

Program Name: next.cpp

Input File: next.dat

There are several applications in which the order of character sequences in the English language is important. For example, when designing the keyboard of a PDA, typewriter, or other text entry machine, the ergonomic movement of fingers can be optimized by properly arranging the keys. In this problem, you will build a program that analyzes some plain text to determine the frequency of combinations of characters in English words.

Your program will read a file of sample English text. For every alphabetic character in the file, your program will determine what alphabetic character (if any) follows. Your program will then count the number of times that each alphabetic character follows all others. The text will contain non-alphabetic characters and all of these including the beginning and end-of-line should be treated as word breaks. You program will not consider combinations that include or span word breaks. After analyzing the entire input file, your program should determine the 5 most frequently occurring alphabetic character paired sequences and print them in order starting with the most frequent. In case of a tie, the 2-character sequence should be listed in alphabetical order.

For example, consider the following line.

This is the sample text for your program. Notice the many "th" sequences.

would yield the following list of 5-most frequently occurring alphabetic paired sequences. Notice that your program must convert all alphabetic characters to lower case to produce the correct answer.

```
1: t h
2: a m
3: c e
4: h e
5: i s
```

Input

Input to your program is a free-form text file that can contain any printable character in the ASCII range 0-127. (Hint: Do not be tempted to make this harder than it is.) The only restriction on the input file is that no single line in the input file will contain more than 80 printable characters.

Output

After reading the input file, your program should print the 5 most frequently occurring pairs of alphabetic characters each on a line by itself. Each line of output should contain: the rank (1-5) in column 1, a colon in column 2, a space in columns 3 and 5, the first character of the sequence at that rank in column 4, and the second character of the sequence at that rank in column 6.

Example: Input File

There are several applications in which the order of character sequences in the English language is important. For example, when designing the keyboard of a PDA, typewriter, or other text entry machine, the ergonomic movement of fingers can be optimized by properly arranging the keys. In this problem, you will build a program that analyzes some plain text to determine the frequency of combinations of characters in English words.

Output to screen

```
1: e r
2: i n
3: t h
4: h e
5: e n
```