Lab 1: NLP
Spring 2023 – Jan 23

Welcome to Natural Language Processing Lab!
The goal of this lab is to remind you of Python! There are a set of exercises that you will do by searching the Internet.
https://www.python.org/

First set up Google Colaboratory (Colab).

- An executable document allowing you to write, run, and share code on
  - Google Drive (Jupyter notebook stored in Google Drive)
  - GitHub
- A notebook document is composed of cells: Code, text, image, or …
- Colab connects your notebook to a cloud-based runtime
  - no required setup on your own machine
- Able to use Code snippets and insert them in your own code

Practice 1: 40%

1. Print odd numbers in range 0 to 100
2. Write a function that returns summation of even numbers in a range
   (start, end) and both start and end are included in the summation
3. Define an empty list
   a. Add 10 random integer numbers in range [1,100] to your list
   b. Print the maximum number in the list
   c. Print the minimum number in the list
   d. Sort the list
   e. Randomly shuffle the list
   f. Try running the sort part of code with “Run the focused cell”

4. Define a dictionary as: {‘a’: 1, ‘d’:4, ‘b’:2, ‘c’:3}
    a. Sort the dictionary by keys
    b. Sort the dictionary by values in descending order
    c. Add new tuple to the dictionary {‘e’: 5}

Practice 2: 40%
Go to TF-IDF Wikipedia page and copy the first 3-Sentences (above motivation)

Test turning the debug mode on and off with command: !pdb on and !pdb off

a. Define a string with this paragraph as the value
b. Write a program that prints the number of total words and the number of total unique words in this paragraph
c. Which word has the most frequency? (write a program for this)
d. Read the TFIDF.txt file as the string and test if you will get the same results
e. Use NLTK library and remove the stop-words. What is the most frequent word now?

Practice 3: 20%
Target file: quotes.tsv (tab-separated values)

Read the TSV file with pandas library. Install the library with command: ! pip install pandas

Answer the following questions

a. How many quotes are from ‘Alexandre Dumas’?
b. Who has the longest quote? (number of words)
c. In whose quote there is the word ‘one’? Name the author(s)
d. What are the most and least frequent words used in all the quotes?

Submit your solution by copying and pasting the code into a text file. Make sure your code is readable; having both indentation and comments.