

TREE PROJECT

There is a real program developed by a computer company that reads a report (running text) and issues warnings on style and partially correct bad style. You are to write a simplified version of this program with the following features:

Statistics

A statistical summary with the following information:

- Total number of words in the report
- Number of unique words
- Number of unique words of more than three letters
- Average word length
- Average sentence length
- An index (alphabetical listing) of all the unique words (see next page for a specific format)

Style Warnings

Issue a warning in the following cases:

- Word used too often: list each unique word of more than three letters if its usage is more than 5% of the total number of words of more than three letters
- Sentence length : write a warning message if the average sentence length is greater than 10
- Word length : write a warning message if the average word length is greater than 5

Input

From the keyboard: The name of the file containing the text to be analyzed

From the file: The report to be analyzed.

Output

1. Write the following information to a file:

- The name of the input file
- The statistical summary of the report (see Statistics above)
- The style warnings (see Style Warnings above)

Data Structures

A BST of unique words in the report, created as the file is read. If a word is not in the list, put it there. If it is, increment a counter showing how many times the word has been used.

Definitions:

Word: Sequence of letters ending in a blank, a period, an exclamation point, a question mark, a colon, a comma, a single quote, or a semicolon. You may assume that numbers do not appear in the words; they may be ignored.

Unique word: Words that are spelled the same, ignoring uppercase and lowercase distinctions.

Sentence: Words between end of markers.

SAMPLE OUTPUT

FILE NAME: chapter.txt

STATISTICAL SUMMARY

TOTAL NUMBER OF WORDS: 987

TOTAL NUMBER OF "UNIQUE" WORDS: 679

TOTAL NUMBER OF "UNIQUE" WORDS OF MORE THAN THREE LETTERS: 354

AVERAGE WORD LENGTH: 8 characters

AVERAGE SENTENCE LENGTH: 12 words

STLE WARNINGS

WORDS USED TOO OFTEN: (WORDS OF MORE THAN 3 LETTERS THAT ARE USED MORE THAN 5% OF THE TOTAL NUMBER OF WORDS OF MORE THAN 3 LETTERS)

- 1) Well
- 2) Total
- 3) Good
- 4) Since
- 5) Because
- 6) Little

AVERAGE SENTENCE LENGTH TOO LONG – 12 words

AVERAGE WORD LENGTH TOO LONG – 8 characters

INDEX OF UNIQUE WORDS

A

and
all
around

...

B

be
because
but

...