I have assumed the query to be a document, so whatever the corpus size is, query will be an additional +1 doc. I have tested my code on three collections.

The query is taken as input and then stored in the doc.

By default the code is set to run on the stories provided with the assignment on slate, but to check it on 2 other collections namely col1 and col2, 2 slight changes have to be made in the code.

1)Uncomment the collection and corpus size which you want to use in **main()**.

```
313
314              //uncomment the collection which you wanna use for checking the code
315              // 1)col1
316      //        FileWriter fw=new FileWriter("col1\\4.txt");
317      //        corpusSize=4;
318
319              // 2)col2
320      //        FileWriter fw=new FileWriter("col2\\4.txt");
321      //        corpusSize=4;
322
323              // 3)col3
         FileWriter fw=new FileWriter("col3\\51.txt");
325          corpusSize=51;
326
```

2)The collections are to be uncommented in **Indexing()**

```
169  //          String data=readFileAsString("col1\\"+(i+1)+".txt");
170
171          /*My Collection col2
172          d1 w1 w2 w4 w6
173          d2 w1 w2 w7 w3
174          d3 w8 w5 w4 w5 w6
175          q   w2 w5 w6
176      */
177  //          String data=readFileAsString("col2\\"+(i+1)+".txt");
178
179          //Sir Collection
180              String data=readFileAsString("col3\\"+(i+1)+".txt");
```

By doing so, code can be tested on 3 different collections.

The results obtained from the collections are:

Col1:

D1: dil dil Pakistan, jan jan Pakistan

D2: Pakistan hum sub ki jan

D3: dil aur jan Pakistan Pakistan
query "dil jan Pakistan".

	df	d1	d2	d3	q	idf	tf-idfd1	tf-idfd2	tf-idfd3	tf-idfq
aur:	1.0,	0.0,	0.0,	1.0,	0.0,	0.48,	0.0,	0.0,	0.48,	0.0,
dil:	2.0,	2.0,	0.0,	1.0,	1.0,	0.18,	0.36,	0.0,	0.18,	0.18,
hum:	1.0,	0.0,	1.0,	0.0,	0.0,	0.48,	0.0,	0.48,	0.0,	0.0,
jan:	3.0,	2.0,	1.0,	1.0,	1.0,	0.0,	0.0,	0.0,	0.0,	0.0,
ki:	1.0,	0.0,	1.0,	0.0,	0.0,	0.48,	0.0,	0.48,	0.0,	0.0,
pakistan:	3.0,	2.0,	1.0,	2.0,	1.0,	0.0,	0.0,	0.0,	0.0,	0.0,
sub:	1.0,	0.0,	1.0,	0.0,	0.0,	0.48,	0.0,	0.48,	0.0,	0.0,

***********Ranking Of Documents:***********

D-1 Sim: 1.0

D-3 Sim: 0.33

D-2 Sim: 0.0

Col2:

```
d1: w1 w2 w4 w6

d2: w1 w2 w7 w3

d3: w8 w5 w4 w5 w6

q: w2 w5 w6
```

	df	d1	d2	d3	q	idf	tf-idfd1	tf-idfd2	tf-idfd3	tf-idfq
w1:	2.0,	1.0,	1.0,	0.0,	0.0,	0.18,	0.18,	0.18,	0.0,	0.0,
w2:	2.0,	1.0,	1.0,	0.0,	1.0,	0.18,	0.18,	0.18,	0.0,	0.18,
w3:	1.0,	0.0,	1.0,	0.0,	0.0,	0.48,	0.0,	0.48,	0.0,	0.0,
w4:	2.0,	1.0,	0.0,	1.0,	0.0,	0.18,	0.18,	0.0,	0.18,	0.0,
w5:	1.0,	0.0,	0.0,	2.0,	1.0,	0.48,	0.0,	0.0,	0.96,	0.48,
w6:	2.0,	1.0,	0.0,	1.0,	1.0,	0.18,	0.18,	0.0,	0.18,	0.18,
w7:	1.0,	0.0,	1.0,	0.0,	0.0,	0.48,	0.0,	0.48,	0.0,	0.0,
w8:	1.0,	0.0,	0.0,	1.0,	0.0,	0.48,	0.0,	0.0,	0.48,	0.0,

```
***********Ranking Of Documents:***********

 D-3 Sim: 0.83

 D-1 Sim: 0.32

 D-2 Sim: 0.08
```

The results obtained from the stories (**Col3**) provided with the assignment are:

Col3:

<u>Crowd busy</u>

```
***********Ranking Of Documents:***********

D-23 Sim: 0.02747

D-7 Sim: 0.0168

D-17 Sim: 0.0144

D-37 Sim: 0.01391

D-12 Sim: 0.01261

D-46 Sim: 0.01243

D-3 Sim: 0.01129

D-27 Sim: 0.01041

D-39 Sim: 0.0096

D-44 Sim: 0.00948

D-6 Sim: 0.00901

D-20 Sim: 0.00616
```

<u>Dog cows</u>

```
***********Ranking Of Documents:***********

D-2 Sim: 0.31992

D-14 Sim: 0.09064

D-32 Sim: 0.02502

D-18 Sim: 0.02338

D-24 Sim: 0.02232

D-19 Sim: 0.0173

D-21 Sim: 0.01558

D-37 Sim: 0.0111
```

King queen

```
***********Ranking Of Documents:***********

  D-31 Sim: 0.18967

  D-7 Sim: 0.02099

  D-34 Sim: 0.01976

  D-50 Sim: 0.01561

  D-43 Sim: 0.0114

  D-49 Sim: 0.00959

  D-40 Sim: 0.0081

  D-48 Sim: 0.00743

  D-25 Sim: 0.0061
```

Love hate

```
***********Ranking Of Documents:***********

  D-43 Sim: 0.03562

  D-6 Sim: 0.02386

  D-1 Sim: 0.01016

  D-9 Sim: 0.00836

  D-26 Sim: 0.00672

  D-49 Sim: 0.00663

  D-45 Sim: 0.0057
```

Travel water

```
***********Ranking Of Documents:***********

 D-45 Sim: 0.04174

 D-11 Sim: 0.03439

 D-31 Sim: 0.03233

 D-21 Sim: 0.02405

 D-13 Sim: 0.01808

 D-19 Sim: 0.01551

 D-33 Sim: 0.00782

 D-28 Sim: 0.00694

 D-48 Sim: 0.0052
```