

6. Übung

19. Januar 2015

Abgabe der Hausaufgaben: per Moodle bis zum 2. Februar 2015, 16:00 Uhr
Bei Fragen und Problemen können Sie sich per E-Mail/Moodle an die Betreuer wenden.

Aufgabe 1 (Aho-Corasick Algorithmus I) (4 Punkte)

Aufgabe 1.1: Präfixbaum (Trie) (1 Punkt)

Konstruieren Sie einen Präfixbaum für die Pattern $p_1 = cdddc$, $p_2 = cdcdc$, $p_3 = cdc$, $p_4 = cd$ und $p_5 = cccd$.

Aufgabe 1.2: Aho-Corasick Automat (1 Punkt)

Konstruieren Sie basierend auf dem Präfixbaum den Aho-Corasick Automat, inklusive aller Transitionen.

Aufgabe 1.3: Pattern Matching (2 Punkte)

Beschreiben Sie die einzelnen Schritte der Suche für den Eingabetext *ccdcdcdcdcdcdcdcdcdcdcd*. Geben Sie dazu an, wie der Text verarbeitet wird, welche Zustände im Automaten jeweils erreicht werden und welche Pattern an welcher Stelle matchen. Die einzelnen Knoten im Automat können Sie anhand des Pfades beschreiben, ein Knoten v_{cdd} wird also vom Startpunkt aus erreicht indem die Kanten c , d , d benutzt werden.

Abgabe: Geben Sie diese Aufgabe entweder digital in Moodle oder schriftlich auf einem Zettel ab. Vor dem Sekretariat befindet sich ein Briefkasten, dort können Sie Ihre Lösung einwerfen.

Aufgabe 2 (Aho-Corasick Algorithmus II) (3 Punkte)

Aufgabe 1.1: Implementierung (2 Punkt)

Implementieren Sie den Aho-Corasick Algorithmus in einer Programmiersprache Ihrer Wahl.

Hinweise: Ihr Programm soll per Kommandozeile gestartet werden können und zwei Parameter entgegennehmen:

1. **-p** <pattern.txt>: Textdatei mit Signaturen, eine Signatur pro Zeile
2. **-t** <text>: Zu analysierender Text

Die Ausgabe des Programms soll die gefundene Signatur sowie deren Position enthalten und die folgende Form haben:

```
Position1: Signatur1  
Position2: Signatur2  
...
```

Aufgabe 1.2: Evaluierung (1 Punkt)

Testen Sie Ihr Programm mit den Signaturen

```
herein  
rein  
ein  
in  
eins
```

und dem Text *indeinhereins*.

Abgabe: Geben Sie Ihre Implementierung und die Ergebnisse im Moodle-Textfeld ab oder laden Sie die Lösung in Moodle hoch.