Readme

The input file for this set are the tweets collected in activity 1, the data is in the form of a csv file in the folder named: Data

hdfs dfs -put ~/Data/ ~/input

hadoop com.sun.tools.javac.Main Stripes\_lemma.java

jar cf Stripes\_lemma.jar Stripes\_lemma\*.class

hadoop jar Stripes\_lemma.jar Stripes\_lemma ~/input/Data ~/op\_stripes

hdfs dfs -get ~/op\_stripes

The output obtained is in the folder op\_stripes

The format of the output is as follows:

Word = (neighbor,freq) = (neighbor, freq)

"1","de = (capoli , 1)= (seminude , 1)

aamazinggameofthrones = (te , 1)= (vicies , 1)= (wellthis , 1)= (just , 1)

aandfan = (webbysht , 1)= (loving , 1)= (years , 1)= (im , 1)