CSC790: Information Retrieval and Web Search
Fall 2021, Assignment #2
Date Assigned: Monday, September 27, 2021
Due Date: Wednesday, October 06, 2021 at 11:59 pm (on blackboard)
(120 points)

Objectives:

- Build biword and positional indexes from a list of documents.

- Retrieve documents related to user phrase queries.

Questions

In this assignment you will use the same dataset from homework 1.

Write the python code that:

1. Build biword index (do not remove stop words).

2. Build positional index (do not remove stop words).

3. Save the index in a file/load the saved index file.

4. Read the user queries from a text(.txt) file (each query is written in one line). This query consists of at least two words. Your code will display the following:

   - The size of the biword index.
   - The size of the positional index.
   - The result (retrieved documents) using the biword index (the output should be sorted) for each query.
   - The result (retrieved documents) using the positional index (the output should be sorted) for each query.

Submission

1. Write your own code. Use as many functions as you can.

2. Make sure you writing you name and assignment number on all files you submit.

3. Your python code and the instructions on how to run it.

4. Enclose all your files in a folder named **HW02_yourlastname.zip**.

5. Submit the zip file using blackboard.