

Übungsblatt 4

1. Es sei das folgende Wort w gegeben:

```
1070490343021602010960235033401530197068039027
3016302670121041002480425035302860254030021001
5902390362034902580580406022002290292029603240
8301440134064011077010202770191026018204000178
0372041903110429038104503150200140010603050387
0870172033001250107049034302160201096023503340
1530197068039027301630267012104100248042503530
2860254030021001590239036203490258058040602200
```

Die Länge von w beträgt 368 Symbole. Beweisen Sie, dass die Kolmogorov-Komplexität $K(w)$ kleiner als 150 ist! Hinweis: Die Zahlen 7 und 437 sind bei der Lösung hilfreich.

2. Die folgende Aufgabe können Sie gerne händisch mit Papier und Bleistift erledigen, wenn Sie wollen. Sie können aber auch gerne ein paar kleine Programme schreiben, die zumindest einen Teil der Aufgaben für Sie bearbeiten (empfohlene Vorgehensweise).

Im Moodle finden Sie eine Datei `text.txt` (Quelle: Wikipedia, von Sonderzeichen bereinigt). Im Folgenden sei Σ die Menge der Symbole, die im Text verwendet werden. Bestimmen Sie für jedes Symbol $s \in \Sigma$ die relative Häufigkeit seines Vorkommens im Text. Lassen Sie dabei Whitespace-Symbole (Spaces, Tabs, Zeilenumbrüche) außen vor. Wir gehen in den folgenden Aufgaben davon aus, dass $\mathcal{P}(s)$ gleich der so bestimmten relativen Häufigkeit ist.

- Bestimmen Sie für jedes $s \in \Sigma$ seinen Informationsgehalt.
- Bestimmen Sie die Shannon-Entropie von Σ .
- Bestimmen Sie die Min-Entropie von Σ .
- Wieviele Bits werden benötigt um ein Wort der Länge 8 zu codieren, wenn jedes Symbol mittels ASCII codiert wird?
- Wie groß ist die zeichenbasierte Shannon-Entropie eines Wortes w der Länge 8, wenn seine Symbole aus Σ gewählt werden und wir die oben bestimmte Entropie von Σ zugrunde legen?
- Wie groß ist die Min-Entropie eines Wortes w wie in der letzten Aufgabe?
- Bestimmen Sie den Shannon-Informationsgehalt der folgenden Wörter:
 - ab
 - Woody
 - einen
 - beherbergt
 - Staatsbetrieben
 - Bestandteil