

Sprachtechnologie

10. Übung

Prof. Tobias Scheffer
Christoph Sawade

Sommer 2009

Ausgabe am: 06.07.09
Besprechung am: 13.07.09

Aufgabe 1

Indexieren und Suchen

Simulieren Sie, wie der auf Folien 9 und 10 vorgestellte Algorithmus einen invertierten Index über alle Terme des folgenden Textes aufbaut. Berücksichtigen Sie dabei, dass nur 10 Indexterme gleichzeitig in den Hauptspeicher Ihres Rechners passen. Der Text lautet:

Ich fliege mit meiner Rakete zum Mars. Ich besuche dort die auf dem Mars lebenden Marsmenschen.

- Simulieren Sie, wie Sie mit Hilfe des Index' alle Fundstellen des Wortes *Mars* suchen.
- Suchen Sie mit Hilfe des Index' nach Textstellen, an denen die Phrase *Rakete zum Mars* vorkommt.
- Bauen Sie einen Suffix-Baum über den oben stehenden Text auf. Verwenden Sie alle Wortanfänge als Indexpositionen. Simulieren Sie, wie der Algorithmus, den wir in der Vorlesung kennen gelernt haben, dabei vorgeht.
- Bauen Sie aus dem Suffix-Baum jetzt ein Suffix-Array.
- Suchen Sie mit Hilfe des Suffix-Baumes und anschließend mit Hilfe des Suffix-Arrays nach Vorkommen des Strings *Rakete zum Mars*.