Homewrok2 summary:

I have a main method in the beginning of the java file. Other methods will be called here.

1. *method readFileByLine*(**"BROWN.pos.all"**);

File will be read per tree. A new target file BROWN-clean.pos.txt will be generated after running the program. This new file has the expected format.

1. *print the top 20 tags I used the similar method as in previous homework.*

*Result:*

The top 20 words(word : frequency) are listed below:

nn : 161407

in : 158025

dt : 116447

jj : 76598

the : 69830

nnp : 62023

nns : 55911

to : 52261

rb : 52039

prp : 47304

vbd : 46683

cc : 38094

vb : 36885

of : 36361

vbn : 29436

and : 28822

a : 23157

vbz : 21627

vbg : 17264

prp$ : 16918

the most frequent TAG is NN

it will appear 161407 times!

1. Generated the hashes of hashes from file BROWN-clean.pos.txt.

Method **public static** HashMap GetHash(String FileName)

A hash map will be returned (hashes of hashes) as required. I set several test cases and found it should be right.

1. take the most frequent tag and use it to tag the words in all the sentences from the BROWN-clean.pos.txt file. Report the performance of this tagger.

Method **public static double** Performance(String FileName, String tag)

I compared the tag with the expected tag and counted the number for each and got the result.

Result:

performance is 0.15432974648566058