

Career Fair 2014

Programming Competition

Biologists have sequenced the genes of a pre historic family of organisms and they have stored the sequences in a text file represented by characters. Each line of the file contains the gene sequence of one organism. They also have learned that these family of organisms came from one organism whose gene sequence is stored in the first line of a text file. The biologist have learned that the gene sequence of the organism can change in one of the 4 ways every time it mutates.

- 1) two of its genes may swap

e.g: abcdeImnop -> abcedImnop

- 2) one gene may vanish

eg: abcdeImnop -> abcdImnop

- 3) a new gene may get introduced

eg: abcdeImnop -> abcde#Imnop

- 4) one gene may be replaced by new one

eg: abcdeImnop -> abcde*nop

Also they have learned that there will be at most two changes that may occur from one generation to another. For example the following two sequences are not in consecutive generations because there are 3 changes between them.

eg: abdceImnop -> abcde#mno1p

Your job is to sort the genetic sequence of the organisms starting from the oldest to the newest organism. Note that there could be an organism gene sequences in the input file that may not be related to the family of organism the biologists are studying. You must identify these sequences too.

Input

A text file with sequences of characters representing the genetic make-up of the organisms under study. Each line represents one organism, and the first line is the oldest organism at the root of the mutation. With this instruction you are provided with a text file input.txt which has the sample data.

```
abcdelmnop!fghijkquxyrstuvwxyz$elmnop*abcjkquxyrstuvwxyz1ghi  
zbcdelmno!pfgh2jkquyxr3stuvwxyz$emlnopabcj#kquxyrstuvwxyz1hi  
zbcdelmno!pfghijkquxyrstuvwxyz$elmnopabcjkquxyrstuvwxyz1hi  
abcdelmnno!pfghijkquxyrstuvwxyz$elmnopabcjkquxyrstuvwxyz1hi  
zcbdelmno!pkgh2jkquyxr3stuvwxyz$emlnopabcj#kquxyrstuvwxyz1hi  
abcdelmnopfghijkquxyrstuvwxyzdelmnop*abcjkqu5xyrstuvwxyz1ghi  
abcdelmnop!fghijkquxyrstuvwxyz$elmnop*abcjkquxyrstuvwxyz1hi  
zcbdelmno!pfgh2jkquyxr3stuvwxyz$emlnopabcj#kquxyrstuvwxyz1hi  
zbcdelmno!pfgh2jkqu6xr3stu8wd$emlnvpabcj#kquydstuvwxyz1hi  
zbcdelmno!pfghijkquyxrstuvwxyz$elmnopabcj#kquxyrstuvwxyz1hi  
zbcdelmno!pfgh2jkquyxrstuvwxyz$elmnopabcj#kquxyrstuvwxyz1hi
```

Output

You should have two versions of your program with different levels of output. The output file should be named output.txt and it should be located in the same folder as the program.

Version 1: In this mode the program should write the steps it takes to sort the gene sequences in the input file. This version of the program will be used to test your algorithm. In this mode the program should output:

1) When the program compares two gene sequences, it should write

comparing x and y

2) When the program finds two gene sequences that are one after the other it should output

x after y

3) When the program changes the order of a gene sequence, it should output

move x to y

4) When the program finds a gene sequence outside of the family, it should write

x not in family

Where x and y are the line number of the sequence starting from 1.

Version 2: In this mode, the program will be tested for speed. The program should write the sorted gene sequences from the oldest to the newest generation into a text file name output.txt. Any gene sequence that is not related to the family should be placed at the end of the text file below a marker that says "NOT IN FAMILY". See example below.

```
abcdelmnop!fghijkquxyrstuvwxyz$elmnop*abcjkquxyrstuvwxyz1ghi  
abcdelmnop!fghijkquxyrstuvwxyz$elmnop*abcjkquxyrstuvwxyzg1hi  
abcdelmno!pfghijkquxyrstuvwxyz$elmnopabckquxyrstuvwxyz1hi  
zabcdelmno!pfghijkquxyrstuvwxyz$elmnopabckquxyrstuvwxyz1hi  
zbcdelmno!pfghijkquyxrstuvwxyz$elmnopabck#kquxyrstuvwxyz1hi  
zbcdelmno!pfgh2jkquyxrstuvwxyz$elmnopabck#kquxyrstuvwxyz1hi  
zbcdelmno!pfgh2jkquyx3rstuvwxyz$elmnopabck#kquxyrstuvwxyz1hi  
zcbdelmno!pfgh2jkquyx3rstuvwxyz$elmnopabck#kquxyrstuvwxyz1hi  
zcbdelmno!pkgh2jkquyx3rstuvwxyz$elmnopabck#kquxyrstuvwxyz1hi  
NOT IN FAMILY  
zbcdelmno!pfgh2jkqu6xr3stu8wd$emlnvpabcj#kquryrstuvwxyz1hi  
abcdelmnopfghijkqxyrstuvwxyzdelmnop*abcjkq5xyrstuvwxyz1ghi
```

Rules:

- You have to use Java SE language
- You have to submit all your source code files and compiled files for the two output versions
- Start Date: February 27, 2014
- End Date: March 9, 2014 at 9 pm

