

---

## Algorithmische Bioinformatik I

---

Abgabetermin: Freitag, der 1. Juli, 9<sup>00</sup> Uhr in Moodle

### Hausaufgabe 1

Bestimme für das Wort  $s = s_0 \cdots s_{15} = \text{babbaababbaababa}$  die Shift-Tabelle für den Boyer-Moore-Algorithmus aus der Vorlesung.

Gib dabei nicht nur die Tabelle, sondern auch alle Zwischenschritte für die Berechnung an (wie im Skript bzw. in der Vorlesung).

### Hausaufgabe 2

Betrachte den unten angegebenen Boyer-Moore-Algorithmus, der für die Bestimmung der Shifts bei einem Mismatch nur die Extended-Bad-Character-Rule berücksichtigt. 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.

---

EBC-Boyer-Moore (char  $t[]$ , int  $n$ , char  $s[]$ , int  $m$ )

---

```
int  $i := 0$ ,  $j := m - 1$ ;
while ( $i \leq n - m$ ) do
  while ( $t[i + j] = s[j]$ ) do
    if ( $j = 0$ ) then return TRUE;
     $j--$ ;
   $i := i + j - \max\{k : (k < j \wedge s[k] = t[i + j]) \vee (k = -1)\}$ ;
   $j := m - 1$ ;
return FALSE;
```

---

### Tutoraufgabe 3 (Vorbereitung bis zum 29. Juni 2022)

Gegeben seien  $s \in \Sigma^m$  und  $t \in \Sigma^n$  mit  $m \leq n$ . Zeige, wie man aus den Z-Werten für eine geeignet gewählte Zeichenreihe (welche?) in Zeit  $O(n + m)$  feststellen kann, an welchen Positionen in  $t$  überall  $s$  als Teilwort auftritt.

*Hinweis:* Korrektheitsbeweis und Laufzeitanalyse nicht vergessen.

Am Montag, den 27. Juni findet um 16 Uhr c.t. per Videokonferenz eine Informationsveranstaltung für Studierende der Bioinformatik zum Ende des Bachelor-Studiums und zum Übergang in den Master Bioinformatik statt.