Subwords 1 X85229_en

A nucleic acid or amino acid sequence of length n can be seen as composed of a number of possibly overlapping k-mers or words of length k, for $1 \leq k \leq n$. An interesting problem is the generation of all the words of length k contained in a genomic sequence with n nucleotides, for all k with $1 \leq k \leq n$. That is, the generation of all the subwords of a genomic sequence of length n.

Write pseudocode, Python code, and C++ code for the subwords problem. The program must implement and use the subwords function in the pseudocode, which must be iterative and is not allowed to perform input/output operations. 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 a sorted list of all the nonempty subwords of s, without repetitions.

Sample input

```
TATAAT
```

Sample output

```
A
AA
AAT
AT
ATA
ATAA
ATAAT
T
TA
TAA
TAAT
TAT
TATA
TATAA
TATAAT
```

Problem information

Author : Gabriel Valiente
Generation : 2021-11-03 10:41:01