**ÇANKAYA UNIVERSITY**

**FACULTY OF ENGINEERING**

**COMPUTER ENGINEERING DEPARTMENT**

**Software Requirements Specification**

**CENG 407**

**Sentiment Analysis in Turkish**

**Naz BAYRAKTAR 201111502**

**Merve TEMÜR 201111050**

**Senior Project**

**Advisor: Murat SARAN**

**Co-advisor: Ulaş GÜLEÇ**

**Table of Contents**

Table of Figures..........................................................................................2

1. Introduction........................................................................................... 3

1.1. Purpose................................................................................................3

1.2. Scope of Project .................................................................................3

1.3. Glossary...............................................................................................4

1.4 Overview of the Document ..................................................................4

2.0. Overall Description..............................................................................5

2.1 System Environment ............................................................................5

2.2 Functional Requirements Specification................................................5

2.2.1 User Use Case....................................................................................6

2.2.2 Admin Use Case................................................................................7

2.2.3 System Use Case...............................................................................8

2.3 User Characteristics..............................................................................9

3.0. Requirements Specification...............................................................10

3.1. External Interface Requirements .......................................................10

3.1.1. User Interface..................................................................................10

3.1.2. Hardware Interface..........................................................................10

3.1.3. Software Interface...........................................................................10

3.1.4. Communication Interface................................................................11

3.2. Functional Requirements....................................................................11

3.3. Non-Functional Requirements............................................................12

3.3.1. Overview of the Data ......................................................................12

3.3.2. Logical Structure of the Data...........................................................13

**1. INTRODUCTION**

**1.1 Purpose**

In this project, we will get tweets and location information from Twitter, and we will keep this information in a database. We will do sentiment analysis on Turkish sentences according to the tweet contents in the database. For example, the positive/ negative emotions in these Tweets according to the hashtags and specific contents (name, organization name, etc.) will be extracted. The main aim of the project is to develop a new algorithm to do sentiment analysis on Turkish sentences. In the scope of the project, Turkish tweets in the health, politics, education , municipalities, and brands areas will be analyzed by using the algorithm that will be developed to figure out the opinions of the users about the selected topics.

**1.2. Scope of Project**

Sentiment analysis has been an active research area for quite some time. Unfortunately, most works are specific to the English language. In this work, we present a framework for unsupervised sentiment analysis in Turkish text documents. In this project, we will get tweets and location information throughout Twitter (is a social networking and microblogging service that allows users to post real time messages, called tweets.), and we will keep this information in a database. We will do sentiment analysis on Turkish sentences according to the tweet contents in the database. For example, positive / negative and neutral emotions , hashtags (Users usually use hashtags to mark topics. This is primarily done to increase the visibility of their tweets) and content (specific name , organization name, etc.) The aim of the project is to develop a new algorithm to do sentiment analysis on Turkish sentences In Turkish, most of the negation is done by the negation suffix in the predicate, which usually affects the whole sentence The added suffixes may change the polarity of words, for example, adding (“-ma”) to the word will change the polarity of that word and so on. Also we will study effect of adjectives and effective words. Moreover, Turkish has several letters that are missing in English (“ç”, “ı”, ”ğ”, “ş”, “ü”, ”ö”) and, in informal writing on the Web especially on Twitter, people tend to substitute these Turkish letters with the closest ASCII English letters (“c”, “i”, “g”, “s”, “u”, “o”). This creates complication in identifying the words. We decided to create a sentiment analysis framework taking into account the above mentioned differences. For this purpose, there are some modules to correct the sentences of the input text that we decided to use in our project such as Sentence extractor, ASCII character converter, Morphological analyzer, Negation handler, and Polarity predictor. These modules could help us to analyze the Turkish sentences and create the algorithm about it. We will also have, Microblog data like Twitter, on which users post real time reactions to and opinions about “everything”, poses newer and different challenges. Also, people may use some emoticons (These are facial expressions pictorially represented using punctuation and letters; they express the user’s mood) in their tweets, they may use tweets ending in positive emoticons like “:)” “:-)” as positive and negative emoticons like “:(” “:-(” as negative so we try to clear these at the same time catch the emoticons and classified they to categories (negative, positive and neutral) in Database. Also, we will take the location of the public tweets that we can access, so we design CSS that could shows the map of Turkey and the emotions in all the cities in turkey. As a result, we decided to design windows form application for this project.

**1.3 Glossary**

|  |  |
| --- | --- |
| Term | Definition |
| Categories | General title is of the topic that the user wants to see the sentiment analysis results. For example; Education. |
| Subcategories | Topics that user's wants to see the result of sentiment analysis. For example; Cankaya University. |
| Hashtag | Tweets that includes hashtag related subcategory which is taken from twitter. |
| Word | Tweets that includes word related subcategory which is taken from twitter. |
| Location | Using information of geographical location of thrown tweet's, according to location sentiment analysis results are reported. |
| Database | Tweets that belong to sub-category are saved in database for sentiment analysis. |
| Sentiment analyses | System will analysis the tweets as positive, negative and neutral sentiments about a topic the user wants. Tweets in database will be analysis. |
| Admin | Member registration is done, add the category and subcategories. |

**1.4 Overview of Document**

The remainder of the document includes three parts. The second part is about the functionality of the system.

