

Aufgabe 1:

Das Wort Hochschule ergibt folgenden Octaldump:

```
0000000 067 510 064 143 061 563 072 550 062 554 000 012
0000013
```

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

Da das Wort keine Umlaute enthält, ist es ASCII konform. Das heißt, dass man aufgrund des Dumps nicht unterscheiden kann, ob es sich um ASCII, UTF-8 oder eine ISO Norm handelt. Da jedes Symbol aber nur ein Byte beansprucht, können UTF-16 und UCS-4 ausgeschlossen werden.

Aufgabe 2:

Das Wort Universität ergibt folgenden Octaldump:

```
0000000 067 125 073 151 071 145 064 563 141 564 072 244 000 012
0000015
```

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

Das Wort Universität enthält einen Umlaut und damit ein nicht ASCII konformes Zeichen. Da der Dump ein Byte mehr als das Wort Buchstaben enthält, kann man davon ausgehen, dass es sich um eine Multibytecodierung handelt. Da die anderen Symbole mit jeweils einem Byte codiert sind, kann man UTF-16 ausschließen und es muss sich um UTF-8 handeln.

Aufgabe 3:

Das Wort Informatik ergibt folgenden Hexdump in verschiedenen Encodings:

```
00000000 ff fe 49 00 6e 00 66 00 6f 00 72 00 6d 00 61 00 |..I.n.f.o.r.m.a.|
00000010 74 00 69 00 6b 00 0a 00 |t.i.k...|
00000018
```

```
00000000 49 6e 66 6f 72 6d 61 74 69 6b 0a |Informatik.|
0000000b
0000013
```

a) Um welches Encoding kann es sich jeweils handeln?

Man sieht, dass die ‘‘sinntragenden’’ Bytes gleich sind und der Rest mit ‘‘leeren’’ Bytes aufgefüllt wurde. Darum handelt es sich bei dem oberen Encoding vermutlich um UTF-16, bei dem unteren um ASCII/ISO oder UTF-8.

b) Gibt es Byteordermarks (BOM) und wenn ja, wo?

Im oberen Dump, also in UTF-16 gibt es ein ff fe (Little Endian).

c) Welche Codierung des Newline Characters wurde verwendet?

Man findet ein 0a in beiden Dumps, also handelt es sich um \n

Aufgabe 4:

Eins der zuvor bereits benutzten Wörter ergibt folgenden Octaldump:

```
0000000 000 000 000 111 000 000 000 156 000 000 000 146 000 000 000 157
0000020 000 000 000 162 000 000 000 155 000 000 000 141 000 000 000 164
0000030 000 000 000 151 000 000 000 153 000 000 000 012
0000040
```

a) Um welches Encoding handelt es sich?

Im Gegensatz zu den bisherigen Dumps existieren hier vor jedem ‘‘sinntragenden’’ Byte jeweils drei ‘‘leere’’ Bytes. Daher muss es sich um UCS-4 handeln.

b) Um welches Wort handelt es sich?

Die ‘‘sinntragenden’’ Bytes decken sich mit keinem des bisherigen Octaldumps, daher muss es sich um das Wort handeln, von dem bisher nur der Hexdump sichtbar war, also ‘‘Informatik’’. Zählt man nach, deckt sich auch die Anzahl der Buchstaben.