Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2015/2016
Prof. Dr. Jens Stoye · M.Sc. Linda Sundermann

http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/SequenzAnalyse

Präsenzübungsblatt 1, Woche 44/2015

Aufgabe 1 (Sequenzwahrscheinlichkeiten)

1. Wie wahrscheinlich ist es, die DNA-Sequenz $s = \texttt{TATAAA}$ per Zufall zu erhalten? Nimm dabei an, dass jede Base mit derselben Wahrscheinlichkeit auftritt. Gib auch die Formel an, die du zur Berechnung benutzt hast.
2. Wieso wird diese Sequenz nicht nur aufgrund von Zufall in einem Eukaryoten-Genom auftauchen?
3. Gegeben sei nun eine DNA-Sequenz der Länge 9. Wie wahrscheinlich ist es, dass genau 4 `A`s darin vorkommen? Gib auch hierzu eine Formel an.
4. Wie wahrscheinlich ist es, dass in einer Sequenz der Länge 9 mindestens ein `T` vorkommt?

Aufgabe 2 (Anzahl von Subsequenzen und Substrings)

1. Wie viele Subsequenzen der festen Länge k hat ein String der Länge n?
2. Wie viele Subsequenzen mit den Längen $0 \leq k \leq n$ hat ein Wort der Länge n insgesamt?
3. Wie viele Substrings der festen Länge k hat ein String der Länge n?
4. Wie viele Substrings mit den Längen $0 \leq k \leq n$ hat ein Wort der Länge n insgesamt?

Versuche, dir die Formeln jeweils anschaulich klar zu machen und fertige kleine Skizzen an.

Aufgabe 3 (Metriken)

1. Drei Eigenschaften müssen erfüllt sein, damit die binäre Operation $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ auf der Menge $\mathcal{X}$ eine Metrik ist:

$$d(x, y) = 0 \leftrightarrow x = y \tag{1}$$
$$d(x, y) = d(y, x), \forall x, y \in \mathcal{X} \tag{2}$$
$$d(x, y) \leq d(x, z) + d(z, y), \forall x, y, z \in \mathcal{X} \tag{3}$$

   Beweise, dass aus diesen Definitionen die Nichtnegativität folgt:

$$d(x, y) \geq 0, \forall x, y \in \mathcal{X}$$

2. Erläutere, wieso die folgende Metrik auf den Elementen der Potenzmenge $\mathcal{P}(Z)$ einer Menge Z auch *Mengentransformationsdistanz* genannt wird:

$$d_\infty(X, Y) = \max\{|X \setminus Y|, |Y \setminus X|\}$$