

# Text Indexing and Information Retrieval

## Übungsblatt 2

Besprechung: 20.10.2014

### Aufgabe 1 (Praxis)

Implementieren Sie einen Inverted Index, der Suchanfragen mit einem Wort unterstützt. Der Einfachheit halber bietet es sich an, eine Hashtabelle (und die in Ihrer Sprache eingebauten Hashfunktionen für Strings) für das Vokabular zu verwenden (Sie können aber auch Ihren Trie aus dem letzten Übungsblatt verwenden). Testen Sie Ihren Index mit der auf der Homepage verlinkten Kollektion, in der Sie den Satz in jeder Zeile als ein einzelnes „Dokument“ ansehen können.

### Aufgabe 2 (Theorie)

In den Übungen hatten wir das Verfahren „Blind Search“ für die Suche in kompakten Tries kennengelernt. Beschreiben Sie, wie mit diesem Verfahren auch Vorgänger-Anfragen (s. Kap. 2 im Skript) beantwortet werden können. Zeigen Sie auch die Korrektheit des Verfahrens.

### Aufgabe 3 (Theorie)

Zeigen Sie, dass bei der exponentiellen Suche  $O(\sum_{i=1}^{|L_1|} \log d_i)$  maximiert wird, wenn für alle  $d_i = |L_2|/|L_1|$  gilt.