

## Übungen zur Algorithmischen Bioinformatik I

### Blatt 5

**Abgabetermin:** Dienstag, 6.6.2017, 10 s.t.

#### 1. Aufgabe (Bonus-Aufgabe):

Betrachten Sie die folgende Suchwortmenge  $S = \{ein, eis, kreisen, reise, steiles, teil\}$ .

- Konstruieren Sie einen Suchwort-Baum nach Aho-Corasick;
- Konstruieren Sie die Failure-Links;
- Markieren Sie nach der in der Vorlesung angegebenen Methode alle Knoten darin, die auf einen Treffer hinweisen;
- Wenden Sie den Algorithmus von Aho-Corasick mit dem konstruierten Suchwort-Baum auf das folgende Wort an, um alle Vorkommen der Suchworte zu finden: *eieinsteilesteilinkreisen*

#### 2. Aufgabe:

In Aufgabe 2 und 3 betrachten wir eine Optimierung des KMP-Algorithmus für den Fall an, dass das vorliegende Alphabet binär ist, d.h. dass  $\Sigma = \{0, 1\}$ . Hierbei sei  $\bar{0} = 1$  und  $\bar{1} = 0$ . Die Tabelle  $border'$  ist wie folgt definiert:

$$border'[j] = \max\{-1, i \in [0 : j - 1] : s_0 \dots s_{i-1} s_i = s_{j-i} \dots s_{j-1} \bar{s}_j\}.$$

Insbesondere ist dann wieder  $border'[0] := -1$ .

Geben Sie einen Algorithmus in Pseudo-Code an, der die Tabelle  $border'$  berechnet und mit linearer Laufzeit auskommt.

Wie sieht die Tabelle  $border'$  für das Beispiel aus Aufgabe 3 auf Blatt 4 aus?

*Hinweis:* Korrektheitsbeweis und Laufzeitanalyse nicht vergessen.

#### 3. Aufgabe:

Geben Sie eine Modifikation des Algorithmus von Knuth, Morris und Pratt in Pseudo-Code an, der auf dieser Tabelle  $border'$  basiert.

Bestimmen sie möglichst exakt eine obere Schranke für die Anzahl der Vergleiche für diesen Algorithmus (die Tabelle  $border'$  darf hierbei als gegeben betrachtet werden).

*Hinweis:* Korrektheitsbeweis und Laufzeitanalyse nicht vergessen.

**4. Aufgabe (Bonus-Aufgabe):**

Bestimmen Sie für das Wort *babbaababbabbaababbaababa* die Shift-Tabelle für den Boyer-Moore-Algorithmus aus der Vorlesung. Geben Sie dabei nicht nur die Tabelle, sondern alle Zwischenschritte für die Berechnung an.