CSC790: Information Retrieval and Web Search
Fall 2021, Programming Assignment #4
Date Assigned: Monday, November 17, 2021
Due Date: Friday, December 03, 2021 at 11:59 pm (on blackboard)
(120 points)

Objectives:

- Implementing probabilistic retrieval model.

Tasks

Download the folder HW04 from Blackboard. 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 28,29,30). Your code should display the top 10 documents with high RSV. Your output should look like.

```
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. Copied code(from internet or from a classmate) will receive a zero grade and will be reported.
3. Make sure you writing you name and assignment number on all files you submit (you will lose 20 points if you do not)
4. Your python code(in .py format) and the instructions on how to run it.
5. Enclose all your files in a folder named **HW04_yourlastname.zip**.
6. Submit the zip file using blackboard.