

1K Generate the Frequency Array of a String

Frequency Array Problem

Generate the frequency array of a DNA string.

Input: A DNA string *Text* and an integer *k*.

Output: The frequency array of *k*-mers in *Text*.

GTTGGTGACGTC →

A	C	G	T
1	2	5	4

Formatting

Input: A DNA string *Text* and an integer *k*.

Output: A space-separated list of integers representing the frequency array of *k*-mers in *Text*).

Constraints

- The length of *Text* will be between 1 and 10^3 .
- The integer *k* will be between 1 and 10^1 .
- *Text* will be a DNA string.

Test Cases

Case 1

Description: The sample dataset is not actually run on your code.

Input:

ACGCGGCTCTGAAA

2

Output:

2 1 0 0 0 0 2 2 1 2 1 0 0 1 1 0

Case 2

Description: This dataset checks if you have an off-by-one error at the end of *Text* (i.e. you are not counting the last k -mer in *Text*). There are three instances of AA (**AAA**AC, AA**AA**C, and AAA**AC**), but there is one instance of AC at the end (AAA**AC**).

Input:

AAAAC

2

Output:

3 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Case 3

Description: This dataset checks if you have an off-by-one error at the beginning of *Text* (i.e. you are not counting the first k -mer in *Text*). There are two instances of AA (TT**AAA** and TTAA**A**), but there is one instance of TA (T**T**AAA) and one instance of TT (**TT**AAA).

Input:

TTAAA

2

Output:

2 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1

Case 4

Description: This dataset checks if your code actually increments each count, or if your code instead just sets the count equal to one each time. In other words, this dataset checks if your code is doing something like $\text{array}[k\text{-mer}] = 1$ instead of $\text{array}[k\text{-mer}] += 1$.

Input:

AAA

2

Output:

2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Case 5

Description: A larger dataset of the same size as that provided by the randomized autograder. Check input/output folders for this dataset.