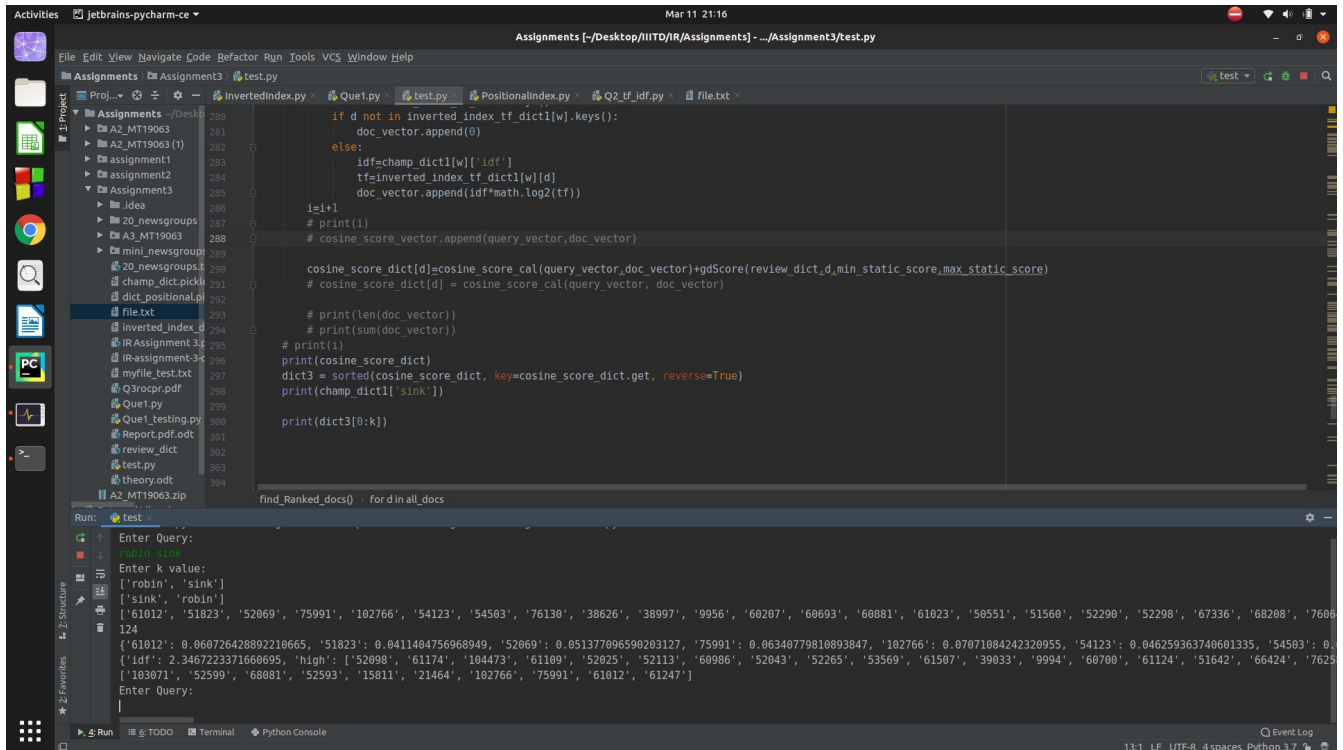


Analysis

Q1.

Output:



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Assignments3 test.py
Project: Assignments3
Assignments3
  A2_MTI9063
  A2_MTI9063 (1)
  assignment1
  assignment2
  Assignment3
  .idea
  20_newsgroups
  A3_MTI9063
  mini_newsgroups
  20_newsgroups
  champ_dict.pickle
  dict_positional.py
  file.txt
  InvertedIndex.py
  IRAssignment3
  IRAssignment3-1
  myfile_test.txt
  Q3rocrpr.pdf
  Que1.py
  Que1_testing.py
  Report.pdf.odt
  review_dict
  test.py
  theory.odt
  A2_MTI9063.zip
  External Libraries
  Scratches and Console

280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304

if d not in inverted_index_tf_dict[w].keys():
    doc_vector.append(0)
else:
    idf=champ_dict[w]['idf']
    tf=inverted_index_tf_dict[w][d]
    doc_vector.append(idf*math.log2(tf))

i=i+1
# print(i)
# cosine_score_vector.append(query_vector,doc_vector)

cosine_score_dict[d]=cosine_score_cal(query_vector,doc_vector)+gdScore(review_dict,d,min_static_score,max_static_score)
# cosine_score_dict[d] = cosine_score_cal(query_vector, doc_vector)

# print(len(doc_vector))
# print(sum(doc_vector))
# print(i)
print(cosine_score_dict)
dict3 = sorted(cosine_score_dict, key=cosine_score_dict.get, reverse=True)
print(champ_dict['sink'])

