



**CSCI 5408 – Data Management, Warehousing, Analytics**

**Assignment 5**

**Work done by,**

**Name: Guturu Rama Mohan Vishnu**

**Banner ID: B00871849**

**Email: [rm286720@dal.ca](mailto:rm286720@dal.ca)**

## **DECLARATION**

**I, Guturu Rama Mohan Vishnu, declare that in assignment 3 of CSCI 5408 course, I declare that all the work done was done by myself and I have not collaborated with anyone for the assignment.**

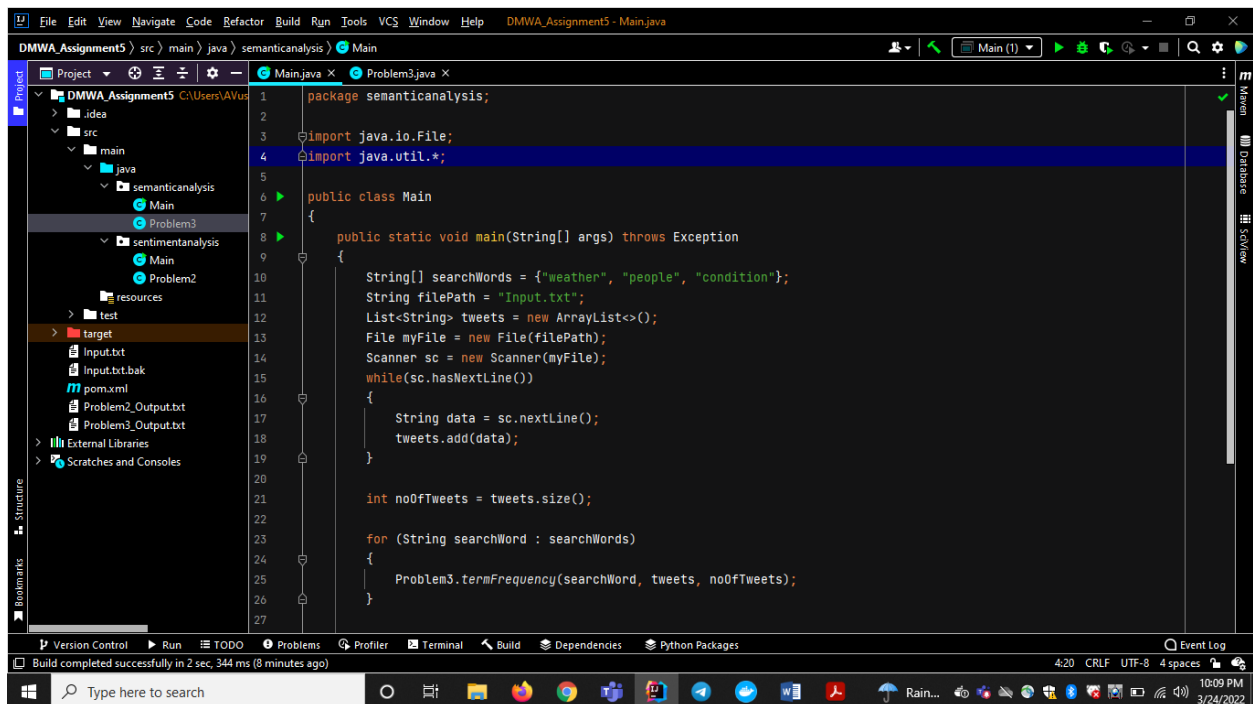
## Problem #3

### Semantic Analysis

The code written for this program has been pushed to git repo :

