Submitted By: Nosherwan

Registration Number: L1S21BSSE0012

Semester: 2nd

Section: N2

Project: Sentiment Analysis

**Introduction:**

My name is Nosherwan. I am from Peshawar. I did my FSC from FG degree College Peshawar Cantt. I have been assigned with Programming Fundamentals Project (Sentiment Analysis).

**Reading of Project:**

I have read the project many times. First, I did project in different way (creating dictionary without help of raw data file). Then one of my classmates told me to create dictionary from raw data txt file. After reading the project for many times, I concluded that first, I will create dictionary out of raw data file without garbage and stop words. Then user will enter a sentence and the program will remove garbage and stop words and compare the sentence with dictionary. In end the program will print the sentiment of sentence i.e., Positive or Negative.

StopWord Remove

Garbage Remove

RawData.txt

Positive Or Negative

Dictionary

**Problem Statement:**

In this project I will am creating dictionary and then taking sentence from user and check the sentence into dictionary and state the sentiment of sentence.

**Dictionary Generating:**

* From Raw data txt file will read word one by one.
* Send the word into garbage function to remove garbage words.
* Send the clean word to case change function to convert the word into small case.
* Send the word to stopwordremove function to remove stop words.
* In end will create dictionary from the words picked from raw data txt file.
* Create dictionary txt file.

**Sentence Input:**

* Take input from user word by word.
* Send the word to garbage Remove Function to remove garbage words.
* Send the word to stop words remove function to remove stop words.
* After removing of stop and garbage words.
* Saving the words into sentence txt file
* Now read the words from sentence txt file.
* Store the words into jagged dynamic array.

**Sentence Sentiment check:**

* Compare the words stored in jagged array one by one with dictionary.
* Count the words total and positive and negative count.
* Display Positive if positive count is greater.
* Display Negative if negative count is greater.

**Development Elements:**

**Pointers:**

I have used pointers variables to store address of dynamic array. While passing the array into other functions. In this project for generating dictionary and taking input from user I have declared dynamic array of size 20. I have passed the array into many other functions for removing of garbage and stop words. So, I used pointers.

**Char\*\* Sentence;**

**Char\* input=new char [];**

**Int \*rows=new int;**

**Int\* Columns;**

**Variables:**

I have also used many variables in this project. For example, for counting the size of word, comparing two words. I have also used bool variables. To check either word is garbage or not.

In this project I have used variables with name sizeOfWord, key, check.

**Int key;**

**Int sizeOfWord;**

**Bool check;**

**Int sizeOfwordTwo;**

**Filing(txt):**

I have also used file handing concepts in this project. For generating dictionary, I have read raw data form txt file. Then remove the stop words and garbage from the raw data and stored the cleaned words in txt file. I have also stored the dictionary int txt file.

The input sentence from user is also stored in txt file after removal of garbage and stop words.

**Sentence.txt**

**cleanData.txt**

**cleandatfull.txt**

**dictionary.txt**

**rawdata.txt**

**stopword.txt**

**Modules (Functions):**

**void clearFile(void);**

This function will delete all data on file except raw data txt file.

**void readRawData(void)**

In this function the program will read data from raw txt file word by word. And then send the word in garbage remove function.

**void rawGarbageRemove (char\*, int);**

Word sent from rad raw data function will be checked if garbage is found the function will terminate. If garbage is not found the word will be send to case change function to change the word into small letters.

**void caseChange(char\*, int);**

Word sent from garbage remove function will be converted into small letter and then will be send to stop words remove function.

**void rawStopWordsRemove(char\*, int);**

Word sent from case function will be compare with stop words if found the function will be terminated otherwise the word will be sent to wrote raw function.

**void writeRaw(char\*);**

The word will write into txt file named cleandatafull.

**void removeDuplicate(void);**

This function will read word from cleandatafull txt file and remove duplicate words from the file. Then save the word into another txt file.

**void DictionaryMaking(void);**

This function will read first word from full clean data txt file and then check the into clean data txt file words if words found its total count, positive count and negative count will be stored in dictionary txt file.

**void ClearSentenceFile(void);**

This function will clear sentence txt file. The txt file is using for storing sentence without garbage and stop words.

**void inputSentence(void);**

This function will ask user to enter sentence word by word. The function will another function named garbage remove to remove garbage word.

**void garbageRemove(char\*,int);**

word sent from input function will be checked if the word has garbage the function will be terminate other wise case change function will be called.

**void stopWords(char\*,int);**

In this the program will compare the word with stop words if the word is stop word the function will be terminate otherwise the word will be sent to write sentence function.

**void WriteSentence(char\*);**

This function will store the clean word into sentence txt file.

**int\* ReadColumnsFromFile(int\*);**

This function will read sentence txt file and initialized columns array.

**char\*\* ReadSentence(int\*,int\*);**

This function will also read sentence txt file and stored the words into jagged array.

