**HACETTEPE UNIVERSITY**

**COMPUTER ENGINEERING**

**BBM497 – Introduction to Natural Language Programming**

**3rd Assignment**

**Name – Surname : Daghan Emre Aytac**

**Number : 21228051**

## Algorithm:

In task 1, i used our lecture-8 note (BBM495-Lecture8-WSD, page:38) for the algorithm. I used hashtables, arraylists, arrays for performing the task. In task 2, i used a regular expression for parsing corpus and removing some useless parts of words to get a valuable corpus.

## Execution:

I have five argument in that experiment. You can use below codes for compiling and running the code in linux.

For compile all java files: ***javac *.java***

For running the program:

***java Main Generous_train.txt Generous_test.txt generous_vocabulary.txt WordSimilarity.txt output.txt***

References :

**http://www.mkyong.com/java/how-to-count-duplicated-items-in-java-list/**

**https://coderanch.com/t/609803/java/Array-elements-frequency-count-hashmap**

**http://stackoverflow.com/questions/14735171/regex-to-tokenize-string-in-java-with-space-and-double-quotes**

**http://stackoverflow.com/questions/1844688/read-all-files-in-a-folder**

**http://stackoverflow.com/questions/24222730/split-a-string-and-separate-by-punctuation-and-whitespace**

**http://stackoverflow.com/questions/18830813/how-can-i-remove-punctuation-from-input-text-in-java**

**https://commons.apache.org/sandbox/commons-text/jacoco/org.apache.commons.text.similarity/CosineSimilarity.java.html**

**https://github.com/tdebatty/java-string-similarity/blob/master/src/main/java/info/debatty/java/stringsimilarity/Cosine.java**