

Prof. Dr. Knut Reinert
Sascha Meiers
Thimo Wellner

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmen und Datenstrukturen in der Bioinformatik

Siebtes Übungsblatt WS 13/14

Abgabe Freitag, 20.12., 12:00 Uhr

Name:

Matrikelnummer:

Niveau I

Aufgabe 1: BLAST/Aho-Corasick

Gegeben ist ein Pattern $P = \text{HILAM}$

- Finden Sie alle k -mere, die zu einem beliebigen k -mer aus P einen Score von mindestens 13 aufweisen. Die Fenstergröße ist auf $k = 3$ festgelegt. Benutzen Sie zum Scoring die BLOSUM62-Matrix.
- Erstellen Sie aus der Liste der k -mere einen Such-Trie. Markieren Sie dabei die finalen Knoten.
- Erweitern sie den Trie zu einem Aho-Corasick-Automaten und zeichnen Sie alle nicht-trivialen Supply-Links ein, also jene die nicht zur Wurzel zurückgehen.

Aufgabe 2: Suche im Aho-Corasick Automaten

Nun sind zwei Datenbanken $D1 = \text{ILHIVAVHIIM}$, $D2 = \text{HLLALAM}$ gegeben.

- Suchen Sie die *seeds*, also alle Vorkommen eines der Tupel in den Datenbanken $D1$ und $D2$ mit dem Aho-Corasick Algorithmus. Als Ergebnis genügt es, die Stellen zu markieren.