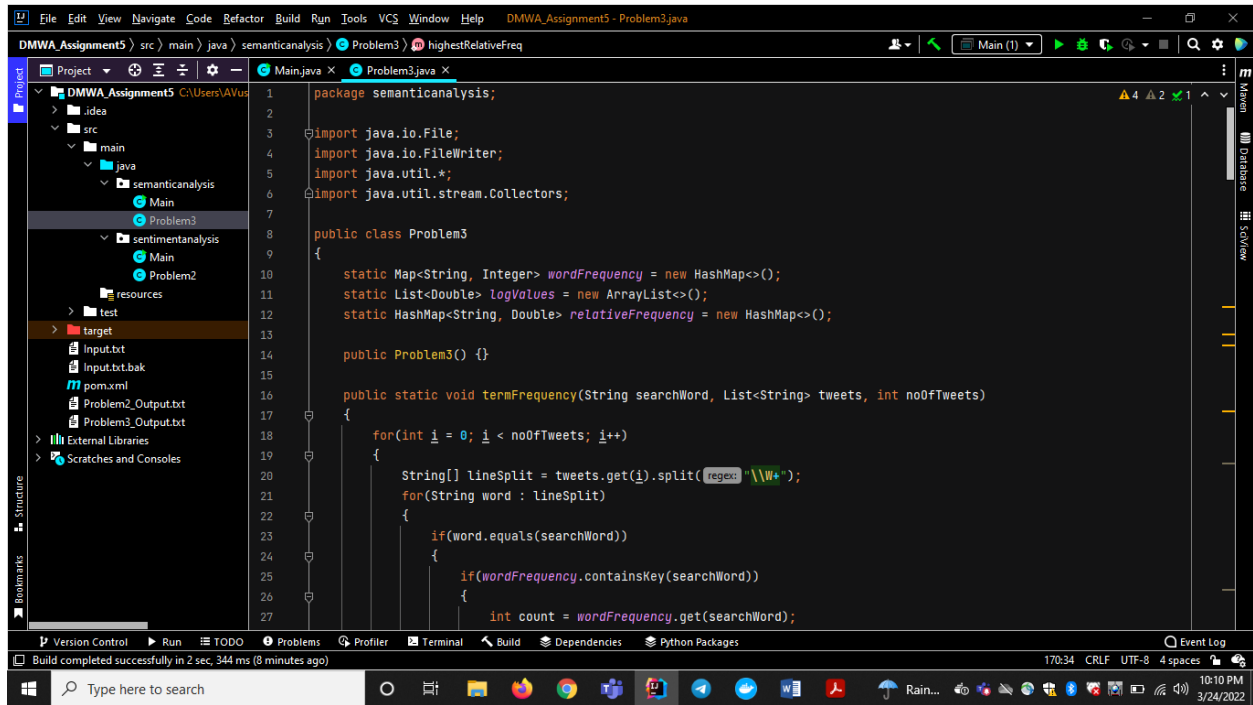
<https://git.cs.dal.ca/rguturu/csci-5408-w2022-b00871849-gutururamamohanvishnu/-/tree/main/assignment-5>

Screenshots of the code written for this program:



The screenshot shows an IDE window titled "DMWA\_Assignment5 - Main.java". The code is in the "semanticanalysis" package and defines a "Main" class. The code reads from "Input.txt", processes tweets, and uses "Problem3.termFrequency" to calculate term frequencies for search words. The IDE interface includes a project explorer on the left, a code editor in the center, and a status bar at the bottom.

```
1 package semanticanalysis;
2
3 import java.io.File;
4 import java.util.*;
5
6 public class Main
7 {
8     public static void main(String[] args) throws Exception
9     {
10         String[] searchWords = {"weather", "people", "condition"};
11         String filePath = "Input.txt";
12         List<String> tweets = new ArrayList<>();
13         File myFile = new File(filePath);
14         Scanner sc = new Scanner(myFile);
15         while(sc.hasNextLine())
16         {
17             String data = sc.nextLine();
18             tweets.add(data);
19         }
20
21         int noOfTweets = tweets.size();
22
23         for (String searchWord : searchWords)
24         {
25             Problem3.termFrequency(searchWord, tweets, noOfTweets);
26         }
27     }
28 }
```



The screenshot shows an IDE window titled "DMWA\_Assignment5 - Problem3.java". The left sidebar displays a project structure for "DMWA\_Assignment5" with folders like "src", "main", "java", "semanticanalysis", "Main", "Problem3", "sentimentanalysis", "Main", "Problem2", "resources", "test", and "target". The main editor area shows the following Java code:

```
1 package semanticanalysis;
2
3 import java.io.File;
4 import java.io.FileWriter;
5 import java.util.*;
6 import java.util.stream.Collectors;
7
8 public class Problem3
9 {
10     static Map<String, Integer> wordFrequency = new HashMap<>();
11     static List<Double> logValues = new ArrayList<>();
12     static HashMap<String, Double> relativeFrequency = new HashMap<>();
13
14     public Problem3() {}
15
16     public static void termFrequency(String searchWord, List<String> tweets, int noOfTweets)
17     {
18         for(int i = 0; i < noOfTweets; i++)
19         {
20             String[] lineSplit = tweets.get(i).split(regex: "\\W+");
21             for(String word : lineSplit)
22             {
23                 if(word.equals(searchWord))
24                 {
25                     if(wordFrequency.containsKey(searchWord))
26                     {
27                         int count = wordFrequency.get(searchWord);
```

The bottom status bar indicates "Build completed successfully in 2 sec, 344 ms (8 minutes ago)". The system tray at the bottom shows the time as 10:10 PM on 3/24/2022.

