

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Präsenzübungsblatt 8, Woche 25/2018

Aufgabe 1 (Exakte Textsuche in BW-transformierten Texten)

Gegeben sei der Text $t = \text{TACGCACGGCAAGCCCGTGC\$}$, seine Borrows-Wheeler-Transformation $bwt(t) = \text{CCTCAGGGGCAACTGCACC\$G}$ und sein Suffix Array

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

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 $bwt(t)$ und gib ihre Startpositionen mithilfe von $pos(t)$ an:

- GCA
- CGTG
- GCAC
- TGC
- CGCAC
- GC
- AAGC
- CGCA
- ACGC
- CA