COSC 2320: Data Structures
Homework 0: The most repeated word

Deadline: **Sept. 9, 2014 11:59 PM**

1. INTRODUCTION

This is a warm-up homework that aims at helping you get familiar with the programming environment and tools (how to write code, compile, debug and run the program etc.).

In this homework, you are asked to write a C++ program to find the most repeated word with a length of **at least 4** in a document. You can use any algorithms and data structures that you prefer, as long as the results are correct. It is preferred, but not necessary, that your algorithm is as efficient as possible, both in time to complete as well as memory management. **You are not allowed to use the STL Library.**

2. INPUT

The input is one text file, with at most **150** different words, separated by spaces (repeated spaces should be considered as one space) or carriage returns. You can assume that words will be at most **50** characters long.

You will have to read the input file **FROM DISK**, the file name will be provided as a command-line argument in the program call. (Details about the arguments will be discussed in Section 4.)

Notice that words can be followed by punctuation signs and that some of the letters can be capitalized. This should not interfere with your results: a capitalized word is the same as a non-capitalized word for the purposes of this homework. Uppercase letters can be anywhere in the word, for example: "Hungry", "hungry" and "hUngry" should be counted as the same word. Punctuations should be ignored.

Example of the input file:

```
The hungry, hungry caterpillar
Eric Carle
```

3. OUTPUT

The output should be printed **ON THE SCREEN**. You should print the most repeated word whose length is **no less than 4** in **LOWERCASE** followed by an equal sign and an integer indicating how many times is that word repeated in the input file. We will make sure in the input file that if such word exist, there will be only **ONE** that is most repeated. If the input has no word that has a length of 4 or more, print "**none**" on the screen.

Given the example of the input file above, the output should be:

```
hungry=2
```

4. PROGRAM AND ARGUMENT SPECIFICATION

The main program should be called "**topWord**". The program should be able to take the first argument as input file name.

The call syntax will be like:

```
topWord.exe input.txt
```

Note that the input file name will not necessarily be the same every time, so your program shouldn't have this "input.txt" hard coded.

5. SUBMISSION REQUIREMENTS

Your submission should be well tested before submitting under Visual Studio 2010 or later versions. You can get a copy of Visual Studio from the UH website using your cougarnet username and password. The URL is http://uh.edu/infotech/php/software/list.php.

We use the UH blackboard system to collect your homework submissions. Before you submit your homework, please make sure to **put everything in a ZIP file** named in the form of **LastName_PeopleSoftID_HW0.zip.**

For example: Zhang_1234567_HW0.zip

The instructions about how to use the blackboard system can be found on the TA's webpage for this course: http://www2.cs.uh.edu/~yzhang/cosc2320-f2014/

Please submit in time, all submissions submitted after the due date will accrue a penalty of **33%** per day.

6. GRADING

The maximum grade for this homework is 100.

You will get 15 pts for submitting the homework in time, 10 pts if your program can be successfully compiled.

We will test your program with 5 easy test cases and 5 hard test cases. Each easy test case will worth 10 pts, and each of the hard ones will worth 5 pts.

When testing, we will compare your program's output with the standard output. Therefore **do not print any content on the screen unless required**, avoid any prompt information like "Please enter the input file name:", "The most repeated word is:" etc.

Last but not least, **no cheating or plagiarism will be tolerated in any graded submissions.**

If you have any other questions, please send email to zhangyiqun9164@gmail.com.