In the third part it will give more information and it will contains more details about the system with different diagrams. Also, it includes functional and non-functional requirements specifications in detailed

These two parts describe of the projects software details. This system appearance to for our user who is used to this system.

**2.0 Overall description**

**2.1 System Environment**

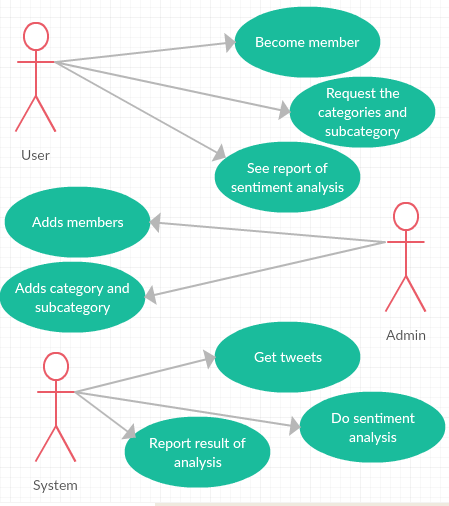
****

Figure 1: System Environment

In our project has three active actors. Admin adds new user records. If an institution wants to see its own sentiment analysis, firstly it need to fill out the membership form members can use the system approved by the admin. Admin adds subcategories that users want. User becomes member for using system. User requests the categories and subcategory. User can see reporting the sentiment analysis of categories that he/she wants .System get tweets on twitter according to subcategories that is wanted by user. System does sentiment analysis and reports the results to the user.

**2.2 Functional Requirements Specification**

Our project has three actors. These actors are user, admin and system.

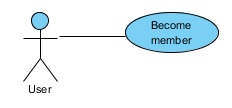
**2.2.1. User Use Case**

**User Use Case**

**Use case**

* Become member

**Diagram:**

****

**Brief Description:**

The user will sing up to our system by filling the mandatory part of login page.

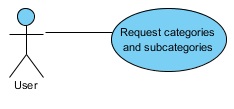
**Initial Step-By-Step Description:**

1. The user fills in the register form.
2. User must get the name, surname, Email, birthday, address, and phone number to register to our system.

**Use case**

• Request categories and subcategories

**Diagram:**

****

**Brief Description:**

After registration the member can access to the categories and subcategories.

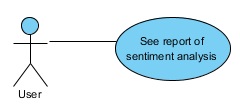
**Initial Step-By-Step Description:**

1. In registration form user can requests the categories and subcategory.
2. Member can access to the categories and subcategories in the system.
3. Members select a subcategory.

**Use case**

• See report of sentiment analysis

**Diagram:**

****

**Brief Description:**

The member can reaches the reporting of sentiment analysis using subcategories.

**Initial Step-By-Step Description:**

1. Members select a subcategory.
2. Member chooses a time interval.
3. Member can see the sentiment analysis reporting which is appropriate to the selected options.

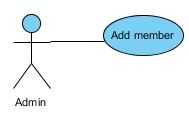
**2.2.2. Admin Use Case**

**Admin Use Case**

**Use case:**

* Add members

**Diagram:**

****

**Brief Description:**

The admin accept the people who register to our system as a members

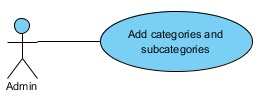
**Initial Step-By-Step Description:**

1. The admin can accept the user that want to sing up to our system as a member.

**Use case:**

* Add categories and subcategories

**Diagram:**

****

**Brief Description:**

The admin create or add new categories and subcategories to the system.

**Initial Step-By-Step Description:**

1. after the accepting of user, the admin can add categories and subcategories to our system
2. the categories can be

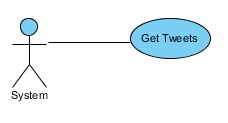
**2.2.3. System Use Case**

**System Use Case**

**Use case:**

* Get Tweets

**Diagram:**

****

**Brief Description:**

Our system will get the tweets form Twitter.

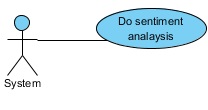
**Initial Step-By-Step Description:**

1. System will access to the open sections on Twitter
2. Get the related tweets
3. Set them to the Database

**Use case:**

* Do sentiment analysis

**Diagram:**

****

**Brief Description:**

Our system will do sentiment analysis to the tweets that gotten.

**Initial Step-By-Step Description:**

1. The system will get the sentences in Turkish

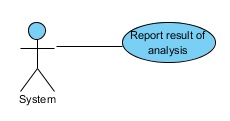
Clean the sentences from the unnecessarily letters or expressions.

1. Classified it into three emotional branch positive, negative and neuter

**Use case:**

* Report result of analysis.

