

---

## Algorithmische Bioinformatik I

---

Abgabetermin: Freitag, der 20. Juni, 9<sup>00</sup> Uhr in Moodle

### Hausaufgabe 1

Konstruiere die Border-Tabelle für das Wort:  $s = babaabaabb$ .

Gib dabei nicht nur die Tabelle an, sondern auch alle Zwischenschritte (also auch die Ränder eines Präfixes von  $s$ , die zu einem eigentlichen Rand eines Präfixes von  $s$  erweitert werden sollten, ähnlich wie in der Vorlesung bzw. im Skript).

### Hausaufgabe 2

Betrachte die folgende Suchwortmenge  $S = \{ab, aba, abca, bc, bcaba, bcabc, c, caa, cabac\}$ .

- Konstruiere einen Suchwort-Baum nach Aho-Corasick;
- Konstruiere die Failure-Links;
- Markiere nach dem in der Vorlesung angegebenen Algorithmus alle Knoten darin, die auf einen Treffer hinweisen (inklusive der Hit-Links);
- Wende den Algorithmus von Aho-Corasick mit dem konstruierten Suchwort-Baum auf das folgende Wort an:  $cacbcababcaa$

*Hinweis:* Verwende verschiedene Farben, aus denen ersichtlich wird, welche Teile des Baumes (bzw. welche Annotationen) zu welchem Aufgabenteil gehören (zeichne ggf. den Baum mehrmals).

### Tutoraufgabe 3 (Vorbereitung bis zum 18. Juni 2025)

Betrachte den Boyer-Moore-Algorithmus, der für die Bestimmung des Shifts bei einem Mismatch nur die Weak-Good-Suffix-Rule berücksichtigt:

$$S'[j] = \min \left\{ \sigma : \begin{array}{l} (s_{j+1} \cdots s_{m-1} \in \mathcal{R}(s_{j+1-\sigma} \cdots s_{m-1}) \wedge \sigma \leq j) \vee \\ (s_0 \cdots s_{m-1-\sigma} \in \mathcal{R}(s_0 \cdots s_{m-1}) \wedge \sigma > j) \vee (\sigma = m) \end{array} \right\}$$

Gib eine Konstante  $c > 0$  sowie eine unendliche Familie  $\mathcal{F} = \{(s, t) : s, t \in \Sigma^*\}$  an, die für jedes Paar  $(m, n) \in \mathbb{N}^2$  ein Paar  $(s, t) \in \mathcal{F}$  mit  $|s| \geq m$ ,  $|t| \geq n$  und  $|s| \leq |t|$  enthält und bei dem diese Variante für eine erfolglose Suche mindestens  $c \cdot (|s| \cdot |t|)$  Zeichenvergleiche ausführt.