I am printing the tweets and their required details for all three parts (a, b, c) in the assignment, on the console for validation purpose. Details of all the tweets are stored in a text file.

Screenshot of tf-idf table on console:

The screenshot shows the IntelliJ IDEA IDE interface. The main window displays the output of a Java program executed from the command line. The output consists of two tables.

**Table 1: tf-idf table**

Search Query	Documents having term(df)	N/df	Log10(I/df)
weather	38	3425/38	1.9548769792116343
people	229	3425/229	1.1748250934885505
condition	5	3425/5	2.8356905714924254

**Table 2: term frequency table**

Term	Frequency(f)	Weather appeared in 38 documents
Total words(m)		
22	3	RT pscunnygc: From cold and rainy to shorts weather in one afternoon--we march for a weather NewDeal4CUNY in any weather
48	2	ashleyroybal My concern is less than how it does in snow and more how does it do in -40 weather days. My car battery has
15	2	Im over this cold weather bull shit , bring on the 80 degree weather alreadyid":"1500922924521476097
25	2	this spring weather is fucking up my allergies and it dont help that it be nice one day and cold af the next id":"15009

Screenshot of term frequency table on console:

The screenshot shows an IDE window titled 'DMWA\_Assignments5 - Main.java'. The main editor area displays a term frequency table for the word 'weather'. The table has three columns: 'Term', 'Frequency(f)', and 'Weather appeared in 38 documents'. The table lists 22 terms with their frequencies and the documents they appear in. The IDE interface includes a menu bar (File, Edit, View, Navigate, Code, Refactor, Build, Run, Tools, VCS, Window, Help), a toolbar, a project explorer on the left, and a status bar at the bottom.

Term	Frequency(f)	Weather appeared in 38 documents
22	3	RT psscunyg: From cold and rainy to shorts weather in one afternoon--we march for a weather NewDeal4CUNY in any weath
48	2	ashleyroybat My concern is less than how it does in snow and more how does it do in -40 weather days. My car battery h
15	2	Im over this cold weather bull shit , bring on the 80 degree weather alreadyid:"1500922924521476097
25	1	this spring weather is fucking up my allergies and it dont help that it be nice one day and cold af the next id:"1500
11	1	tan123 Man global warming has resulted is some unusually cold weather
12	1	soapachu Someone who cant handle cold weather... grew up in Mid Wales
10	1	my baby's warm hugs for today's cold weather
28	1	Libertybroad BonoDoesLiberty Guys, the weather today was perfect. My game however was total shit. I might as well been
53	1	katx01 catherine___c H'mmm \"global warming\" Malaysia has just had snow Vietnam having it's coolest weather in 40 year
33	1	farmy_prx7 Usually 4 coat over them for me XD I just need layers to shed because I can't stand cold weather, but warm
14	1	clairesworld4 Snow is bad I couldnt agree more there we just have windy weather
12	1	bro a BUG landed on me please bring back the cold weather
15	1	awtlivv Our hot and cold weather is killing me rn. Make up your damn mind
49	1	ashleyroybat My concern is less than how it does in snow and more how does it do in -40 weather days. My car battery ha
16	1	Weather sudden increase in gusts and snow . Wilded a tiny bit. ShareYoua Weather PEI Canada Canadian PrinceEdwardIsl
40	1	We picked up another good win yesterday in a settled weather game. All our guys played tough through the snow and rain to
43	1	WINTER WEATHER ADVISORY IN EFFECT FROM 1 AM TO 10 AM EST WEDNESDAY. Snow expected. Total snow accumulations of 2 to 4 i
28	1	This the kind of weather I LOVE, but its the cold in the morning thats upsetting me cause Im ready to have my feet out i
14	1	The weather is cold. Na to lie down for bed sleep.

