

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024

Prof. Dr. Jens Stoye · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2024winter/sa>

Präsenzübungsblatt 5, vom 18./19.5.2024

Aufgabe 1 (Approximative Textsuche)

Gegeben seien das Muster $s = \text{CDBA}$ und der Text $t = \text{CACDABADCDBACBD}$.

1. Verwende die *Cutoff*-Variante von Sellers' Algorithmus mit Einheitskosten und berechne die Edit-Matrix D von s und t mit bis zu $k = 1$ Fehlern. Markiere in jeder Spalte den *last essential index*.
2. Markiere in D alle Positionen, an denen ein Vorkommen von s mit bis zu $k = 1$ Fehlern in t endet. Gib für jede dieser Endpositionen ein korrespondierendes Alignment an.
3. Warum ist es sinnvoll, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? (Stichwort: *Runs*.) Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?