**Software Requirements Specification**

**For**

**Opinion Analysis Of Textual Data**

**15/03/2022**

**Prepared by**

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| CSE CCVT | 500075285 | Utkarsh Bahuguna |
| CSE CCVT | 500076356 | Gaurav Porwal |
| CSE CCVT | 500076608 | Sudip Maji |
| CSE CCVT | 500077157 | Amresh Garg |



Department of Informatics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007, Uttarakhand

Table of Contents

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | | 2 |
| Revision History | | 3 |
| 1 | Introduction | 4-5 |
|  | 1.1 Purpose of the Project | 4 |
|  | 1.2 Target Beneficiary | 4 |
|  | 1.3 Project Scope | 4 |
|  | 1.4References | 5 |
| 2 | Project Description | 5-19 |
|  | 2.1 Reference Algorithm | 5-7 |
|  | 2.2Data/ Data structure | 7 |
|  | 2.3 SWOT Analysis | 7-8 |
|  | 2.4Project Features | 8 |
|  | 2.5 User Classes and Characteristics | 8 |
|  | 2.6 Design and Implementation Constraints | 8-9 |
|  | 2.7 Design diagrams | 10-18 |
|  | 2.8 Assumption and Dependencies | 19 |
| 3 | System Requirements | 19-20 |
|  | 3.1 User Interface | 19 |
|  | 3.2 Software Interface | 19 |
|  | 3.3 Protocols | 19-20 |
|  |  |  |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. **INTRODUCTION**
   1. **Purpose of Project**

Its main objective is to determine the **emotional tone behind a series of words,** used to gain an understanding of the attitudes, opinions, and emotions expressed within an online mention. The primary goal of our project is to apply basic language processing and computational analysis to extract sentiments from source data. identifying and categorizing the opinions conveyed in a text to establish if the opinions or attitudes expressed are good, negative, or neutral.

Our sentiment analysis system's purpose is to produce an output value that indicates how positive, negative, or neutral the sentence or document is. Because of the complex nature of the writings, we want to perform sentiment analysis on them. We'll examine the user-generated material and categorize the comments as favorable, negative, or neutral.

* 1. **Target Beneficiary**

Sentiment analysis has a wide range of applications because opinions are highly essential to practically all human activities Whenever we need to make a decision, we seek the advice of others. The goal of this project is to check the feelings and polarity in a sentence.

Sentiment analysis is highly valuable in social network monitoring because it helps us to get a sense of what the general public thinks about certain topics. We evaluate user-generated material on social networks and categorize comments as favorable, negative, or neutral.

* 1. **Project Scope**

Sentiment analysis is highly valuable in social network monitoring because it helps us to get a sense of what the general public thinks about certain topics.

**Social media monitoring:** As social media grows in popularity, its frontiers expand. As a result, social media and blogging platforms are the best places to share your thoughts. As a result, it is critical to discern the type of perspective or sentiments the person is attempting to portray, as well as whether or not the material is healthy. We have an impact on our minds when we read social blogs, which can be pleasant or harmful.

**Customer support (Analyzing reviews):** Before making any decisions, we read other people's reviews and thoughts. So, it’s difficult to check what type of responses are they?

If we are purchasing something online there are types of responses that are difficult to judge. So, this can help us by telling the polarity of reviews.

* 1. **References**
* Bo Pang, Lillian Lee, and ShivakumarVaithyanathan, 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP Proceedings.

* Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL Proceedings.

* Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization for rating scales. ACL Proceedings.
* P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews,” Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002

1. **PROJECT DESCRIPTION**

2.1**Reference Algorithm**

**Proposed Algorithm:**

The Proposed algorithm contains the following features:

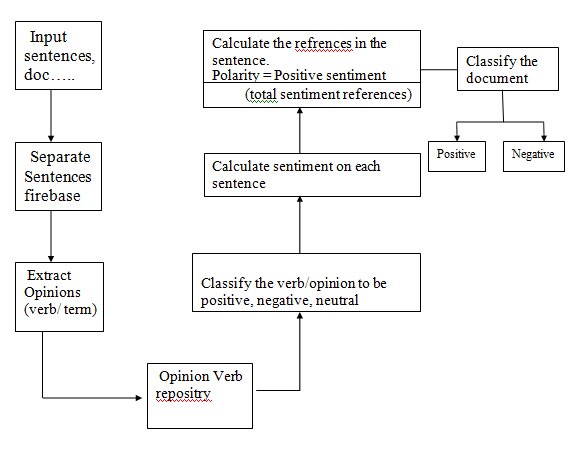
• Function to extract comments.

