Jutge.org

The Virtual Learning Environment for Computer Programming

Read mapping 2

X38913_en

Sequence reads are mapped to a reference genome using some indexing data structure, one of which is the suffix array. The suffix array of a string is a sorted array of the suffixes of the string, that is, an array of the (starting) positions of the sorted suffixes of the string.

For example, the genomic sequence fragment TATAAT has the suffixes TATAAT at position 1, ATAAT at position 2, TAAT at position 3, AAT at position 4, AT at position 5, and T at position 6. The sorted suffixes are AAT at position 4, AT at position 5, ATAAT at position 2, T at position 6, TAAT at position 3, and TATAAT at position 1. Therefore, the suffix array of TATAAT is (4,5,2,6,3,1).

Write pseudocode, Python code, and C++ code for the read mapping problem. Make two submissions, including the pseudocode as a comment to both the Python and the C++ code.

Input

The input is a string s (a genomic sequence) over the alphabet $\Sigma = \{A, C, G, T\}$.

Output

The output is the suffix array of s.

Sample input	Sample output
TATAAT	4
	5
	2
	6
	3
	1

Hint

Implement and use the bubble-sort algorithm to sort suffixes.

Problem information

Author: Gabriel Valiente

Generation: 2021-10-07 18:58:32

© *Jutge.org*, 2006–2021. https://jutge.org