Assignment 1- NLTK Lab (Due Oct 7 5pm)

Exercise 1:

1) find out the similar word of "freedom" in text4
2) how many times that the word "computer" appeared in text5
3) what's the lexical diversity for text6
4) generate frenquency dictionary for text7 & plot the frequency distribution for the top 20 words
5) list all the words in text 8 that appeared more than 10 times and less than 50 times

Exercise 2:

1) Extract & parsing the text data from "https://www.nyit.edu/news/features/new_york_tech_mini_research_grants_program_expands_to_focus_on_girls_in_stem"
2) Apply word tokenization to the file
3) Use NLTK stop-word list to remove the stop-words
4) Normalize all tokens into lowercase & apply wordnet Lemmatization to the tokens
5) Generate vocabulary (distinct words after stop-word removal, normalization & lemmatization), and write the vocabulary to the text file "vocab.txt"