Assignment 2. Fit Alignment

Due date: Feb. 5, 11:59pm.

Write a program to compute both the optimal fit alignment of two sequences.

Input:
A single FASTA-like file with exactly two non-empty sequences in it, for example:

```
>seq 1
KQKFFHI
>seq 2
MEKQQEFFHILOP
```

The program will find a substring $s$ from the second sequence, so that the alignment score between $s$ and the first sequence is maximized. Also, the program should output the alignment. The scoring scheme used for the alignment is Match = 1, mismatch = -1, indel = -1.

Output:
Print the alignment score and actual alignment to standard output in the following format:

```
4
KQK-FFHI
KQKFFEHH
```

The first line is the alignment score. The second line is the first sequence (possibly with inserted dash symbols), and the third line is the substring from the second sequence (possibly with inserted dash symbols). Ensure that the last line contains a line break `\n` at the end.

Command line:
The command line used to call your program will be:

```
run.sh inputFile
```

Test:
For each test case, you need to provide two files: n.in, and n.out. Here the n is to be replaced by 0-9. The content of n.out should be the same as the output of the correct program with input files n.in.

When a program is tested against a test case, several cases can happen:

1. If the test case’s score does not match the alignment’s actual score, then the test case is regarded as invalid.
2. If the program’s output is invalid (either not a valid alignment or the score is wrong), then the program fails.
3. Else if the program's score is better than the test case's score, then the test case is invalid.
4. Else if the score is less than the test case’s score, then the program fails.
5. Else the program passes.

The marking is the same as Assignment 1. Suppose your test cases covers $x$ different test classes and fails $x'$ different students' programs. Moreover, your program passes $y$ test classes. Additionally, the TA determines that $z$ of your test cases are invalid cases. Your will receive $p = \frac{x + x'}{2} + y - z$ points for the assignment. $p$ will be normalized so that the best student will receive 100% of the assignment mark. The normalization formula will be determined during the marking.

**Detailed specifications:**
The main purpose of this assignment is to test your alignment program is correctly implemented. Therefore, most of these specifications are listed here to restrict the test cases. If a test case does not follow these specifications it is regarded as invalid. So your program can assume the input files are valid.

The FASTA files in the test cases must follow the following format.
1. It contains exactly two non-empty protein sequences.
2. Each protein entry’s first line is a header line that starts with a ‘>’ sign; followed by one or more lines for the amino acid sequences.
3. The sequence can be in multiple lines, in which case, the line breaking characters should be discarded and the multiple lines are concatenated together to form the sequence.
4. The sequence should contain only English letters, digits, and punctuations (except for the greater sign (>) and dash (-)). Whitespace characters (whitespace and tab) are not allowed. Greater sign (>) and dash (-) are not allowed to appear in the sequence.
5. Empty lines are allowed and will be discarded.