

Project 1

Due date: 1/25

Description

For this project you will read words from a given file into your program. Each word consists strictly of the letters A-Z and a-z. Any whitespace, punctuation, digit, etc. will separate words. An example would be this!is A&Word. This text would represent the 4 words this, is, A, and Word. Each word should be converted to lower case so that only lower case letters remain.

Once you have read all of the words, you should output the total number of words read. Your next task is to determine how many distinct words appear in the file and output that value. Finally, you will prompt the user to enter a word and report how many occurrences of the given word are in the file. You will continue to prompt the user until they enter the EOF character.

When the user enters a word to search for, a '?' character may be present as some of the letters for the search. This special character represents the ability to match any character, including the empty character. For example, colo?r matches both "color" and "colour". This query should report every word that matches.

Keep in mind that the ? character in the source input file is not a wildcard match. In the original source, it is just another non-letter character.

There will be no more than 10,000 distinct words in the source file, and when multiple words match a query, the results should appear in the order that the result words appear in the original text.

Requirements

Please carefully read the following requirements:

- You must supply a makefile. We will discuss makefiles in class (1/25)
- You must use C++ streams for all I/O
- You must format your output as shown in the example below.
- You must do your own work, you must not share code.
- You must submit your project in a zip file as specified under submission

Example

Here is an example document "sample.txt"

Cryptography is both the practice and study of the techniques used to communicate and/or store information or data privately and securely, without being intercepted by third parties. This can include processes such as encryption, hashing, and steganography. Until the modern era, cryptography almost exclusively referred to encryption, but now cryptography is a broad field with applications in many critical areas of our lives.

You must be able to run the program as shown below and get the identical output

```
$ make
g++ -Wall -std=c++11 main.c -o project1
$ ./project1 sample.txt
The number of words found in the file was 64
The number of unique words found in the file was 52
```

```
Please enter a word: of
The word of appears 2 times in the document
```

```
Please enter a word: is
The word is appears 2 times in the document
```

```
Please enter a word: or
The word or appears 2 times in the document
```

```
Please enter a word: a?d
The word and appears 4 times in the document
```

```
Please enter a word: a??
The word and appears 4 times in the document
The word as appears 1 time in the document
The word a appears 1 time in the document
```

```
Please enter a word: ^C
$
```

We highly recommend that you create your own tests to make sure you have covered all possibilities. Those students that do not test their code have a higher chance of having errors during grading.

Submission

Remember that you must submit the project before due date. Late submissions will cause a 10% per day late. No submissions will be accepted after 2 days late. Please read the syllabus for more information.

For this project you must compress the contents of your project directory into a zip file. This means if you unzip this file you would get only the contents of the directory and not the directory itself. The format of the zip file should be LAST-NAME_FIRSTNAME.zip. You can test this on the cs-intro server with the command (change to your name)

```
$ unzip Dixon_Brandon.zip Archive:
Dxion_Brandon.zip
  inflating: main.cpp inflating:
  makefile inflating: sample.txt
$ ls
main.cpp makefile Dixon_Brandon.zip sample.txt
$
```

Once you have created your zip file please submit it to blackboard for grading. Failure to follow the submission instructions will result in a reduction of your grade.