• Function to divide and categorize the text into positive or negative.

• Function that matches words (from comments) to the built list of words.

• Function to calculate the polarity of statements.

1. **Architecture:**

The system architecture is as shown below:****

1. **Modules:**

This algorithm can be divided into 5 modules:

1. Creation of vocabulary and data sets:

* Build two separate lists namely
* Positive/Good words. Ex: good, awesome, useful, pretty
* Negative/Bad words. Ex: rude, dirty, worst, horrible

1. Input: Extract comments from web pages.
2. Matching:

* After extracting the context

-divide it into a sentence

-perform word-by-word analysis of each sentence.

-categorize each word to be positive, negative, or neutral.

1. Calculate Polarity:

• Polarity indicates the percentage of positive sentiment references among total sentiment references. **Polarity = (positive\_sentiment)**

**(total\_sentiment\_references)**

1. Output:

Display the calculated polarity to the end-user

. Ex: Total words, Positive and Negative words, Polarity.

**2.2 Characteristic of Data**

Here, we are going to use:-

* Arrays
* Strings
* ArrayLists
* LinkedList
* Hash Maps

**2.3 SWOT Analysis**

1> Strength: -

1. Opinion analysis is a vital tool for any organisation or group whose public perception or attitude toward them is critical to their success - in whatever sense success is defined.
2. Millions of individuals are discussing and reviewing businesses and products on social media, blogs, and online forums. These thoughts, evaluations, and his massive quantity of data are gathered using computer programmes that not only find all mentions of their products, services, or company, but also determine the feelings and attitudes underlying the words.
3. The results of opinion analysis assist organisations in comprehending conversations and discussions in order to react and act appropriately.They can immediately detect any negative attitudes and transform bad client experiences into excellent ones.
4. Local government departments can evaluate public attitude about their department and the services they provide by listening to and analysing comments on Facebook and Twitter, and using the data to improve services such as employment.
5. Universities can employ opinion analysis to examine student opinions and comments gathered from surveys or internet sources such as social media. They can then use the information to identify and resolve any areas of student unhappiness, as well as to identify and expand on areas where students are expressing good feelings.

2> Weakness: -

1. Computer software have issues perceiving things like sarcasm and irony, negations, jokes, and exaggerations, which a human would have no trouble recognising. And failing to notice them can cause the results to be skewed.
2. There's a good chance that an automated opinion analysis technology wouldn't classify sentences or phrases expressions like sarcasm like ‘someone fails exam and teacher commented good job’. It’s most likely classified it as a positive statement.
3. There may not be enough context for a reliable opinion analysis using short sentences and fragments of text, such as those found on Twitter and occasionally on Facebook. However, Twitter has a reputation for being a strong source of information for opinion analysis, and with the new higher word count for tweets, it'll become even more valuable.
4. Automated opinion analysis techniques do a fantastic job at detecting opinion and attitude in text, but they aren't flawless.

**2.4 Project Features**

Overcome from the project is that we determine the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions, and emotions expressed within an online mention. It also helps in understanding public opinion, companies use sentiment analysis in doing market research and figuring out if their customers like a particular product or not.

**2.5 User Classes and Characteristics**

Users will be given a sentence or a paragraph to find the opinion of the sentence or the paragraph through which the user would be able to:

* evaluate the person's opinion in certain cases
* determine the emotional tone behind the series of the word
* teach the machine to analyze the various grammatical nuances
* implement an algorithm for the automatic classification of text into positive or negative.
  1. **Design and Implementation Constraints**

**PSEUDOCODE**

Step1: Define the files containing the words with their sentiment values.

Step2: Extract/Input the comment from user i.e. text[ ]

Step3: Divide each sentence in the text to form an array of words i.e. input[ ]

Step4: For each word in the string input[ ], check whether it is present in the list

of goodWords[ ] and badWords[ ] arrays .

1. If the word is found in goodWords[ ] list, increment the count of positive\_sentiment and also the total\_sentiment\_references
2. If the word is found in badWords[ ] list, increment the count of negative\_sentiment and also the total\_sentiment\_references

Step5: Calculate the general sentiment count

Step6: Display the calculated general sentiment count and the general sentiment of

text.

