CS 3423.02
Assignment 2: Perl language
Due: Midnight next Monday (12:05am of 09/18/2014)

1 Overall Description:

Find out the occurences of each word (A Word Count Example) inside a directory or a file. Those files will be normal text files. You will turnin one perl script, "wordcount.pl" in the end. The first line of this script will be:

```
#!/usr/bin/perl -w
```

This script can be run like this:

```
./wordcount.pl INPUT NUMBER OUTPUT
```

More description about different parameters can be seen as follows.

- The INPUT parameter can be a directory or a file. If it is a directory, we should find out all files inside the given directory and count all words for these files. If it is a file, then we only count the occurrence of words on this file.
- The NUMBER parameter is used to tell the program to output the NUMBER of words with most occurrences.
- We will redirect all of these found words into a OUTPUT file, but only display the NUMBER of words on the screen.

1.1 General Rule

- You need to only submit "wordcount.pl" in the end. Otherwise, you can only get 80% percent even if you do it all right for an assignment. Since we are using the autograding system to grade all program assignments in this semester, not following these rules can greatly increase the amount of workload for me and TA.
- All programs should be able to handle those errors in the first place. Failing to do this will loose some points. For example, if an input has a problem, your program should be able to point out the errors. When you point out errors, you will use "die" keyword. The following is an example for this.

  ```
  die "ERROR: the INPUT is not existing";
  ```

  Since we are going to check your error report using script, then you should use the specific format for this error report. The error report will always start as

  ```
  die "ERROR: YOUR_SPECIFIC_DESCRIPTION";
  ```

- Readablity of a program. We will give some additional points to those programs with good comments and styles, while those ones are not readable can be deducted from some points.

1.2 Suggested Tests:

Make sure that you can verify your scripts using the following test cases:

1. The number of arguments is not correct. We only accept three arguments.
2. INPUT is the same as OUTPUT.
3. OUTPUT is already existing.
4. This given file or directoy can be not existing. Thus, OUTPUT will be empty.
5. INPUT is a file.
6. File name, either INPUT or files in the specified directory, can include space in the middle.
7. Directory may have subdirectories.

1.3 Implementation Tips

- We suggest to use arrays not lists here. The difference between array and list can be seen at http://friedo.com/blog/2013/07/arrays-vs-lists-in-perl.

2 Submission Requirements

In this assignment, you will have to turnin two parts.

1. wordcount.pl: this the program that you are writing. (70%)
2. WordcountDesignForPerl.pdf: more about this file are described below. (30%)

For WordcountDesignForPerl.pdf, you will have to include the following things:

1. **Design Description:** This part describes how you implement your program, such as the steps to resolve this problem. How many functions have you used? Why you think that is good to use a function.
2. **Performance Comparison:** This part will get the performance of your script on a middle size input, provided at http://www.cs.utsa.edu/ tongpingliu/teaching/cs3423/middleinput (which I will upload by Sept. 18). Also, you will try to get the performance of SHELL script on the same input. To get the performance data, you simply run

   ```
   time ./wordcount.pl middleinput 0 OUTPUT
   ```

   Similarly, you run this for the SHELL script. If you can not finish your first assignment successfuly, I will upload a SHELL script for the first assignment at
   http://www.cs.utsa.edu/ tongpingliu/teaching/cs3423/wordcount.sh
   This script only appear after Sept. 21th since we have to wait someone to turnin their assignment.
3. **Advantage of Using Perl:** You simply lists some advantage of using perl based on your observation or comparing these two assignment. The answer can be open-minded.

The writeup don't need to be very formal. It is about one page of word document.