

M03S01 [Arrays and Strings]: Exercise 09

CS110: Computing Lab
Department of CSE, IIT Guwahati
Jan-May 2018

Problem Description

A life science researcher is interested in a particular genomic region of length n in a bacterial genome. Bacterial genome consists of the four nucleotides represented by 'A', 'T', 'G' and 'C'. In other words the genomic region of interest is a string of characters x , where $x \in \{A, T, G, C\}$. The researcher wants to mask the sub-region where no nucleotide has been repeated, i.e., she wants to mask the longest sub-region which does not have repeating nucleotides. Write a program to help the life science researcher. Your program can expect that the genomic region of length n is given.

Your program should output the following:

- start index of a longest sub-region
- end index of the same sub-region and
- the longest sub-region without repeating nucleotides.

[Remark: Your program must print at least one such sub-region if there are more than one longest sub-region.]

Example 1:

Input Genomic region: ATAAATT

Expected Output: Start-index = 0, End-index = 1, sub-region = AT Start-index = 4, end-index = 5, sub-region = AT start-index = 1, end-index = 2, sub-region = TA