Approximate Pattern Matching Problem

Input: Strings *Pattern* and *Text* along with an integer d.

Output: All starting positions where *Pattern* appears as a substring of *Text* with at most d mismatches.

SAMPLE DATASET:

Input:

ATTCTGGA

CGCCCGAATCCAGAACGCATTCCCATATTTCGGGACCACTGGCCTCCACGGTACGGACGTCAATCAAAT

3

Output:

6 7 26 27

The sample dataset is not actually run on your code.

TEST DATASET 1:

<u>Input</u>:

AAA

TTTTTTAAATTTTAAATTTTTT

2

<u>Output</u>:

4 5 6 7 8 11 12 13 14 15

      This dataset checks if you are only counting instances where the number of mismatches is exactly equal to d (i.e. ignoring instances where $mismatch < d$).

TEST DATASET 2:

Input:

GAGCGCTGG

GAGCGCTGGGTTAACTCGCTACTTCCCGACGAGCGCTGTGGCGCAAATTGGCGATGA
AACTGCAGAGAGAACTGGTCATCCAACTGAATTCTCCCCGCTATCGCATTTTGATGC
GCGCCGCGTCGATT

2

Output:

0 30 66

       This dataset checks if your code has an off-by-one error at the beginning of *Text* (i.e. your code is not checking the the leftmost substring of *Text*).

TEST DATASET 3:

<u>Input</u>:
AATCCTTTCA

CCAAATCCCCTCATGGCATGCATTCCCGCAGTATTTAATCCTTTCATTCTGCATATAA
GTAGTGAAGGTATAGAAACCCGTTCAAGCCCGCAGCGGTAAAACCGAGAACCATGA
TGAATGCACGGCGATTGCGCCATAATCCAAACA

3

<u>Output</u>:
3 36 74 137

       This dataset checks if your code has an off-by-one error at the end of *Text* (i.e. your code is not checking the the rightmost substring of *Text*).

TEST DATASET 4:

CCGTCATCC

CCGTCATCCGTCATCCTCGCCACGTTGGCATGCATTCCGTCATCCCGTCAGGCATACT
TCTGCATATAAGTACAAACATCCGTCATGTCAAAGGGAGCCCGCAGCGGTAAAACC
GAGAACCATGATGAATGCACGGCGATTGC

3

Output:
0 7 36 44 48 72 79 112

This dataset checks if your code is correctly accounting for overlapping instances of *Pattern* in *Text*.

TEST DATASET 5:

<u>Input</u>:
TTT
AAAAAA
3

<u>Output</u>:
0 1 2 3

This dataset checks if you are only counting instances of *Pattern* with less than *d* mismatches (as opposed to instances of *Pattern* with less than <u>or equal to</u> *d* mismatches).

TEST DATASET 6:

<u>Input</u>:
CCA
CCACCT
0

<u>Output</u>:
0

This dataset checks if your code works with input where $d = 0$ (i.e. only perfect matches are allowed).