

# Text Indexing and Information Retrieval

## Übungsblatt 8

Besprechung: 30.1.2014

### Aufgabe 1 (Praxis)

Select-Anfragen können auch durch binäre Suche über der Rank-Datenstruktur in  $O(\lg n)$  Zeit gelöst werden. Implementieren Sie diese Idee, aufbauend auf Ihrer Implementierung aus der 1. Aufgabe vom 7. Übungsblatt.

### Aufgabe 2 (Theorie)

Beschreiben Sie, wie man mithilfe von Wavelet-Trees und select-Anfragen auf Bitvektoren select-Anfragen auf allgemeinen Sequenzen lösen kann (also das  $i$ -te Vorkommen eines Buchstabens  $a \in \Sigma$  finden).

### Aufgabe 3 (Theorie)

Eine weitere Operation in komprimierten Suffixbäumen sind *level ancestor queries* (LAQs): zu einem gegebenen Knoten  $v = [v_\ell, v_r]$  soll  $\text{LAQ}(v, d)$  den zur Wurzel nächsten Vorfahren von  $v$  liefern, dessen String-Tiefe mindestens  $d$  ist. Beschreiben Sie, wie diese Operation in  $O(\lg n)$  Zeit auf dem Suffix- und LCP-Array simuliert werden kann (wobei hier wie in der Vorlesung davon ausgegangen werden kann, dass Suffix- und LCP-Array-Zugriffe sowie Range Minimum-Anfragen in konstanter Zeit beantwortet werden).