



# Algorithmen auf Sequenzen

SoSe 2022

Projekt 1

## Patternmatching

**Abgabetermin: 04.05.2022**

**Ansprechpersonen:**

- Georg Gruenert: [georg.gruenert@mni.thm.de](mailto:georg.gruenert@mni.thm.de)
- Sujaya Shrestha: [sujaya.shrestha@mni.thm.de](mailto:sujaya.shrestha@mni.thm.de)
- Cornelia Meckbach: [cornelia.meckbach@mni.thm.de](mailto:cornelia.meckbach@mni.thm.de)

**Abgabe:** Beschreiben Sie Ihr Vorgehen und Ihre wichtigsten Erkenntnisse in einem Dokument. Fügen Sie dem Dokument auch Ihren kommentierten Quellcode hinzu, sowie wichtige Ausgaben und speichern Sie es als PDF. Vergessen Sie nicht Ihren Namen dazuzuschreiben. Speichern Sie den Quellcode inkl. ReadMe-file (für Ausführungsdetails), sowie das oben beschriebene PDF-Dokument in einem Ordner und laden Sie diesen entweder als .zip oder als .tar.gz in Moodle hoch. (Quellcode kann gerne auch als Jupyter Notebook (.ipynb) abgegeben werden)

### Aufgabe 1: SimpleSearch und Horspool-Algorithmen

Implementieren Sie in der Programmiersprache Python die in der Vorlesung behandelten Algorithmen

- a) SimpleSearch
- b) Horspool

so dass für ein beliebiges übergebenes Pattern  $p$  und einen beliebigen Text  $t$  das exakte Vorkommen aller Pattern  $p$  in  $t$  mit Positionen zurückgegeben wird.

### Aufgabe 2: Vergleiche zählen

Ergänzen Sie Ihre beiden Algorithmen aus Aufgabe 1, so dass alle getätigten Vergleiche des jeweiligen Algorithmus gezählt werden. Geben Sie für jeden Lauf des Algorithmus die Vergleichsanzahl aus.



### Aufgabe 3: Testen: Schneewittchen



Gegeben ist Ihnen die Datei *Schneewittchen\_(1850).txt*.

- Überführen Sie alle Buchstaben in Kleinbuchstaben.
- Ändern Sie alle Umlaute (ä,ü,ö,ß) um in (ae,ue,oe,ss), sowie Anführungszeichen „...“ in "...".
- Finden Sie in der Datei alle Vorkommen von „zwerg“ und geben Sie es mit der entsprechenden Position zurück (entweder Zeile+Position in Zeile oder Position in gesamten Text).
- Geben Sie zusätzlich für alle Treffer die kompletten Sätze aus, in denen die Treffer enthalten sind.
- Finden Sie in dieser Datei alle Vorkommen von „zwerg“, die nicht zum Wort „zwerge“ gehören. Wie könnten Sie hier vorgehen?
- Wie viele Vergleiche benötigen Sie jeweils in c) und e)?

### Aufgabe 4: Vergleich SimpleSearch & Horspool



Gegeben ist die Datei *ABAB.txt*. und die Pattern

- ABA
- CCC

Betrachten sie die Datei und berechnen Sie im Vorfeld, wie viele Vergleiche Sie für den jeweiligen Algorithmus benötigen.

Geben Sie für diese Datei

- die Anzahl der Treffer zurück, die Sie mit SimpleSearch bzw. dem Horspool Algorithmus finden
- die Anzahl der benötigten Vergleiche für jeden Algorithmus an.

Vergleichen Sie die Performance von SimpleSearch und Horspool und erläutern Sie Ihre Beobachtungen. Vergleichen Sie Ihre zu vorigen Berechnungen mit Ihren tatsächlichen Ergebnissen.