

Foundations of Sequence Analysis
Winter 2005/2006

Übungen

Beispielfragen zu Kapitel 3, Besprechung am 05.01.2006 bzw. 09.01.2006.

Die folgende Liste von Beispielfragen soll eine Hilfestellung bei der Wiederholung des Stoffs von Abschnitt 3 der Vorlesung „Grundlagen der Sequenzanalyse“ sein. Alle Fragen lassen sich mit Hilfe des Skripts *Foundations of Sequence Analysis* lösen. Die Liste definiert nicht den vollständigen Prüfungsstoff zu der Vorlesung. Die Grundlage dafür ist das Skript.

1. Gegeben ein endliches Alphabet \mathcal{A} , wie ist die Menge \mathcal{A}^* aller Worte über \mathcal{A} definiert? Ist sie endlich oder unendlich? Enthält sie nur Worte endlicher oder auch Worte unendlicher Länge?
2. Was sind Teilworte, Suffixe und Präfixe einer Sequenz s ?
3. Was ist eine Metrik, was eine Pseudo-Metrik? Geben Sie drei Beispiele für Metriken.
4. Welche Distanzmaße auf Sequenzen kennen Sie?
5. Welche Edit-Operationen gibt es, was ist ein Alignment zweier Sequenzen?
6. Ist die Menge aller Alignments zweier gegebener Sequenzen endlich oder unendlich? Wovon hängt die Größe dieser Menge ab?
7. Wie definiert man Kostenfunktionen über Alignments, welche Kostenfunktionen kennen Sie?
8. Wie definiert sich die Edit-Distanz zweier Sequenzen? Was ist ein optimales Alignment?
9. Wie berechnet man Edit-Distanz zweier Sequenzen? Welchen Zeit- und Platzbedarf hat Ihr Algorithmus?
10. Wie heißt das generelle Schema, nach dem dieser Algorithmus vorgeht? Was muss man beachten, um einen exponentiellen Zeitaufwand zu vermeiden?
11. Wie berechnet man ein optimales Alignment zweier Sequenzen, wie alle optimalen Alignments?
12. Was ist die grundlegende Idee bei der Beschleunigung der Berechnung der Edit-Distanz für einfache Kostenfunktionen wie die *Simple Levenshtein Distance* oder die *Unit Edit Distance*? Wie funktionieren die Algorithmen im Detail? Was sind ihre Laufzeiten im *worst case*?
13. Wie lautet das Problem der Berechnung lokaler Alignments? Was muss dabei für die Score-Funktion gelten? Wie kann man lokal optimale Alignments effizient berechnen? Was ist die Laufzeit, was der Speicherbedarf?
14. Was ist die Idee von Filter-Algorithmen bei der Sequenzdatenbanksuche?
15. Wie ist die *Maximal Matches Distance* zwischen zwei Sequenzen definiert? Ist diese eine Metrik? Wie ist der Zusammenhang zur Edit-Distanz mit Einheitskosten?

16. Wie ist die q -Gramm-Distanz zwischen zwei Sequenzen definiert? Ist diese eine Metrik? Mit welchem Aufwand lässt sie sich berechnen?
17. Wie ist das Fasta-Ähnlichkeitsmodell definiert? Wie lässt es sich berechnen?
18. Wie ist das BlastP-Ähnlichkeitsmodell definiert? Wie lässt es sich berechnen? Welche Nachverarbeitungsschritte sind in BlastP implementiert?