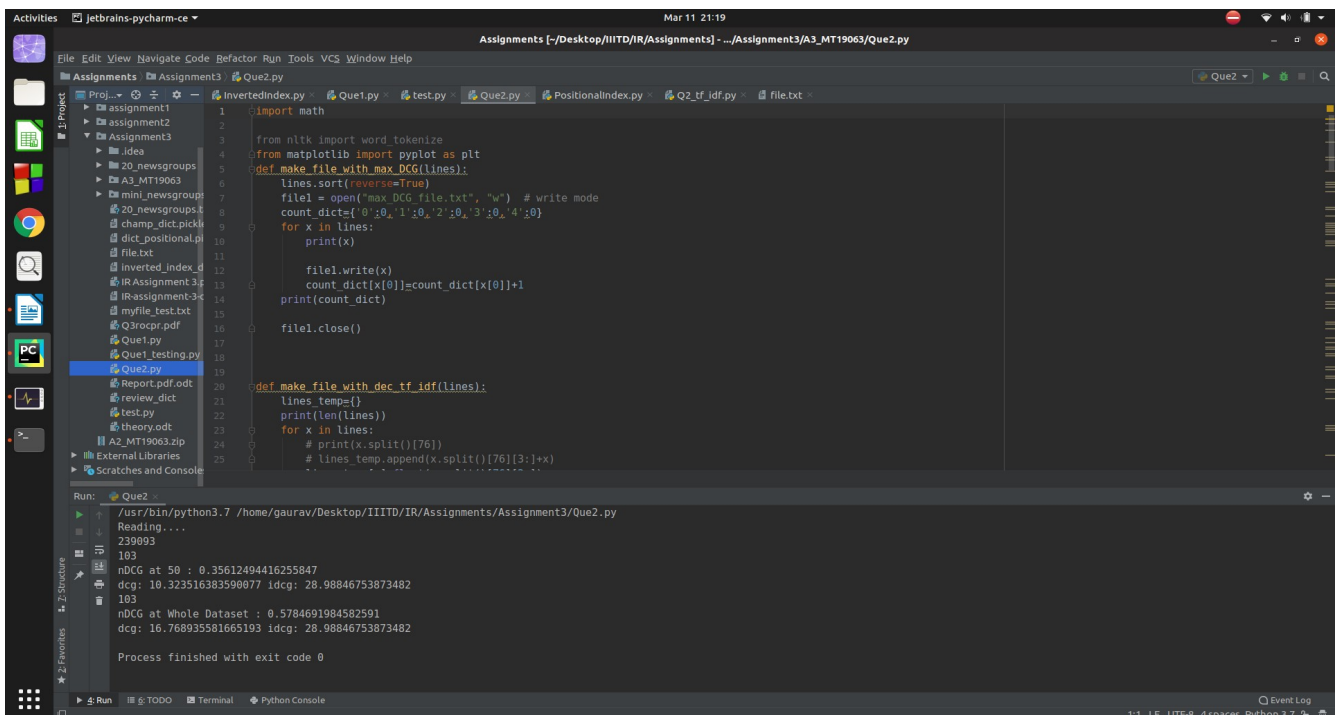
print(dict3[0:k])

Find_Ranked_docs() for d in all_docs

Run: test
Enter Query:
Enter k value:
['robin', 'sink']
['sink', 'robin']
{'161012': '51823', '52069', '75991', '102766', '54123', '54503', '76130', '38626', '38997', '9956', '60207', '60693', '60881', '61023', '50551', '51560', '52290', '52298', '67336', '60208', '7606
124
{'161012': 0.060726428892210665, '51823': 0.0411404756968949, '52069': 0.051377896590203127, '75991': 0.06340779810893847, '102766': 0.07071084242320955, '54123': 0.046259363740601335, '54503': 0.
{'idf': 2.3467223371660695, 'high': ['52098', '61174', '104473', '61109', '52025', '52113', '60986', '52043', '52265', '53569', '61587', '39033', '9994', '60700', '61124', '51642', '66424', '7625
['103071', '52599', '68081', '52593', '15811', '21464', '102766', '75991', '61012', '61247']
Enter Query:

13:1 UTF-8 4 spaces Python 3.7
```

Q2. Output:



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Assignments3 Que2.py
Project: Assignments3
Assignments3
  assignment1
  assignment2
  Assignment3
  .idea
  20_newsgroups
  A3_MTI9063
  mini_newsgroups
  20_newsgroups
  champ_dict.pickle
  dict_positional.py
  file.txt
  InvertedIndex.py
  IRAssignment3
  IRAssignment3-1
  myfile_test.txt
  Q3rocrpr.pdf
  Que1.py
  Que1_testing.py
  Que2.py
  Report.pdf.odt
  review_dict
  test.py
  theory.odt
  A2_MTI9063.zip
  External Libraries
  Scratches and Console

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

import math
from nltk import word_tokenize
from matplotlib import pyplot as plt
def make_file_with_max_DCG(lines):
    lines.sort(reverse=True)
    file1 = open("max_DCG_file.txt", "w") # write mode
    count_dict={'0':0,'1':0,'2':0,'3':0,'4':0}
    for x in lines:
        print(x)
        file1.write(x)
        count_dict[x[0]]=count_dict[x[0]]+1
    print(count_dict)
    file1.close()

def make_file_with_dec_tf_idf(lines):
    lines_temp=[]
    print(len(lines))
    for x in lines:
        # print(x.split()[76])
        # lines_temp.append(x.split()[76][3:]+x)

Run: Que2
/usr/bin/python3.7 /home/gaurav/Desktop/IIITD/IR/Assignments/Assignment3/Que2.py
Reading....
239093
103
ndcg at 50 : 0.35612494416255847
dcg: 10.323516383590077 idcg: 28.98846753873482
103
ndcg at Whole Dataset : 0.5784691984582591
dcg: 16.768935581665193 idcg: 28.98846753873482
Process finished with exit code 0

1:1 UTF-8 4 spaces Python 3.7
```