Algorithmische Grundlagen und Vermittlung der Informatik
Lehrstuhl 11 (Algorithm Engineering)
Prof. Dr. Johannes Fischer

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 9

Besprechung: 17.12.2018

Aufgabe 1

Zeigen Sie alle Datenstrukturen für den FM-Index für den Text $T =$ `YABBADABBADOO$` (nur für counting-queries):

a) Unkomprimierter FM-Index

b) H_0-komprimierter FM-Index (mit Shannon-Fano code)

c) H_2-komprimierter FM-Index (mit Huffman-code)

Aufgabe 2

Implementieren Sie ein Verfahren, das für einen Eingabetext T und einen Eingabeparamter k die Entropie k-ter Ordnung von T berechnet. Führen Sie das Programm auf dem Text von Übungsblatt 1 (Wikipedia titles) für die Werte $k = 0, 1, 2, 3$ aus. Hinweis: Suffix-Arrays!

Aufgabe 3

Zeigen Sie, dass (trotz der Optimalität) manche Codewörter bei Huffman ein längeres Codewort bekommen als bei Shannon-Fano. Warum widerspricht das nicht der Optimalität von Huffman?