**Diagram:**

****

**Brief Description:**

Our system will tender the result of sentiment analysis to the tweets that gotten as positive, negative and neutral emotions.

**Initial Step-By-Step Description:**

1. Report the analyses according to the categories in all of the cities in Turkey
2. The analysis will be on the map with three colors green for the positive emotions, yellow for the neuter emotions and red for the negative emotions.
   1. **User Characteristics**

All users should register to our system and be a member that can pursue systems categories and subcategories. Also, user can show the top tweets on our system.

**3.0 Requirement Specification**

**3.1 External Interface Requirement 3.1.1 User interfaces**

* **A Colorful Map of Turkey:** In the project we will have a map of Turkey in our home page that shows the mood of people if they generally are sad, happy or neutral. These moods will be as emoji symbols on the map. The colors in the map will be divisible into three major colors (Red for sad mood, Green for happy mood, and Yellow for neutral mood).
* **Login Page:** Another requirement will be login page for the members or the people who want to join our application.
* **The List of Categories:** Users can access the categories such as health or education that they want to search for.
* **Results Reporting:** Detailed results report should be provided to the user about the analysis.
  + Results report should include the searched keyword or hashtag
  + Results report should include the number of tweets related with the searched keyword or hashtag that were analyzed
  + Results report should include the timestamp of the analysis
  + Results report should include city location information of the tweets (e.g. Ankara) and the sentimental results of the analysis (e.g. 70% positive opinions)

**3.1.2 Hardware interfaces**

We do not need any hardware interfaces in our project.

**3.1.3 Software interfaces**

* **C# Language:** We will implement the project using C # language.
* **Visual Studio 2015:** We will use Visual Studio 2015 IDE and Windows form application project template.
* **SQL Server Management Studio 2012:** Beside this, we use SQL Server Management Studio 2012 for our Database ınterface. For our application we will create the Tables which are (Tweet, User, Categories, Hashtag, and Admin Tables).
* **NuGet:** We will implement our application in C# language as a Windows form application and we will use NuGet package manager to use Twitter API library.
* **GitHub:** We will use GitHub which is a Web-based Git repository hosting service in our project. Beside this, we will use SourceTree as a tool to connect to the Git repository. These tools will allow us to access the project resources exactly where we want to be.

**3.1.4 Communications interfaces**

* **Twitter and Dev Twitter APIs:** We will use a Twitter account (proje20152016) for our application to access the tweets by using the Twitter and Dev Twitter APIs.

**3.2 Functional Requirement**

**3.2.1 Sentiment Analysis**

**3.2.1.1 Introduction/Purpose of feature**

Tweets will be collected and analyzed according to the Turkish language supported algorithm and it should obtain a result that expresses sentiment as positive, negative or neutral emotion.

**3.2.1.2 Stimulus/Response sequence**

**i.** SearchTurkish tweets

**ii.** Store the tweets in database

**ii.** Tweet cleaning: Correct the spelling and grammar errors

**iii.** Tweet analysis for extracting emotions

**iv.** Provide sentimental result

**3.2.1.3 Associated functional requirements**

**3.2.1.3.1 Functional requirement 1**

Only the Tweets that are in the Turkish language can be transferred to the system.

**3.2.1.3.2 Functional requirement 2**

The contents of tweets are corrected according to the Turkish language grammar and spelling rules. Spelling and grammar errors should be corrected.

**3.2.1.3.3 Functional requirement 3**

Tweets should be analyzed in order to extract the positive/ negative emotions in these Tweets via the algorithm developed for the Turkish language.

**3.2.1.3.4 Functional requirement 4**

The system should give a result that expresses the sentiment of the sentence in percent.

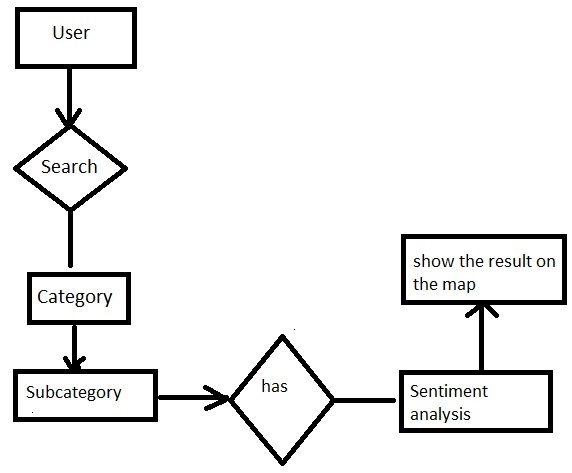
**3.2.1.3.5 Functional requirement 5**

The system should store the tweets and sentimental results in a database.

**3.3 Non-Functional Requirements**

The system should find the sentimental result above 70% accuracy.

**3.3.1. Overview of the Data** The general logic of our system is given below.



**3.3.2. Logical Structure of the Data**

The logic structure of data is in the Database for system.

Database Diagram

