

### 3. Übung „Algorithmen der Bioinformatik I“ Abgabe am 14. bzw. 15. Mai 2003 in der Übung



Institut für Informatik  
Martin-Luther-Universität Halle-Wittenberg

---

**Aufgabe 1.** Wir betrachten das *Good Suffix Preprocessing* im Boyer-Moore-Algorithmus.

- (i) Beweisen Sie, dass  $N_j(P) = Z_{n-j+1}(P^r)$  gilt.
- (ii) Beweisen Sie, dass die Berechnung der  $L'(i)$  aus den  $N_j(P)$  korrekt ist.

**Aufgabe 2.** Implementieren Sie den Boyer-Moore-Algorithmus. Überlegen Sie sich dazu einen Algorithmus, der die  $l'(i)$  in linearer Laufzeit berechnet und erklären Sie Ihr Vorgehen. Vergleichen Sie anhand von selbstgewählten Beispielen die Laufzeiten zwischen naivem Algorithmus, Z-Algorithmus, naivem Algorithmus mit bad character rule und Boyer-Moore-Algorithmus. Dabei sollen nach der Anwendung eines Algorithmus auf ein Text-Muster-Paar folgende Werte ausgegeben werden: Anzahl und Anfangspositionen der gefundenen Muster sowie Anzahl der ausgeführten Zeichenvergleiche. Die Anzahl der ausgeführten Zeichenvergleiche soll ermittelt werden, indem ein Zähler vor oder nach **jedem** erfolgten Zeichenvergleich um eins erhöht wird. Der Grund für diese (und keine möglicherweise effizientere) Zählung ist, dass wir nun die Operation *Zeichenvergleich und Zählerinkrementierung* als eine komplexere Vergleichsoperation betrachten können und somit die Laufzeiten für die verschiedenen Algorithmen in gleichem Maße beeinflusst werden.

*Hinweis:* Ein paar mögliche Testfälle sind auf der Internetseite zur Vorlesung verfügbar.

**Aufgabe 3.** Überlegen Sie sich, warum der Boyer-Moore-Algorithmus angewandt auf englische oder deutsche Texte für längere Muster schneller läuft, als für kürzere.