1. What is the question?

To generate the unique words with their occurrences for macbeth.txt and romeoAndJuliet.txt and compare the first 50 most occurring words in the two output files that are generated.

2. Why is it important?

It is important to compare the 50 most occurring unique words of output file for macbeth.txt and romeoAndJuliet.txt

3. What have others done to try and solve the question?

Unknown.

4. What will I do to solve the problem?

a) Login to fast.cs.odu.edu

b) Secure shell to the Hadoop server.

c) Have created the file name UniqueWords.java. I implemented the functionalities to remove stopwords, punctuation and have converted all words to lowercase to handle condition for case insensitive. Further, I have met all the undergrad level requirements.

c) Run the make.sh bash script to generate the output files for macbeth.txt and romeoAndJuliet.txt

d) Compare the first 50 most occurring words of play macbeth, romeoandjuliet and both the plays.

5. What will I do to prove that I have solved the problem?

I will generate the graph in such a way that it compares the 50 most occurring unique words for macbeth and romeoandjuliet.

6. What is the conclusion?

Generated the output file for macbeth.txt and romeoAndJuliet.txt having unique words with their number of occurrences and was able to compare first 50 most occurring words in macbeth, romeoandjuliet and both the plays in the form of graph.