CSC790: Information Retrieval
Spring 2024, Programming Assignment #4
Date Assigned: Monday, April 1, 2024
Due Date: Friday, April 12, 2024 at 11:59 pm (on Brightspace)
(150 + 30 points)

Objectives:

- Implementing probabilistic retrieval model.

Tasks

Download the folder HW04 from Brightspace. The folder contains the following:

1. **documents:** contains the text files.

2. **file_label:** contains file relevance labeling, it has the following format: [file_name, rel] where $rel \in \{0, 1\}$

3. **query:** a file contains the query.

Write a python code that implements the binary independence model (slides 26-38). Your code should display the top 10 documents with high RSV. Your output should look like.

```
======================= CSC790–IR  Homework  04 =============
First  Name:  your  first  name
Last  Name  :  your  last  name
==============================================================

RSV{file_name1} = 22.9    file_label
RSV{file_name2} = 14      file_label
RSV{file_name3} = 12      file_label
RSV{file_name4} = 10.9    file_label
RSV{file_name5} = 9       file_label
RSV{file_name6} = 4       file_label
```

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 (in .py format) and the instructions on how to run it.

4. Enclose all your files in a folder named **HW04_yourlastname.zip**.

5. Submit the zip file using Brightspace.