Python Tutorial for Natural Language Processing
Homework

Exercise #1 Please save it as ex1_yourinitial.py

From the NLTK pick one of the corpuses by your choice and….
A) Read the Raw text into a list
b) Print length of the text before and after data cleansing
c) import the punctuations from string module and stop words from the NLTK Module, then remove them from the imported text file. Save the cleaned text file as "cleaned.txt*"
d) Stem and Lemmatize your text, by considering the POS tag of each words
f) Print number of the sentences using sentence tokenizer
e) Randomly select five of the sentences from your text and print or draw the NER tags.

Exercise #2 Please save it as ex2_ yourinitial.py

a) Read clean.txt* file and find the top 20 unigrams, and bigrams using the frequency of unigrams and append them into a set.
b) From the List of top 20 unigrams that you collected in a)* create the synonyms and antonyms set for each by using the Wordnet synset. (try to use A Python dictionary to keep the synonyms and antonyms which is a set of key-value pairs)

Exercise #3 Please save it as you ex3_ yourinitial.py

Using the provided regex.txt* file and print the following items using regular expression.
a)Name and last name of people
b)Phone number
c)email address
d)homepages
e)age
f)employer
Each items inside the dictionary (hash table) should contain the information about each person such as age, employer, email address, webpage and…