1995 ACM MID-CENTRAL REGIONAL PROGRAMMING CONTEST

Problem #6 - Letter Sequence Analysis

Source File:	**sequence.{c|cpp|pas}**
Input File:	**sequence.in**
Output File:	**sequence.out**

Cryptographic analysis makes extensive use of the frequency with which letters and letter sequences occur in a language. If an encrypted text is known to be in english, for example, a great deal can be learned from the fact that the letters E, L, N, R, S, and T are the most common ones used in written english. Even more can be learned if common letter pairs, triplets, etc. are known.

For this problem you are to write a program which accepts as input a text file of unspecified length and performs letter sequence analysis on the text. The program will report the five most frequent letter sequences for each set of sequences from one to five letters. That is it will report the individual characters which occur with the five highest frequencies, the pairs of characters which occur with the five highest frequencies, and so on up to the letter sequences of five characters which occur with the five highest frequencies.

The program should consider contiguous sequences of alphabetic characters only, and case should be ignored (e.g. an 'a' is the same as an 'A'). A report should be produced using the format shown in the example at the end of this problem description. For each sequence length from one to five, the report should list the sequences in descending order of frequency. If there are several sequences with the same frequency then all sequences should be listed in alphabetical order as shown (list all sequences in upper case). Finally, if there are less than five distinct frequencies for a particular sequence length, simply report as many distinct frequency lists as possible.

When a text file containing simply the line "Peter Piper Picks Pickles!" is used as input, the output should appear as shown here:

```
Analysis for Letter Sequences of Length 1
-----------------------------------------
Frequency = 5, Sequence(s) = (P)
Frequency = 4, Sequence(s) = (E)
Frequency = 3, Sequence(s) = (I)
Frequency = 2, Sequence(s) = (C,K,R,S)
Frequency = 1, Sequence(s) = (L,T)

Analysis for Letter Sequences of Length 2
-----------------------------------------
```

```
Frequency = 3, Sequence(s) = (PI)
Frequency = 2, Sequence(s) = (CK,ER,IC,PE)
Frequency = 1, Sequence(s) = (ES,ET,IP,KL,KS,LE,TE)

Analysis for Letter Sequences of Length 3
-----------------------------------------
Frequency = 2, Sequence(s) = (ICK,PIC)
Frequency = 1, Sequence(s) = (CKL,CKS,ETE,IPE,KLE,LES,PER,PET,PIP,TER)

Analysis for Letter Sequences of Length 4
-----------------------------------------
Frequency = 2, Sequence(s) = (PICK)
Frequency = 1, Sequence(s) = (CKLE,ETER,ICKL,ICKS,IPER,KLES,PETE,PIPE)

Analysis for Letter Sequences of Length 5
-----------------------------------------
Frequency = 1, Sequence(s) = (CKLES,ICKLE,PETER,PICKL,PICKS,PIPER)
```

When the first three paragraphs of this problem description are used as input, the output should appear as shown here:

```
Analysis for Letter Sequences of Length 1
-----------------------------------------
Frequency = 201, Sequence(s) = (E)
Frequency = 112, Sequence(s) = (T)
Frequency = 96, Sequence(s) = (S)
Frequency = 90, Sequence(s) = (R)
Frequency = 84, Sequence(s) = (N)

Analysis for Letter Sequences of Length 2
-----------------------------------------
Frequency = 37, Sequence(s) = (TH)
Frequency = 33, Sequence(s) = (EN)
Frequency = 27, Sequence(s) = (HE)
Frequency = 24, Sequence(s) = (RE)
Frequency = 23, Sequence(s) = (NC)

Analysis for Letter Sequences of Length 3
-----------------------------------------
Frequency = 24, Sequence(s) = (THE)
Frequency = 21, Sequence(s) = (ENC,EQU,QUE,UEN)
Frequency = 12, Sequence(s) = (NCE,SEQ,TER)
Frequency = 9, Sequence(s) = (CES,FRE,IVE,LET,REQ,TTE)
Frequency = 8, Sequence(s) = (ETT,FIV)
```

```
Analysis for Letter Sequences of Length 4
-----------------------------------------
Frequency = 21, Sequence(s) = (EQUE,QUEN)
Frequency = 20, Sequence(s) = (UENC)
Frequency = 12, Sequence(s) = (ENCE,SEQU)
Frequency = 9, Sequence(s) = (FREQ,NCES,REQU)
Frequency = 8, Sequence(s) = (ETTE,FIVE,LETT,TTER)

Analysis for Letter Sequences of Length 5
-----------------------------------------
Frequency = 21, Sequence(s) = (EQUEN)
Frequency = 20, Sequence(s) = (QUENC)
Frequency = 12, Sequence(s) = (SEQUE,UENCE)
Frequency = 9, Sequence(s) = (ENCES,FREQU,REQUE)
Frequency = 8, Sequence(s) = (ETTER,LETTE)
```