Readme

Hadoop Commands:

hdfs dfs -put ~/Data4\_5/ ~/input

hadoop com.sun.tools.javac.Main Lemma2.java

jar cf Lemma2.jar Lemma2\*.class

hadoop jar Lemma2.jar Lemma2 ~/input/Data ~/op\_lemma

hdfs dfs -get ~/op\_lemma

The input file Data4\_5 has 49 files of latin.

These files were lemmatized using the la.lexicon.csv file.

The outputs are in the following format:

Word/Lemma and the location

"At <apuleius. de\_deo\_soc. 23>

"Christum <augustine. contra\_faustum. 12.31>

"Christus <augustine. contra\_faustum. 12.3>

"Deus <ambrose. epistula\_ad\_sororem. 36>

"Deus, <ambrose. epistula\_ad\_sororem. 23>

"Dives <apuleius. de\_deo\_soc. 23>

"a <augustine. confessions. 8.26>

<augustine. confessions.8.26> <augustine. confessions. 3.21> <augustine. contra\_secundinum. 19.2>