

5. Übung zur Vorlesung “Algorithmische Massenspektrometrie”

Wintersemester 2018/2019

Sebastian Böcker, Martin Hoffmann

Ausgabe: 22. November 2018, Abgabe: 29. November 2018

- Gegeben sei das Alphabet $\Sigma = \{a, b, c, d\}$ mit $\mu(a) = 5$, $\mu(b) = 9$, $\mu(c) = 13$ und $\mu(d) = 17$.
 - Berechnen Sie mit dem Round-Robin Algorithmus die ERT Tabelle und bestimmen Sie anhand der Tabelle alle Compomere mit Masse 31.
 - Was ist die Frobenius-Nummer und wie lässt sie sich mit Hilfe der ERT Tabelle berechnen?

(5 Punkte)
- Beweisen Sie, dass die Anzahl der (nicht-negativen) Zahlen, für die es keine Zerlegung aus $\Sigma = \{a_1, \dots, a_k\}$ mit $\mu(a_1) < \dots < \mu(a_k)$ gibt, gleich $\sum_{r=0}^{\mu(a_1)-1} \lfloor \frac{N[k,r]}{\mu(a_1)} \rfloor$ ist, wobei N die ERT Tabelle über Σ ist.

(3 Punkte)
- Bei der Zerlegung von Massen in Compomeren ist manchmal die Forderung sinnvoll, die Vorkommen eines Buchstabens im Compomer auf ein Intervall zu begrenzen. Beispielsweise macht es Sinn von einem organischen Molekül zu verlangen, dass seltene Elemente wie Jod nur maximal 3 Mal vorkommen dürfen. In diesem Fall spricht man von einem Upperbound für den Buchstaben I . Andersherum ist manchmal bekannt (beispielsweise durch eine Isotopenmusteranalyse), dass ein bestimmtes Element wie Schwefel mindestens einmal vorkommen muss. Dann gäbe es einen Lowerbound für den Buchstaben S .
 - Wie lassen sich Lowerbounds implementieren ohne den eigentlichen Algorithmus zum Aufzählen von Compomeren zu ändern. Hinweis: Ein einfaches Preprocessing der Eingabe und Postprocessing der Ausgabe reichen hierfür aus.
 - Beschreiben Sie, wie man den Algorithmus zum Aufzählen von Compomeren über die ERT Tabelle umändern müsste, um einen Upperbound zu implementieren.

(3 Punkte)
- Üblicherweise läuft das Backtracking einer DP-Tabelle asymptotisch schneller ab als die Berechnung der Tabelle. Beispiel: Das Backtracking in Smith-Waterman hat Laufzeit $O(m+n)$, die Tabellenberechnung hingegen $O(m \cdot n)$. Sowohl beim Round-Robin Algorithmus als auch beim DP-Ansatz zum Zählen der Compomere benötigt das Backtracking (Aufzählen der Compomere) in der Praxis (für Alphabete von Aminosäuren oder chemischen Elementen) deutlich mehr Zeit als die Berechnung der Tabelle selbst.
 - Begründen Sie, warum dem so ist.

- (b) In der letzten Übungsserie Aufgabe 2.) ging es um einen Algorithmus zum Zählen von Compomeren. Ein Backtracking über dessen DP-Tabelle ermöglicht ebenfalls das Aufzählen aller Compomere. Warum ist der Round-Robin Algorithmus in der Praxis dennoch schneller - insbesondere, wenn man viele Massen nacheinander zerlegen will?

(3 Punkte)