556 053 440 064 545 067 150 061 541 072 150
0000020 075 163 064 545 020 564 005 012
0000030
```

ISO-Latin

Aufgabe 2:

Das Wort “Computerlinguistik” ergibt folgenden Hexdump:

```
00000000 ff fe 43 00 6f 00 6d 00 70 00 75 00 74 00 65 00    |..C.o.m.p.u.t.e.|
00000010 72 00 6c 00 69 00 6e 00 67 00 75 00 69 00 73 00    |r.l.i.n.g.u.i.s.|
00000020 74 00 69 00 6b 00 0a 00 0a 00                      |t.i.k.....|
0000002a
```

a) Lässt sich eine Aussage über das Encoding treffen? Wenn ja, um welches Encoding handelt es sich?

Vor jedem ‘‘sinntragenden’’ Symbol befindet sich genau eine Nullstelle (auch in der ASCII-Darstellung zu sehen), darum muss es sich um ein 2 Byte Encoding handeln (UTF-16).

b) Gibt es Byteordermarks (BOM)? Wenn ja, wo befinden sie sich und was sagen sie aus?

Ja, und zwar ‘‘ff fe’’ in der ersten Zeile.

Aufgabe 3:

Wie viele Byte pro Zeichen hat:

- a) UCS-4? 4
- b) UTF-8? 1-6
- c) UTF-16? 2
- d) ISO 8859-15? 1

Aufgabe 4:

Der Satz “Schöne Weihnachtszeit” ergibt hier folgenden Hexdump:

```
00000000 53 63 68 c3 b6 6e 65 20 57 65 69 68 6e 61 63 68 |Sch..ne Weihnach|
00000010 74 73 7a 65 69 74 21 0d 0a                      |tszeit!..|
00000019
```

a) Um welches Encoding handelt es sich?

Der Umlaut ‘‘ö’’ ist mit zwei Byte codiert, während die restlichen ASCII-konformen Buchstaben nur ein Byte groß sind. Also muss es sich um UTF-8 handeln.

b) Welche Zeilenendcodierung wurde verwendet (NL, LF/CR, CR)?

Man sieht am Ende der Zeile 0d 0a, also muss es sich um einen Zeilenumbruch handeln, der aus zwei Zeichen besteht. Dementsprechend CR LF.