

Universität Leipzig Institut für Informatik Automatische Sprachverarbeitung	Algorithmen und Datenstrukturen II SS 2019 – Serie 1		
Prof. Dr. Gerhard Heyer, Dr. Jochen Tiepmar	Ausgabe am 10.04.2018	Abgabe am 17.04.2018	Seite 1/2

Algorithmen und Datenstrukturen II

SS 2019 – Serie 1

1 (10 Punkte) Huffman-Kodierung

Gegeben sei die Zeichenkette

$$S_1 = \text{BELLO } _ \text{BELLT } _ \text{WENN } _ \text{BELLO } _ \text{FAELLT}$$

- a) Zählen Sie die Häufigkeiten der Zeichen und geben Sie diese in einer Tabelle an. Versichern Sie sich, dass Ihre Zählung stimmt, damit Sie die beiden folgenden Teilaufgaben auch richtig lösen können. (1 Punkt)
- b) Konstruieren Sie anhand der in a) ermittelten Häufigkeiten den entsprechenden Huffman-Kodierungs-Baum. Gibt es mehr als zwei in Frage kommende Teilbäume, so werden diejenigen ausgewählt, die die am frühesten im Alphabet erscheinenden Zeichen enthalten. Die Zuordnung als linker und rechter Teilbaum bestimmt sich alphabetisch: der Baum mit dem früher vorkommenden Zeichen wird linker Teilbaum. Das Leerzeichen $_$ kommt alphabetisch vor dem Zeichen A.
Hinweis: Es müssen keine Heap-Bäume oder PriorityQueues für die Sortierung gezeichnet werden. (2 Punkte)
- c) Geben Sie die Code-Werte jedes Zeichens an. Jedes Kantenlabel ist nach links mit 0 und rechts mit 1 zu belegen. Kodieren Sie anschließend die ursprüngliche Zeichenkette. (2 Punkte)
- d) Gegeben sei die Zeichenkette

$$S_2 = _ \text{ABEFLNOTW}$$

Erstellen Sie den Huffman-Kodierungs-Baum für die Zeichenkette S_2 und kodieren Sie die in a) angegebene Zeichenkette S_1 damit. Folgen Sie dem Vorgehen aus a), b) und c) und geben Sie die Zwischenschritte analog an. Welchen Unterschied stellen Sie nach der Kodierung fest? (5 Punkte)

Universität Leipzig Institut für Informatik Automatische Sprachverarbeitung	Algorithmen und Datenstrukturen II SS 2019 – Serie 1		
Prof. Dr. Gerhard Heyer, Dr. Jochen Tiepmar	Ausgabe am 10.04.2018	Abgabe am 17.04.2018	Seite 2/2

2 (5 Punkte) Burrows-Wheeler Transformation

- a) Gegeben sei die Zeichenkette

$$S = \text{NANAMAGANANAS}$$

Wenden Sie die Burrows-Wheeler-Transformation wie in der Vorlesung besprochen auf S an. Alle Zähler beginnen bei 0. Geben Sie die alphabetisch sortierte Matrix der Rotationen von S und die Ausgabe der Transformation (letzte Spalte und Zeilennummer des Originalblocks) an. (3 Punkte)

- b) Führen Sie die Rücktransformation durch. Hierbei ist es ausreichend, den Gesamtpfad zu zeichnen und Start- und Endpunkte zu markieren. (2 Punkte)

3 (6 Punkte) Move-To-Front Kodierung

- a) Gegeben sei die Zeichenkette

$$S = \text{KUDELWUDEL}$$

Wenden Sie Move-To-Front Coding wie in der Vorlesung besprochen auf S an. Verwenden Sie die selbst zu recherchierenden ASCII Codes (dezimal) als initiale Arrayindizes analog zur Vorlesung. Beachten Sie, dass beim ersten Auftreten eines Buchstabens dessen Codewert nicht mehr mit dem ursprünglichen ASCII Wert des Buchstabens uebereinstimmen muss. (3 Punkte)

- b) Gegeben sei die Codefolge

$$C = 83, 68, 74, 90, 77, 85, 0, 5, 5, 5, 5, 70, 5, 5$$

Dekodieren Sie diese unter der Annahme, dass sie analog zu a) erzeugt wurde. Geben Sie die dekodierte Zeichenfolge sowie die erzeugte Zeichensortierung der Codewerte an. (3 Punkte)

4 (4 Punkte) Entropie

Gegeben sei die Zeichenkette

$$S = \text{KUDELWUDEL}$$

Berechnen Sie die Entropie dieser Zeichenkette. Nehmen Sie an, dass Sie mit einer unerforschten Sprache arbeiten, Alphabet und Zeichenwahrscheinlichkeiten also nur aus der gegebenen Zeichenkette extrapoliert werden können. Geben Sie die berechnete Entropie H sowie die zugrundeliegenden Wahrscheinlichkeiten und Informationsgehalte der Zeichen an.

Hinweis: Zur Vollständigkeit ist anzumerken, dass die Zeichen als stochastisch unabhängig angenommen werden können.