

90 Find All Approximate Occurrences of a Collection of Patterns in a String

Multiple Approximate Pattern Matching Problem

Find all approximate occurrences of a collection of patterns in a text.

Input: A string *Text*, and a collection of strings *Patterns*, and an integer *d*.

Output: All positions in *Text* where a string from *Patterns* appears as a substring with at most *d* mismatches.

```
$BANANAS
ANANAS$B
ANAS$BAN
AS$BANAN 0 2 4
BANANAS$ BANANAS$
NANAS$BA NAN NAS
NAS$BANA NAN
S$BANANA
```

Formatting

Input: A string *Text*, a space-separated list of strings *Patterns*, and an integer *d*

Output: A newline-separated list of strings from *Patterns*. Each *Pattern* in *Patterns* is followed by a colon (":") and a space-separated list of starting indices in *Text* where *Pattern* appears as a substring with at most *d* mismatches.

Constraints

- The length of *Text* will be between 1 and 10^4 .
- The number of patterns in the string-set *Patterns* will be between 1 and 10^3 .
- The length of any one pattern in *Patterns* will be between 1 and 10^2 .

Test Cases

Case 1

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

Input:

```
ACATGCTACTTT
ATT GCC GCTA TATT
1
```

Output:

```
ATT: 2 7 8 9
GCC: 4
GCTA: 4
TATT: 6
```

Case 2

Description: *Patterns* contains partial and complete matches.

Input:

```
ACGT
GG AC
1
```

Output:

```
GG: 1 2
AC: 0
```

Case 3

Description: *Patterns* contains no matches.

Input:

```
GGGGG
TT AA
1
```

Output:

```
TT:
AA:
```

Case 4

Description: *Patterns* contains no matches, but has exactly d mismatches.

Input:

GG

TT AA

2

Output:

TT: 0

AA: 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.