

Prof. Dr. Knut Reinert
Enrico Siragusa
Sascha Meiers
Christoph Hartmann

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmen und Datenstrukturen in der Bioinformatik

Erstes Übungsblatt WS 11/12

Abgabe Montag, 24.10., 15:00 Uhr

Name: _____ Übungsgruppe: A B C

Matrikelnummer: _____

Niveau I

Aufgabe 1: Pattern Matching

Finden Sie das Pattern $P = \text{ATAC}$ im Text $T = \text{GATATACC}$ sowohl mit dem **Horspool-** als auch mit dem **Shift-Or-**Algorithmus.

Erstellen Sie dazu erst eine Tabelle mit den **Safe-Shifts** respektive den **Bitmasken**.

Notieren sie den Zustand der Bitmaske d des Shift-Or-Algorithmus nach jedem Schritt. Wieviele Buchstabenvergleiche werden bei den Algorithmen jeweils benötigt?

Aufgabe 2: Approximate Matching

Modifizieren Sie den unten angegebenen Algorithmus für den **naiven Pattern-Matching-Algorithmus**, so dass auch Vorkommen des Patterns mit bis zu k Mismatches gefunden werden.

Für $k = 1$ matcht RABE z.B. mit ROBE und GABE, aber nicht mit RUEBE.

```
FOR i in 0..|T|-|P|:
  matches=0
  FOR j in 0..|P|-1:
    IF T[i+j]==P[j]:
      matches++
    ELSE:
      BREAK
  IF matches==|P|:
    print "Pattern gefunden an Position "+i
```

Niveau II

Aufgabe 3: Wildcards

Erweitern Sie den Shift-Or-Algorithmus so, dass eine beliebige Anzahl von ‘?’ Wildcards im Text und Pattern erlaubt ist (Pseudocode angeben). Das Zeichen ‘?’ matcht jedes andere Zeichen des Alphabets.

Zeigen Sie, wie der modifizierte Algorithmus auf dem folgenden Beispiel arbeitet:

$$P = A?TT$$
$$T = AGTAATT?G$$

Programmieraufgabe (Abgabe Montag, 31.10.2011, 15:00)

Aufgabe 4: Implementieren Sie den Horspool-Algorithmus in C++ und geben sie alle Stellen (Zählung beginnt bei 0) an denen P im Text T vorkommt untereinander aus. T und P werden Ihrem Programm über die Kommandozeile übergeben. Beispiel:

```
sorglos@bitterfeld:~$ ./aufgabe1 "Simon, send more money" "mo"
2
12
17
sorglos@bitterfeld:~$
```

Beachten Sie die Praktikumsanweisung unter <https://www.mi.fu-berlin.de/w/ABI/AlDaBiWS11>.