

## Analysis:

Q1:

output:

On homogenous queries:

The screenshot shows the PyCharm IDE with the file `InvertedIndex.py` open. The code defines a `dict1_all_doc` dictionary and a `run_query` function. The output window displays the results for a homogenous query: `A AND B AND C`. The output shows a list of document IDs (DOCs) and their count.

```
dict1_all_doc=loadData()
# print(len(all_doc))
dict1={}
all_doc=[]

query=input()
while(query!="-1"):
    query = input()
    run_query(query,dict1_all_doc)
# print(lists)
# print(lists)
```

Output:

```
19991
19992
19993
19994
19995
19996
19997
A AND B AND C
DOCs: [188, 9542, 9543, 9544, 9704, 9830, 9977, 9978, 9979, 9980, 9981, 9982, 9983, 9984, 9985, 10140, 10156, 10170, 10698, 15248, 15545, 38248, 38339, 38633, 38898, 38957, 51189, 51198, 51206,
Count : 4510
```

One Hetrogenious Queries:

The screenshot shows the PyCharm IDE with the file `InvertedIndex.py` open. The code is the same as in the previous screenshot. The output window displays the results for a heterogeneous query: `A OR B OR C OR D`. The output shows a list of document IDs (DOCs) and their count.

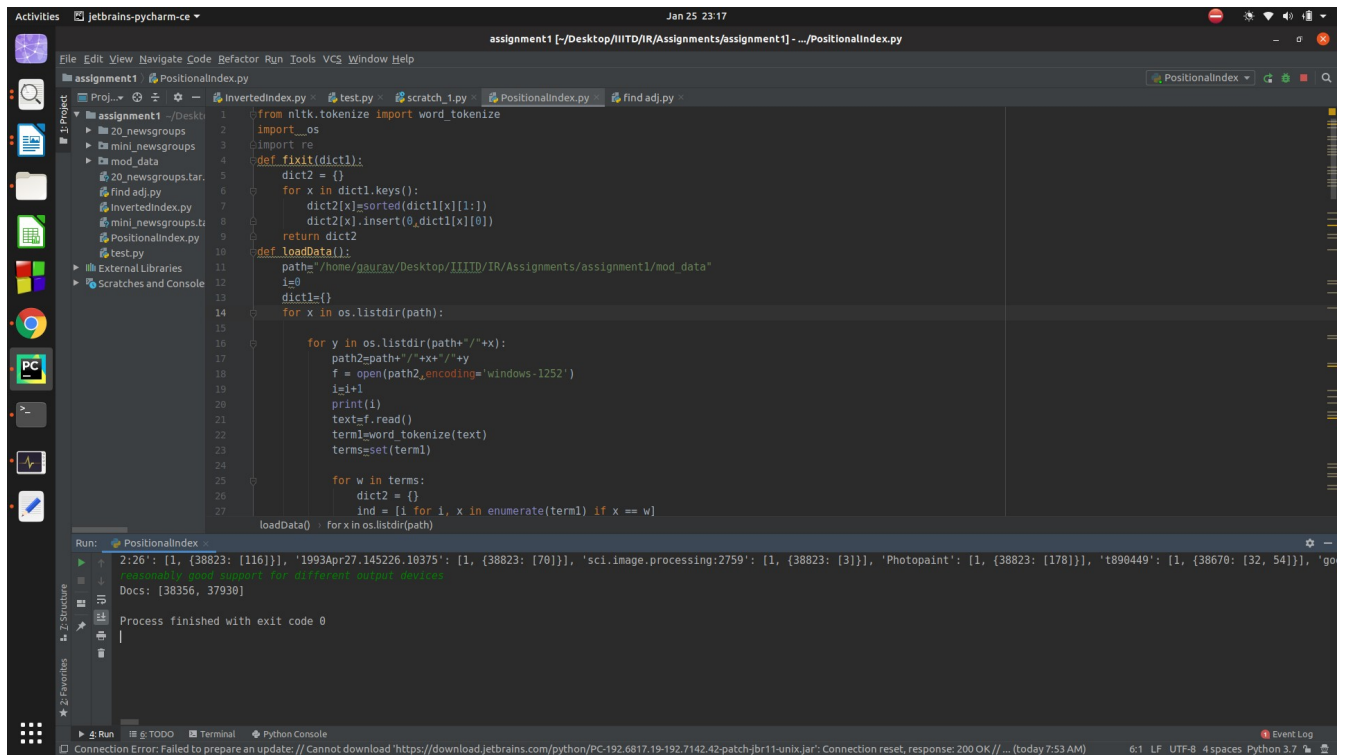
```
dict1_all_doc=loadData()
# print(len(all_doc))
dict1={}
all_doc=[]

query=input()
while(query!="-1"):
    query = input()
    run_query(query,dict1_all_doc)
# print(lists)
# print(lists)
```

Output:

```
DOCs: [433, 9534, 9542, 9543, 9544, 9619, 9626, 9627, 9704, 9755, 9774, 9801, 9816, 9830, 9976, 9977, 9978, 9979, 9980, 9981, 9982, 9983, 9984, 9985, 9986, 9987, 9988, 9990, 10052, 10140, 10156,
Count : 4940
A OR B OR C OR D
DOCs: [4226, 9137, 9143, 9144, 9151, 9152, 9156, 9166, 9168, 9170, 9174, 9450, 9477, 9482, 9493, 9503, 9517, 9523, 9529, 9533, 9534, 9542, 9543, 9544, 9553, 9555, 9569, 9574, 9594, 9601, 9603,
Count : 11971
A OR B OR C AND D
DOCs: [3742, 9137, 9143, 9144, 9151, 9152, 9156, 9166, 9174, 9450, 9477, 9482, 9493, 9517, 9523, 9533, 9542, 9543, 9544, 9553, 9555, 9569, 9594, 9601, 9607, 9609, 9622, 9624, 9640, 9644, 9654,
Count : 8166
```

## Q2:Positional Index:



The screenshot displays a JetBrains IDE window titled "assignment1 [-/Desktop/IIITD/IR/Assignments/assignment1] - .../PositionalIndex.py". The editor shows the following Python code:

```
1 from nltk.tokenize import word_tokenize
2 import os
3 import re
4 def fixit(dict1):
5     dict2 = {}
6     for x in dict1.keys():
7         dict2[x]=sorted(dict1[x][1:])
8         dict2[x].insert(0,dict1[x][0])
9     return dict2
10 def loadData():
11     path="/home/gaurav/Desktop/IIITD/IR/Assignments/assignment1/mod_data"
12     i=0
13     dict1={}
14     for x in os.listdir(path):
15
16         for y in os.listdir(path+"/"+x):
17             path2=path+"/"+x+"/"+y
18             f = open(path2,encoding='windows-1252')
19             i=i+1
20             print(i)
21             text=f.read()
22             term1=word_tokenize(text)
23             terms=set(term1)
24
25             for w in terms:
26                 dict2 = {}
27                 ind = [i for i, x in enumerate(term1) if x == w]
```

The Run window shows the output of the script:

```
Run: PositionalIndex
2:26: [1, {38823: [116]}], '1993Apr27.145226.10375': [1, {38823: [70]}], 'sci.image.processing:2759': [1, {38823: [3]}], 'Photopaint': [1, {38823: [178]}], 't890449': [1, {38670: [32, 54]}], 'go
reasonably good support for different output devices
Docs: [38356, 37930]
```

Below the output, it states "Process finished with exit code 0".