

# Übung zu der Vorlesung Text-Indexierung (WS20/21)

Hinweise

Abgabe bis spätestens Montag, den 07.12.2020 um 12:00 Uhr in elektronischer Form.

**Bitte geben Sie auch ein (leeres) PDF ab, wenn Sie nur die Programmieraufgabe lösen.** Dies hilft bei der Verwaltung der Punkte in Moodle.

Des Weiteren gelten alle organisatorischen Hinweise vom ersten Übungsblatt.

## Programmieraufgabe

(6 Punkte)

Die genaue Aufgabenstellung entnehmen Sie bitte der `README.md` des Aufgaben-Repositories. Um das Ihr Repository zu erzeugen folgen Sie bitte den folgenden Link (Link nur in der Moodle-internen Version). Die einzelnen Teilaufgaben geben hierbei die folgenden Punkte:

- (a) Implementierung des naiven LCP-Array-Konstruktionsalgorithmus. (1 Punkt)
- (b) Implementierung des in der Vorlesung vorgestellten Algorithmus. (2 Punkt)
- (c) Implementierung des Algorithmus aus Abschnitt 3.7.1 im Vorlesungsskript. (2 Punkte)

Im Skript steht, dass der Vorteil des Algorithmus in Abschnitt 3.7.1 die geringere Anzahl an Cache-Misses ist. Bei einem Cache-Miss handelt es sich um den Zugriff auf Daten, die nicht im Cache vorhanden sind.

- (d) Beschreiben Sie, an welchen Stellen in den Algorithmen aus den Aufgabenteilen (b) und (c) zu Cache-Misses kommen kann. (1 Punkt)

Die Beschreibung kann entweder in Textform mit Bezug auf den Pseudocode im Skript erfolgen oder durch *deutliche* Kommentare im Quellcode.

## Eigenschaften des LCP-Arrays

(4 Punkte)

- (a) Sei  $H$  ein LCP-Array. Zeigen oder widerlegen Sie folgende Aussage. (2 Punkte)

$$\exists i \in [1, n] \text{ mit } H[i] = \ell \Rightarrow \exists j \in [1, n] \text{ mit } H[j] = \ell - 1.$$

- (b) Sei  $H$  das LCP-Array von einem Text  $T = T[1] \dots T[n]\$$  und  $H^R$  das LCP-Array des reversen Textes  $T^R = T[n] \dots T[1]\$$ . Zeigen Sie, dass folgende Aussage gilt. (2 Punkte)

$$\exists i \in [1, n] \text{ mit } H[i] = \ell \iff \exists i' \in [1, n] \text{ mit } H^R[i'] = \ell.$$