G51PRG Exercise Four-A:
Word Lengths

Steven R. Bagley

Introduction

This exercise involves finding the length of words in a plain text file. This will require you to process the file character-by-character, and once it has processed the file it should print out a count of word lengths, i.e. the total number of words of length 1 which occurred, the number of length 2, and so on.

Finding words

For this exercise, you are to develop a C program that can open a text file (using `fopen()` and with the name of the file passed in on the command line) and to find the length of each word in the file. We'll define a word as a sequence of letters and hyphens (you may want to look at using the `isalpha()` function here).

You'll need an array of `int`s to store the count of each word length — you can assume that no word will be longer than 25 characters. Initialise all its elements to zero, and then measure each word's length as it is read in, and increment the appropriate counter. *Remember* though that your array starts at zero…

If a word of length greater than 25 is read, print an error message to the *standard error* output stream, and exit the program.

Once you've successfully processed the entire file, you should output the count of words for each length. However, you should only print the output values up to the longest word encountered, not all the way up to 25. So, for the input:

```
The cat sat on the mat whence all but he had fled.
```

your program should output:

```
length  1 :     0 occurrences
length  2 :     2 occurrences
length  3 :     8 occurrences
length  4 :     1 occurrences
length  5 :     0 occurrences
length  6 :     1 occurrences
```

A selection of text files for you to test your program on can be found at:

`http://g51prg.cs.nott.ac.uk/Distribution/Coursework/cswk4a.zip`