Homework Submission # 1
TCSS590: Natural Language Processing
Winter 2018
Chhaya Choudhary

Program can be run from command line as:

python <trigram_model.py> <filepath> <word1> <word2> <word3>

where filepath contains file having combined data of positive files and negative files in movie review dataset

word1, word2 and word3 are candidate tokens for the model. The output represents probability of word3 appearing after sequence of word1 and word2.

Example:

Python trigram_model.py dataset/hw1 hello world yahoo

Probability of [yahoo] appearing after [hello] and [world] is 0.00022527596305474206

python trigram_model.py dataset/hw1 ruben santiago hudson

Probability of [hudson] appearing after [ruben] and [santiago] is 0.00041291291291291294

python trigram_model.py dataset/hw1 three ten year

Probability of [year] appearing after [three] and [ten] is 0.00045045045045045046

python trigram_model.py dataset/hw1t vincent price peter

Probability of [peter] appearing after [vincent] and [price] is 0.00033783783783783786

python trigram_model.py dataset/hw1 albert finney tom

Probability of [tom] appearing after [albert] and [finney] is 0.0003153153153153153