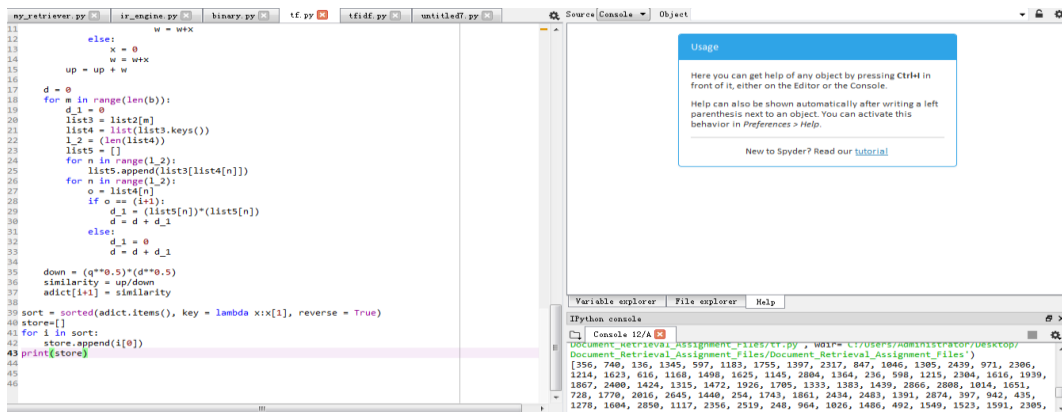


Assignment Report

Xin Sun

1. implementation achieved

I have finished the programming of the three term weighting parts. I have calculated the $|D|$, $\log(|D|/df_w)$, $|d|$ and $\cos(q, d)$. I have calculated the similarity for an example query and sorted the documents according to the similarities in the spyder script.



```
12 else:
13     x = 0
14     w = w+x
15     up = up + w
16
17 d = 0
18 for n in range(len(b)):
19     d_1 = 0
20     list3 = list2[n]
21     list4 = list(list3.keys())
22     l_2 = (len(list4))
23     list5 = []
24     for n in range(1,2):
25         list5.append(list3[list4[n]])
26     for n in range(1,2):
27         o = list4[n]
28         if o == (i+1):
29             d_1 = (list5[n])*(list5[n])
30             d = d + d_1
31         else:
32             d_1 = 0
33             d = d + d_1
34
35 down = (q**0.5)*(d**0.5)
36 similarity = up/down
37 adict[i+1] = similarity
38
39 sort = sorted(adict.items(), key = lambda x:x[1], reverse = True)
40 store=[]
41 for i in sort:
42     store.append(i[0])
43 print(store)
```

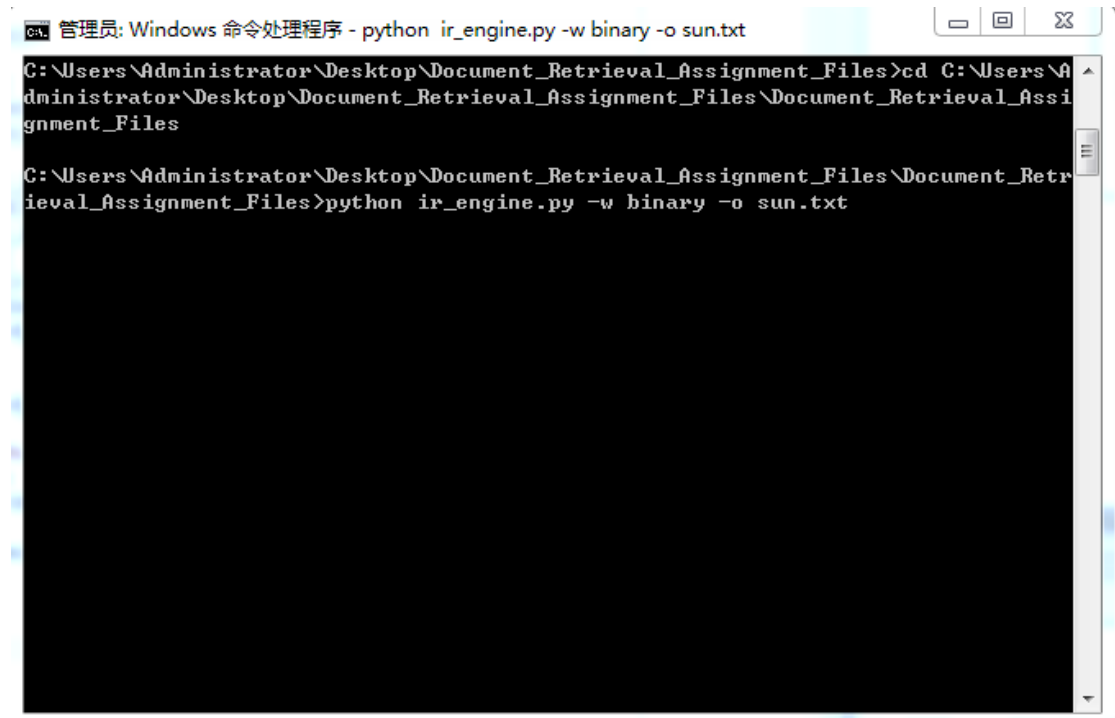
2. Code structure

I looped through the $|D| = 3204$ document ID. In each cycle, the code could be in three parts, calculating the $\sum q * d$, calculating the $\sum_{i=0}^n d^2$, and calculated the $\cos(q,d)$. Then I use sorted function to sort the documents because the same class.

3. Problem

The reason why I cannot complete the sufficient implementation is the complex code. I have tried my best to make an optimization, but I have no idea how to make code easier to run, especially the part of calculating the $|d|$. When the code is running, this part spends lots of

time and when I run it in CMD, I cannot get the result.



```
ca. 管理员: Windows 命令处理程序 - python ir_engine.py -w binary -o sun.txt
C:\Users\Administrator\Desktop\Document_Retrieval_Assignment_Files>cd C:\Users\Administrator\Desktop\Document_Retrieval_Assignment_Files\Document_Retrieval_Assignment_Files
C:\Users\Administrator\Desktop\Document_Retrieval_Assignment_Files\Document_Retrieval_Assignment_Files>python ir_engine.py -w binary -o sun.txt
```

There is no error, but lasting in this window.

I think I should try to make my code clearer and simpler because although there is no error with the code, it hard to read and spend lots of time for running.