

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa>

Präsenzübungsblatt 10, vom 20./21.6.2022

Aufgabe 1 (Exakte Textsuche in BW-transformierten Texten)

Gegeben sei der Text

$$t = \text{TACGCACGGCAAGCCCGTGC\$},$$

seine Borrows-Wheeler-Transformation

$$\text{BWT}(t) = \text{CCTCAGGGGCAACTGCACC\$G}$$

und sein Suffix Array

$$\text{POS}(t) = [20, 10, 1, 5, 11, 19, 9, 4, 13, 14, 2, 6, 15, 18, 8, 3, 12, 7, 16, 0, 17],$$

wobei hier der Text t mit Startposition 0 indiziert wurde.

1. Konstruiere F , die lexikographische Sortierung von t .
2. Suche die folgenden 10 Muster in $\text{BWT}(t)$ und gib ihre Startpositionen mithilfe von $\text{POS}(t)$ an:
 - GCA
 - CGTG
 - GCAC
 - TGC
 - CGCAC
 - GC
 - AAGC
 - CGCA
 - ACGC
 - CA