**CA07**

Due 3/24 11:59 pm

On the 2015 US speech in the UN-Debate dataset, generate syntagmatically related lemmatized nouns and verbs using mutual information on sentences as segments. The following snippet segments sentences, where speech is a string of the text for the speech:

import spacy  
  
nlp = spacy.load("en\_core\_web\_sm")  
doc = nlp(speech) ### tokenization  
\_text = []  
for sent in doc.sents: ### segment sentences  
    \_text.append(sent)  
  
Then remove tokens that are not nouns or verbs from each sentence and lemmatize them. After this, perform the following:

1. For each pair of different words appearing in a segment, compute its mutual information over all segments and store the value and the pair.
2. Store the word pairs with their mutual information values greater than or equal to the median value of all mutual information values computed above (this will be used for searching below). These pairs of words are considered syntagmatic.
3. Enter a word to retrieve all words in alphabetical order that are syntagmatic to the given word. Use keyword = input("Enter a keyword: ")to capture user input to vaviable keyword. Some running samples are given below:

Enter a keyword: development

Syntagmatic words found:

agreement|emergence|enable|encompass|environment|growth|liberty|market|meet|one|open|promote|protect|standard|  
sustain|trade|will|worker

Enter a keyword: environment   
Syntagmatic words found:   
agreement|escalate|growth|kill|market|meet|open|promote|protect|protest|react|repression|standard|strife|sustain|  
trade|turn|worker  
  
Enter a keyword: isolate

Syntagmatic words found:

leverage|neighbour|promise|region|strengthen|system|trade|want|whole

Enter a keyword: endanger

Syntagmatic words found:

fuel|isolate|leverage|neighbour|office|promise|regime|region|trade|violation