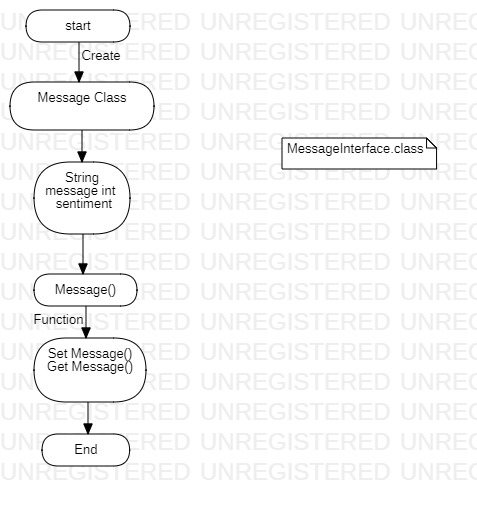
**2. ALGORITHM**

1. Start
2. Import Message file.
3. Message files take input from the user
4. Import MessageHolder
5. MessageHolder file stores the messages till the user enters \*\*\*
6. Calls PopulateMessage() function
7. Messages populated.
8. An import Dictionary file.
9. It checks sentiments.
10. Add the value of sentiments.
11. Open AF1NN file.
12. Search for input words and set sentiment values accordingly.
13. Import Analysis class
14. Check values of sentiments.
15. Differentiate positive, negative, and neutral sentiments.
16. Calculate overall sentiment. percentage of positive, negative, and neutral.
17. Delete the message holder.
18. Show output general sentiment on the output screen.
19. End