**void DictionaryRead(char\*\*, int\*, int\*);**

This function will compare the words stored in jagged array with dictionary file and count its total count positive count and negative count and display the sentiment of sentence.

**void DisplaySentence(char\*\*, int\*, int\*);**

The function will display the sentence.

**void deleteSentence(char\*\*, int\*, int\*);**

This function will delete the sentence array.

**Real Life Applications:**

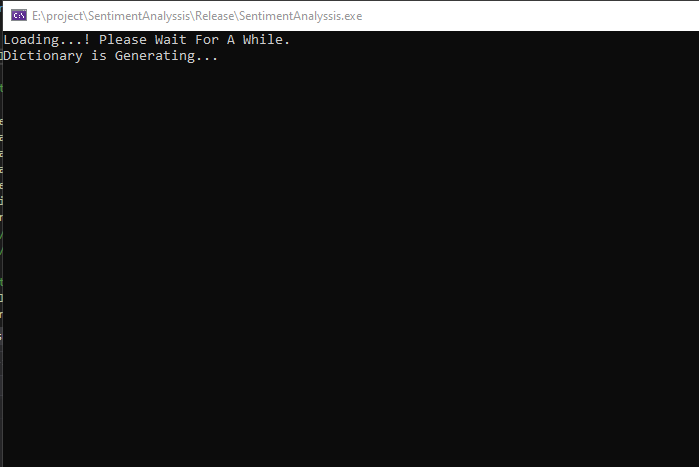
Almost all software companies use sentiment analysis on reviews or comments of the users who uses these software.

* Amazon has a data set of about 142.8 million reviews.
* IMBD has 50,000 review datasets.

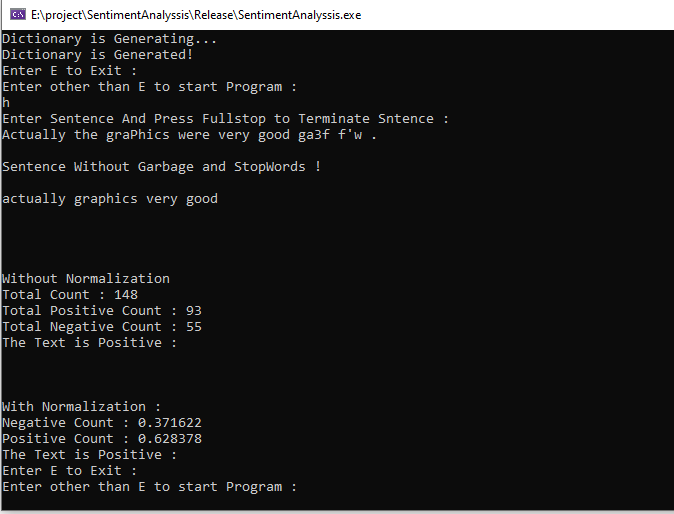
**Testing:**

* The Dictionary generates in 18 to 20 seconds.
* Once the dictionary is generated then no need to generate it again and again.
* In start of program the data on txt files is removed except raw data txt file.
* Sentence txt file is cleared every time
* If word is not found in dictionary the program will display message that the dictionary has no word.

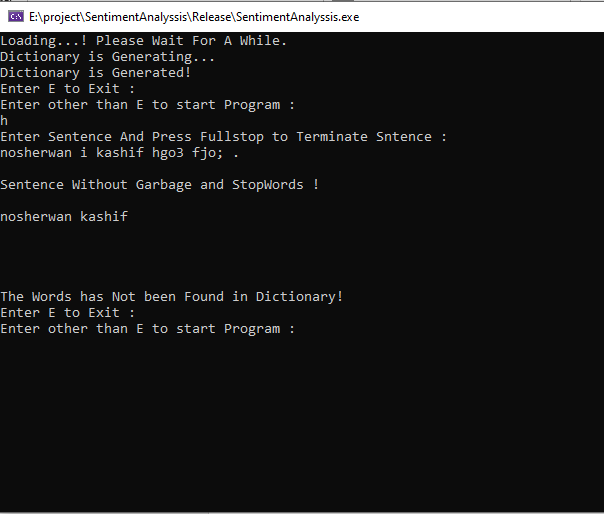
**Generates in about 20 seconds**.



**Dictionary has the words.**



**Dictionary Has no word found.**



**Learn Statement of Project:**

|  |  |  |  |
| --- | --- | --- | --- |
| Components | Use | Average | Best |
| For Loop | Y |  | Y |
| If else | Y |  | Y |
| Pointers | Y | Y |  |
| Functions | Y |  | Y |
| Variables | Y |  | Y |
| Jagged Array | Y |  | Y |
| While loop | Y | Y |  |
| 2D array | Y | Y |  |

**Reference:**

<https://archive.ics.uci.edu/ml/machine-learning-databases/00331/>

https://analyticsindiamag.com/10-popular-datasets-for-sentiment-analysis/