Below is the tweet which has the highest relative frequency (f/m):

```
RT BTSArmy_47: \"The weather is cold, wear warm clothes and gloves\" JUNGKOOK:
Perfect weather, perfect snow, perfect grooming, perfect day source:\"Twitter for iPhoneauthor_id\":\"95763778id\":\"150019
Not sure what omicron or covid i have coz I don't feel sick.. Only flu because the weather so cold .. And little cough.
No way its really boutta snow tmrw after having this nice weather makes no sense
My latest story for MedillChicago is out I took a look at how homeless people are being forced to seek shelter on the C
If you missed the countryfile weather forecast, then here's a summary Monday - Cold and dry Tuesday - Cold and windy W
How exactly will the F-35 help with Arctic security Its range is pretty limited for that purpose. Its basing and support
How Far South Could Wintry Weather
ErieGreenDeep Bright sunny, but cold here in the eastern Sierra foothills just the beginning taste of spring and perff
Environment Canada weather advisory urges motorists to be cautious this afternoon
Dry conditions today will give way to snow tomorrow. Be sure to watch this installment of Erich's Weather Outlook for t
Captaincoby00 It depends on where you live and how long your growing season is. You might want to check what you want t
```

Screenshot of the tweet with highest relative frequency on console:

Below is the tweet which has the highest relative frequency (f/m):

Tweet: How Far South Could Wintry Weather

Relative Frequency: 0.16666666666666666

Process finished with exit code 0

All tweets and details are stored in “Problem3\_Output.txt” file. Screenshot of the output file generated for Problem 3:

```
C:\Users\AVuser\IdeaProjects\DMWA_Assignment5\Problem3_Output.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
Problem3_Output.txt
1 Below is the tf-idf table:
2 -----
3 Total Documents 3425
4 -----
5 Search Query Documents having term(df) N/df Log10(N/df)
6 -----
7 weather 38 3425/38 1.9548769792116343
8 people 229 3425/229 1.1748250934885565
9 condition 5 3425/5 2.8356905714924254
10 -----
11
12 Below is the term frequency table:
13 -----
14 Term weather
15 -----
16 Total words(m) Frequency(f) Weather appeared in 38 documents
17 -----
18 22 3 RT pscounyc: From cold and rainy to shorts weather in one afternoon--we march for a weather NewDeal4CUNY in any weather
19 Allin4SUNYCUNY
20 48 2 ashleyroybal My concern is less than how it does in snow and more how does it do in -40 weather days. My car battery has
21 died a couple times this past winter. When a car relies solely on battery power, will it run in this cold of weather
22 15 2 Im over this cold weather bull shit , bring on the 80 degree weather alreadyid:"1500922924521476097
23 25 1 this spring weather is fucking up my allergies and it dont help that it be nice one day and cold af the next
24 id:"1500922883866038280
25 11 1 tan123 Man global warming has resulted is some unusually cold weather
26 12 1 soapachu Someone who cant handle cold weather... grew up in Mid Wales
27 10 1 my baby's warm hugs for today's cold weather
28 28 1 libertyBroad BonoDoesLiberty Guys, the weather today was perfect. My game however was total shit. I might as well been
29 playing in snow because I couldnt have done worse
30 53 1 katx01 catherine__c H'mmm \"global warming\" Malaysia has just had snow Vietnam having it's coolest weather in 40 years.
31 The climate is controlled by ocean currents which in turn are controlled by the Moon amp Sun. No looney-leftie will ever be more powerful than the
32 Sun. The Sun wins every time
33 33 1 iarmy_prx7 Usually 4 coat over them for me XD I just need layers to shed because I can't stand cold weather, but warm up
34 quickly so going -5 to 13 is kinda.....
35 14 1 clairesworld4 Snow is bad I couldnt agree more there we just have windy weather
36 12 1 bro a BUG landed on me please bring back the cold weather
37 15 1 awliuv Our hot and cold weather is killing me rn. Make up your damn mind
38 49 1 ashlevroybal Mv concern is less than how it does in snow and more how does it do in -40 weather days. My car battery has
39 -----
Normal text file length: 8,742 lines: 63 Ln:1 Col:1 Pos:1 Unix (LF) UTF-8 INS
Type here to search Rain... 10:05 PM 3/24/2022
```

## References:

- [1] Alam, R. and Fedorenko. V., “*Sorting HashMap by values,*” Stack Overflow [Online]. Available at:  
<https://stackoverflow.com/questions/8119366/sorting-hashmap-by-values>  
[Accessed: March 24, 2022].
- [2] Skinner, C. and Storey, J., “*Java HashMap: How to get a key and value by index?,*” Stack Overflow [Online]. Available at:  
<https://stackoverflow.com/questions/3973512/java-hashmap-how-to-get-a-key-and-value-by-index> [Accessed: March 24, 2022].
- [3] Gene B., “*Java Format String for Table Display: Wrap at 80 chars,*” Stack Overflow [Online]. Available at:  
<https://stackoverflow.com/questions/32996910/java-format-string-for-table-display-wrap-at-80-chars> [Accessed: March 24, 2022].