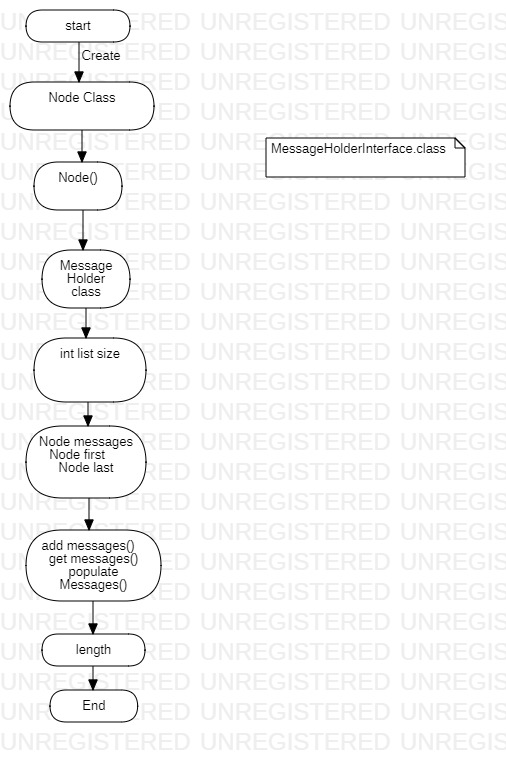
**2.7 Design diagrams**

**Flow chart**

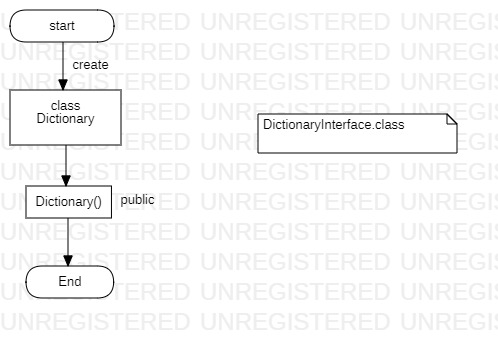
1. **MessageInterface.class**

****

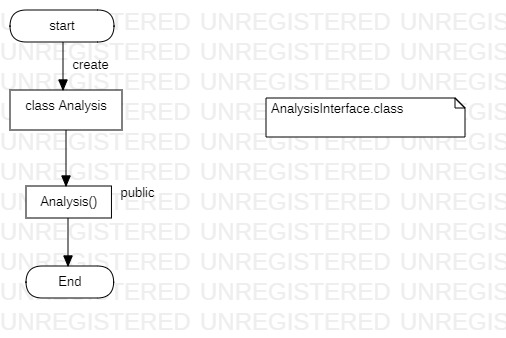
1. **MessageHolderInterface.class**

****

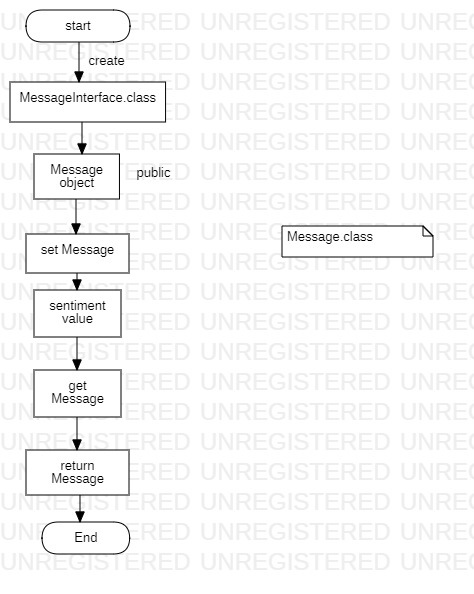
1. **DictionaryInterface.class**

****

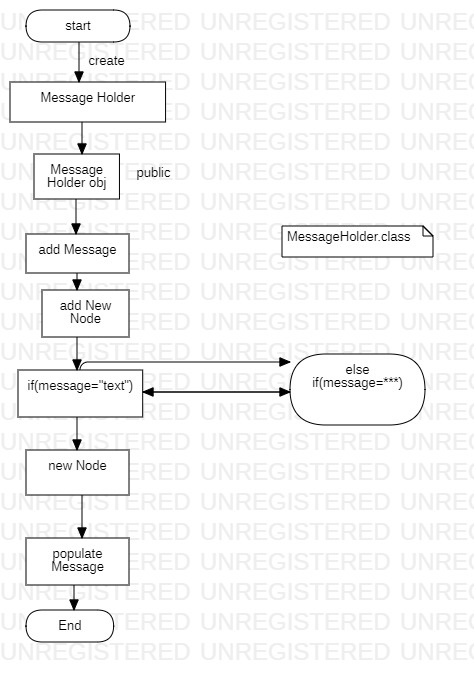
1. **AnalysisInterface.class**

****

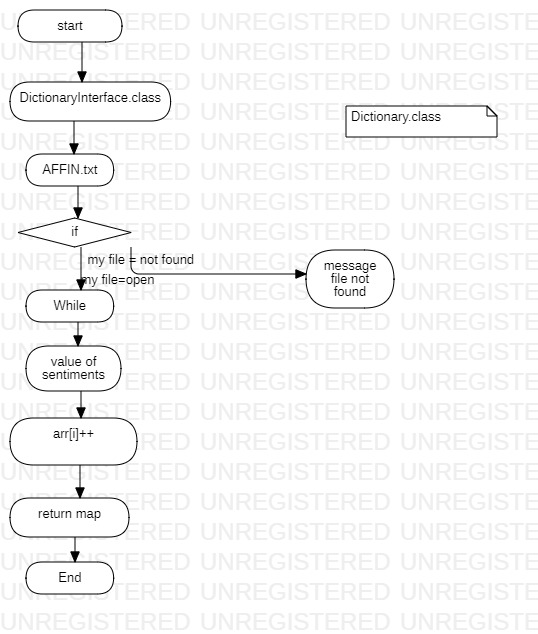
1. **Message.class**

****

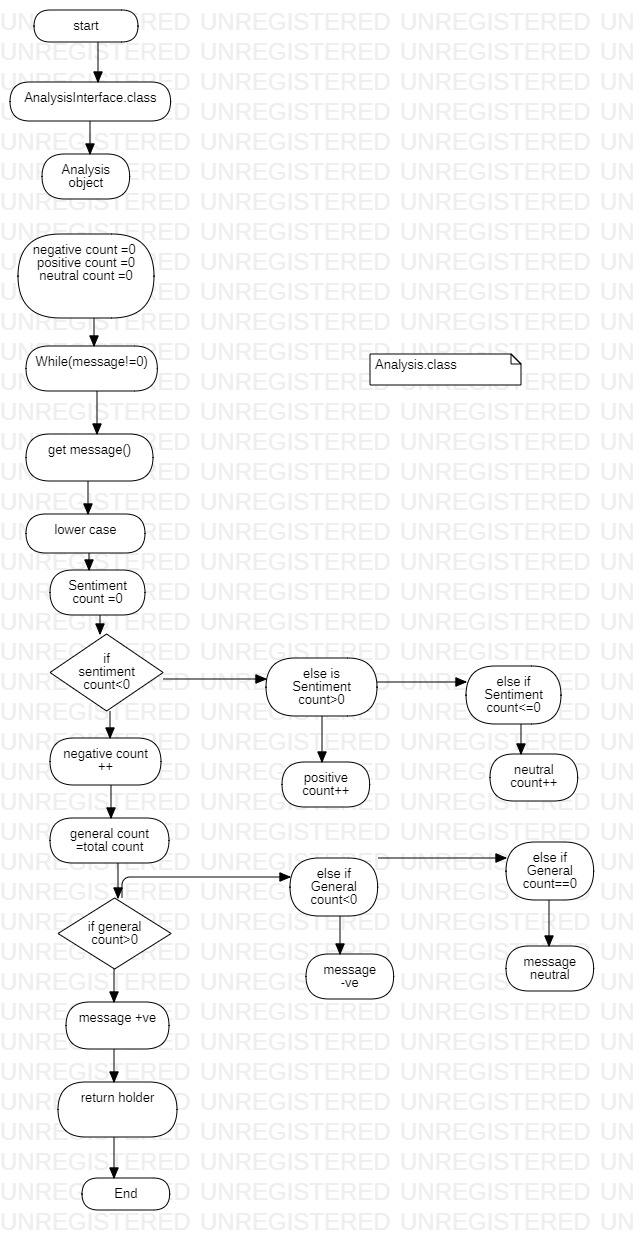
1. **MessageHolder.class**

****

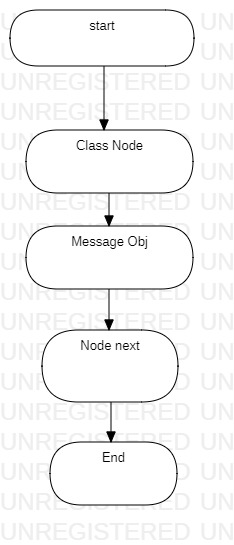
1. **Dictionary.class**

****

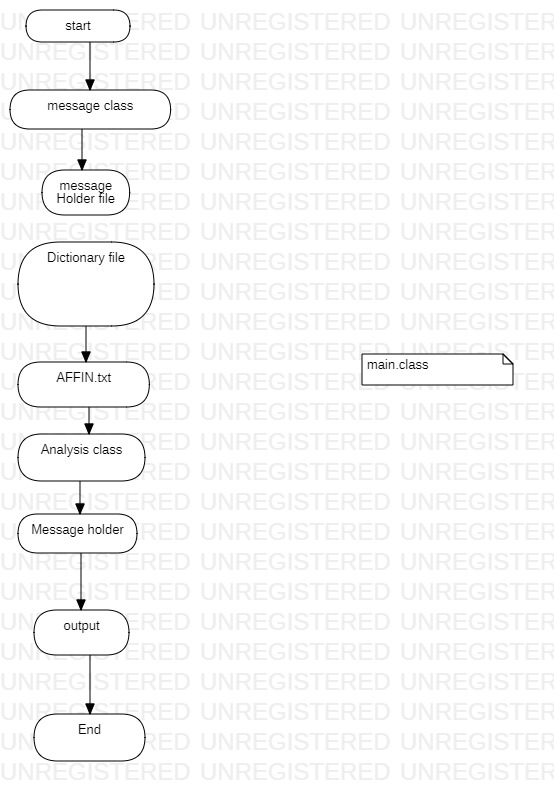
1. **Analysis.class**

****

1. **Node.class**



**10. Main.class**

****

**2.8 Assumption and Dependencies**

* Each word is separated through a space
* The value of each word is compared with the words through the AF1NN file to obtain the outcome.
* Due to new types of words and slang originating in social media and text messaging language it becomes problematic if a word is not present in the dictionary and the program can’t analyze the sentiment value of the word.

1. **SYSTEM REQUIREMENTS**

**3.1 User Interface**

**Hardware Interface:**

1. Minimum RAM requirement for proper functioning is 2 GB.

2. Required input as well as output devices.

**Operating system:**

Windows/Linux/OSX

**3.2 Software Interface**

**Software Interface:**

1. JDK 16

2. Eclipse IDE

**3.3 Protocols**

To apply opinion analysis: -

* Identifying whether a given textual entity is subjective or objective.
* Identifying the polarity of that subjective textual entity after removing the objective content.

**Implementation Issues: -**

1. Due to new types of words and slang originating in social media and text messaging language it becomes problematic if the word is not present in the dictionary and the program can’t identify the word.
2. It is still in the development phase so it still is inaccurate and inconsistent at times.