CCO '97 P6 - Aligning DNA

Time limit: 0.15s   **Memory limit:** 64M

Canadian Computing Competition: 1997 Stage 2, Day 2, Problem 3

You are a biologist studying various patterns in DNA sequences. You need a program that will find regions of similarity in a pair of sequences, and align the similar regions by inserting spaces at the prefix or suffix of the sequences.

A DNA sequence is represented as a string of characters from the set $\{A, C, G, T\}$.

For example, if you were given the sequences:

```
AATCGA
TCGAGG
```

your program would insert two spaces before the second sequence to align the `TCGA`. The aligned sequences would be:

```
AATCGA--
--TCGAGG
```

where the dashes represent spaces. In every alignment, the lengths of the two sequences (including spaces) must be equal, although the original sequences need not be of the same length.

We will assign a number of points to each alignment, and we will call the alignment with the highest number of points the best alignment. To calculate the number of points, we proceed character by character from the beginning of the sequence. For each position, we get a certain number of points:

- if one (or both) of the sequences contains a space at that position, we lose one point;
- if the two characters in a given position are different, we lose one point;
- if the two characters in a given position are the same, we get three points.

For example, with the alignment:

```
ACT
-GT
```

we would lose one point for the first position `(A,-)`, another point for the second position `(C,G)` and gain three points for the last position `(T,T)`, for a total of one point.

Your program will be given a pair of DNA sequences. It must find the best alignment, along with the number of points given to that alignment. If there is more than one alignment with the highest number of points, any one of the best alignments is acceptable.

Input Specification

The input will consist of two lines, each line no longer than 100 characters. Each line represents a DNA sequence, and you are to align the two sequences (i.e., align the second line with the first line; the length of the second line will be the same or smaller length than the first line). The input will only contain letters from `ACGT`.

Output Specification

The output is to consist of three lines. The first line will be a single integer representing the number of points assigned to that alignment. The next two lines will consist of the aligned sequences, in the same order as they were read in, with spaces represented by dashes `-`.

Sample Input

```
ACT
GT
```

Sample Output

```
1
ACT
-GT
```