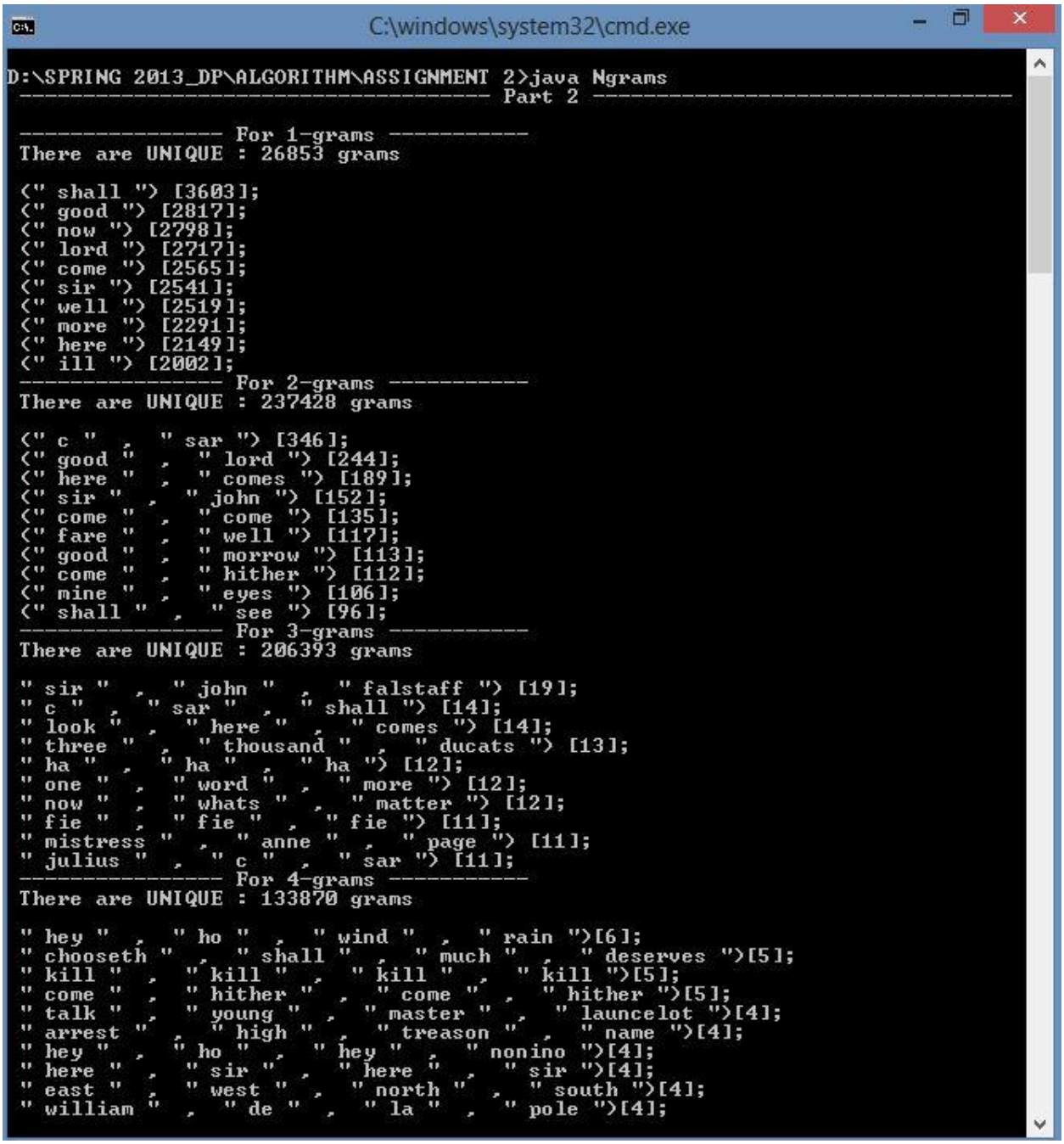
# CSCI – 3412 ALGORITHMS
# SPRING 2013
# PROBLEM SET -2 PROGRAMMING

## PART 1

According to instructions given in Problem Set -2, the program should find the stop words from the file and eliminate them while counting the n-grams. So, for that my program reads the .csv file which contains the stop words and read each line in the loop of counting the n-grams. Basically, while reading the main file (shakespear.txt) for its each and every line, this program will compare with each of the comma separated words from the .csv file each time. If it matches, then it just ignores it and continue for the next one. Otherwise it will store the n-grams.

At the end of the loop, this program has used the Hash Map to store n-grams. Then it uses TreeHashMap for sorted array.

Here is the Result of this program:




## PART 2

In this section, for 2 grams (n=2), it generates all the 2-grams and creates new file (.dot file), and puts all the values to this newly created file in the following manner:

digraph words {

" shall " -> " find " [weight=90];

" mine " -> " honour " [weight=79];

" whats " -> " matter " [weight=76];

" give " -> " leave " [weight=75];

" good " -> " sir " [weight=74];

.

.

.

}

Now, this program only executes the graphviz command and gives this file as an input and generates pdf file which gives the Directed graph for the given 2-grams.

Pdf file is attached with the program.

Let me know in case of any changes.

Thanks,

Darpan Shah.