Assignment 02

Algorithms for Sequence Analysis

Sven Rahmann and Jens Zentgraf

30.04.2024, due **6.05.2024**

02.1: Knuth-Morris-Pratt Example (4 Theory)

- $\Sigma := \{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$,
- $P :=$ BBABBBAB,
- $T :=$ BBABBBACBBBBABBBAB.
1. Compute the lps table for P.
2. Which DFA state are we in after reading each text character?

Algorithmic Bioinformatics

UNIVERSITÄT
DES
SAARLANDES

ZBI ZENTRUM FÜR
BIOINFORMATIK

2

02.2 Aho Corasick (4 Theory)

Given the set of patterns $P = \{aho, house, sea, rhesus, emu, user, us, use\}$,
- build the trie,
- compute the lps and extend the trie with the lps links,
- compute the output function.

Algorithmic Bioinformatics

UNIVERSITÄT
DES
SAARLANDES

ZBI ZENTRUM FÜR
BIOINFORMATIK

3

02.3: Recognizing cyclic permutations (4 Theory)

Given two strings s, t of the same length n,
how can you efficiently decide if one is a cyclic permutation of the other?
Def.: s is a cyclic permutation of t, if `s == t[i:] + t[:i]` for some i.

For example, `0123456` and `2345601` are cyclic permutations of each other,
but `6543210` is not a cyclic permutation of the other two (it's a reversal).
Also, `babbababbba` and `ababbbababb` are cyclic permutations of each other.

Tasks

1. Show that "is a cyclic permutation of" is an equivalence relation on strings.
2. Give an algorithm that takes two strings as input and outputs `True` or `False`.
   Your algorithm should run in $O(n)$ time.

Algorithmic Bioinformatics

UNIVERSITÄT
DES
SAARLANDES

ZBI ZENTRUM FÜR
BIOINFORMATIK

4

02.4: Shift-Or Algorithm (4 Programming)

- Implement the Shift-Or algorithm.
  You can adapt the code for the Shift-And algorithm from the slides.
- Make sure it runs correctly by running some tests,
  e.g. the example pattern and text from the lecture.