SchoolNova Computer Science 202
Homework 25
Due 4/24/2021 on Google Classroom

Task 1

Pick any two books from the https://www.gutenberg.org/browse/scores/top. Download the text versions of the book and delete all the extra text which is not part of the original book. Load both books in Python.

Task 2

Replicate the class work tasks: Split the text into sentences. Calculate the VADER sentiment scores given the number of parts determined by the user (e.g., parts = 5).

Task 3

Compare the sentiment scores in the two books. Plot the sentiment scores as two lines using Matplotlib. Make sure to add labels/legend so that the user can distinguish between the books.

Task 4*

NRClex is similar to the VADER sentiment analysis but it provides more details about emotional characteristics of the words in a text. Import using:

from nrclex import NRCLex

To generate the NRC object use:

nrc = NRCLex(sometext)

To see the results use:

print(nrc.raw_emotion_scores)

Can you replicate Task 3 using NRClex? Notice that you don't have to divide the text by the number of sentences. You can simply divide it by the number of characters, which simplifies things. Also, notice that instead of a single score you will have results for 10 different emotions. Feel free to choose what you find most interesting.