THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Natural Language Proceesing
DATS 6312
Due: *Oct-02-2023*

Homework 4:

- **Show ALL Work, Neatly and in Order.**
- **No** credit for Answers Without Work.
- For written questions, submit a single pdf file includes all of your solutions.
- **DO NOT** submit individual files or images.
- For coding questions, submit **ONE** .py file and include your comments.

E.1:

In part of this exercise, you will use Spacy software to explore the unrealistic news dataset called data.txt.

i. Use pandas to load the data.csv data.

ii. Use Spacy to find the word level attributions (Tokenized word, StartIndex, Lemma, punctuation,white space ,WordShape, PartOfSpeech, POSTag). Use one of the titles in the dataframe and create a dataframe which rows are the word and the columns are attributions.

iii. Use spacy and find entities on the text in part ii.

iv. Grab a different title and use spacy to chunk the noun phrases, label them and finally find the roots of each chunk.

v. Use SPacy to analyzes the grammatical structure of a sentence, establishing relationships between ”head” words and words which modify those heads. Hint: Insatiate the nlp doc and then look for text, dependency, head text, head pos and children of it.

vi. Use spacy to find word similarly measure. Spacy has word vector model as well. So we can use the same to find similar words. Use spacy large model to get a decent results.

E.2:

In part of this exercise, you will use Spacy software to explore the tweets dataset called data1.txt.

i. Use pandas to load the data1.csv data.

ii. Let’s look at some examples of real world sentences. Grab a tweet and explain the text entities.

iii. One simple use case for NER is redact names. This is important and quite useful. Find a tweet which has a name in it and then redact it by word [REDACTED].

E.3:

Use spacy to answer all the following questions.

i. Apply part of speech Tags methods in spacy on a sentence.

ii. Apply syntactic dependencies methods in spacy on same sentence that you used on part i.

iii. Apply named entities methods in spacy on the following sentence "*Apple is looking at buying U.K. startup for 1 billion dollar*".

iv. Apply document similarity on two separate document (2 sentences).

E.4:

Answer all the class exercise questions and submit it (Check the instructions).