CS 203 Computing and Algorithms III
ASSIGNMENT I
Finding the most repeated word

Requirements and Assumptions (for a max of 95% grade)

In this assignment, you will find the most used word(s) in a given document. That is, given a text file, you should read the file and display the most used word(s). Assume that text file has no punctuations except for a period after each sentence.

In your write up, include the 8 steps of algorithm design. You get to select your own data structures and algorithm for solving the problem. I am looking for a reasonably efficient algorithm that is correct and your analysis of the time complexity is correct.

Note that most likely you will get the word "the" as the most repeated word.

To test your program, you may choose news articles from the Internet that have no audio or video files. Alternatively, if, they do you can ignore the video and audio attachments. NOTE that the step of getting the article and turning it into a text file is not a part of the program. It is done manually.

Additional Requirements (for a max of 100% grade)

Of course what is more interesting would be find which person/topic was on the headline news. For instance, if the name "Bush" appears most number of times in an article, we could conclude that most likely "Bush" was discussed in the article. In this part, therefore you are looking most repeated proper noun or name. If you can accomplish this, then you can get up to 100 points.

The algorithm for this part still needs to be reasonably efficient. As far as correctness goes, you might have more flexibility. You can list the limitations of your algorithm.

Even here, you do not have to automate the conversion of article to a text file, you can do it manually.

Sample Run

The commander in chief's priority is preventing another terrorist attack in the United States, President Bush told CNN's "Larry King Live" Tuesday. "The most important job I have had -- and the most important job the next president is going to have -- is to protect the American people from another attack," Bush said. Accompanied by first lady Laura Bush and engaging in a wide-ranging conversation with King, Bush also said that his administration has been involved in stopping specific threats against the United States since the September 11, 2001, terrorist attacks.

The most used word in the above article is of course "the".

Internal Requirements:

Your program must follow the style guide provided.

You may use any data structure you feel is appropriate, but you must be able to justify it.

You may also call any java method on those data structure provided you know the time complexity of the call.

Assignment Write-up.

You will provide a write up of the assignment consisting of the steps described in Section 1.2 of your textbook.