



## Genomanalyse

10. Übungsblatt WS 2009

Abgabe: Dienstag 05.01.2010 vor der Übung, Besprechung: in der Übung

Bitte geben Sie auf Ihren abgegebenen Lösungen Name und Matrikelnummer an und arbeiten Sie in Gruppen von 2 oder 3 Personen zusammen.

---

### Aufgabe 22: [4+8=12 Punkte] Aho-Corasick Algorithmus

Wenden Sie den Aho-Corasick Algorithmus an auf die Patterns

$$P = \{ababc, abc, acb, ba, cba\}$$

und den Text

$$T = abaababcbabcb.$$

Wichtig: Bitte bezeichnen Sie die Knoten mit ihren path labels, das heißt: Knoten  $v$  wird mit  $L(v)$  identifiziert! Beispiel:  $v = aba$ . Bitte keine laufenden Nummern, etc., als Knotennamen vergeben.

**a) Vorbereitung (preprocessing):**

Zeichnen Sie den Aho-Corasick Automaten mit den Supply-links.

**b) Eigentliche Suche (search phase):**

Geben Sie für jeden gelesene Buchstaben  $\sigma = T[i], i = 1, \dots, |T|$  des Textes an:

- Den "aktuellen" Knoten  $current$  (in der Animation:  $q$ ) bevor  $\sigma$  bearbeitet wird.
  - Gegebenenfalls die durchlaufene Folge der Knoten  $S(current), S(S(current)), \dots$
  - Den "aktuellen" Knoten  $current$  (in der Animation:  $q$ ) nachdem  $\sigma$  bearbeitet wurde.
  - Falls es einen Treffer gibt, zusätzlich die Folge der bei der Ausgabe durchlaufenen Knoten *ohne* Verwendung von output links.
- 

Ich wünsche Ihnen allen frohe Weihnachten und ein gutes neues Jahr!