CSCE 361 Fall 2016 Software Specifications

Group 9 - Austin, Christian, Thien, Tony, Andrew

October 21, 2016

1. Introduction

The project implements sentiment analysis on a person’s Tweets.

* 1. Purpose

This tool is intended to analyze a Twitter user’s general mood. It is intended to be used *in addition to* a human analysis component. Its purpose is to provide a basic overview of a user’s average mood, with the knowledge that human behavior is difficult to perfectly categorize, and therefore the results should be taken with a grain of salt.

* 1. Document Conventions

Since we are using Twitter and Tweets in this project, there are some specific conventions that we will be using. One of these conventions is specific to Twitter while others are used by other apps. The one specific to Twitter is called a ‘handle’ or ‘Twitter handle’. These are denoted by the form’@myhandle’ where ‘myhandle’ would be the username of the specific person you’re looking for.

We will also be using emojis. Emojis are small images or icons that are used to represent emotions/actions. There are thousands of different emojis that are supported by Twitter (2,477 when including skin tone and gender modifiers). Each emoji can possibly represent different things depending on how the user is using it. Examples of emojis are linked in the References section1.

The other major convention we will be using are called hashtags. These are denoted by #hashtag where the hashtag portion is a specific word or phrase. These can be related to a specific topic or idea. Hashtags are used to show if something is trending on Twitter.

* 1. Intended Audience

The intended audience of this project are human resources professionals looking for a tool to help them analyze potential hires’ social media posts. Taking time to look through tens or hundreds of posts could take hours, so having an automated tool to give an overview (which should be used to augment human analysis) can save valuable time.

A secondary group that may require some additional development could be school counselors interested in monitoring online bullying happening through Twitter. By identifying malicious Tweets and the users that write them (and to whom they are written), counselors would have a useful tool to help recognize students (both bullies and victims) that may need special attention.

Additionally, since this project is open source, middle and high school students will most likely utilize this tool to expand their social media stalking abilities.

* 1. Product Scope

The specific scope of this project is limited to a command line interface which is used to enter a Twitter handle and view output analysis. The design is limited so the user can only input a valid Twitter handle that is found in the database. Time permitting, a user interface will be added, as well as functionality to search for specific Twitter users based on their name (and potentially other criteria).

* 1. References

1Twitter Emojis, <https://www.piliapp.com/twitter-symbols/>

1. Overall Description
   1. Product Perspective

This product will be able to return an analysis of a Twitter user’s stored Twitter activity. First, each Tweet is analyzed and assigned a sentiment from a list, then a percentage breakdown by sentiment is returned by the user Tweet analysis engine, in addition to an overall sentiment expressed in a single user’s Twitter activity.

* 1. Product Functions

The following lists are requirements that the program will meet.

2.2.1 General Requirements

1. The program will take a valid Twitter handle as input from the command line interface
2. The program will generate output with:
   1. Percentage breakdown of individual Tweet sentiments
   2. Overall sentiment of the user
3. The program will update the database as it analyzes each individual Tweet with the associated sentiment as shown in table 1.

Table 1 - List of Sentiments

|  |  |
| --- | --- |
| **Positive:** | **Negative:** |
| Excited | Upset |
| Happy | Stressed |
| Relaxed | Nervous |
| Content | Sad |
| Elated | Bored |
| Enthusiastic | Angry |
| Love | Depressed |
| Grateful | Hate |
| Humour | Frustrated |
| Amused | Pissed off |
| Joyful | Exhausted |
| Optimistic | Overwhelmed |
| Peaceful | Anxious |

2.2.2 Command Line Requirements

4. The product will implement a command line interface that will take input from the user. The syntax of using the command line is as follows:

analyze *handle*

In this case, *handle* is a known Twitter user in the database.

5. The command line will print out the result of the sentiment analysis.

* 1. User Classes and Characteristics

2.3.1 Command Line User

A user accessing the command line will type in a Twitter handle. After confirming that the handle exists in the database, the program will hand off execution to the analyzer. After analysis of a user’s profile is complete, the output will be printed out to the command line for the user to view.

* 1. Operating Environment

The expected operating environment is Windows, OS X or a Linux/Unix based operating system. This is required for the command line interface.

* 1. Design and Implementation Constraints

Twitter does not allow free access to a Twitter user’s Tweets so a database of Tweets will be used. A data collector could in addition be set up to collect data live or a function could be added to allow the connection of paid access to twitter. These are potential improvements if time allows.

* 1. Assumptions and Dependencies

This product assumes that Tweets are all stored in a local database, and that any handle used is stored in that database. In reality, data will have to be taken from Twitter and the system will have to be adjusted to scale accordingly. The system depends on a database populated with Twitter handles and tweets.

1. External Interface Requirements
   1. User Interfaces

The product will implement a command line interface. Should time permit, a graphical user interface will be implemented and extra functionality added to the report generator and user search classes to produce a more robust and intuitive user experience.

3.1.1 Command Line Interface

The command line interface that will be implemented is described in 2.3.1. Use case diagrams for this are located in Appendix A. Section 2.2.2 contains specific requirements for the interface.

* 1. Hardware interfaces

No custom hardware will be implemented for this project.

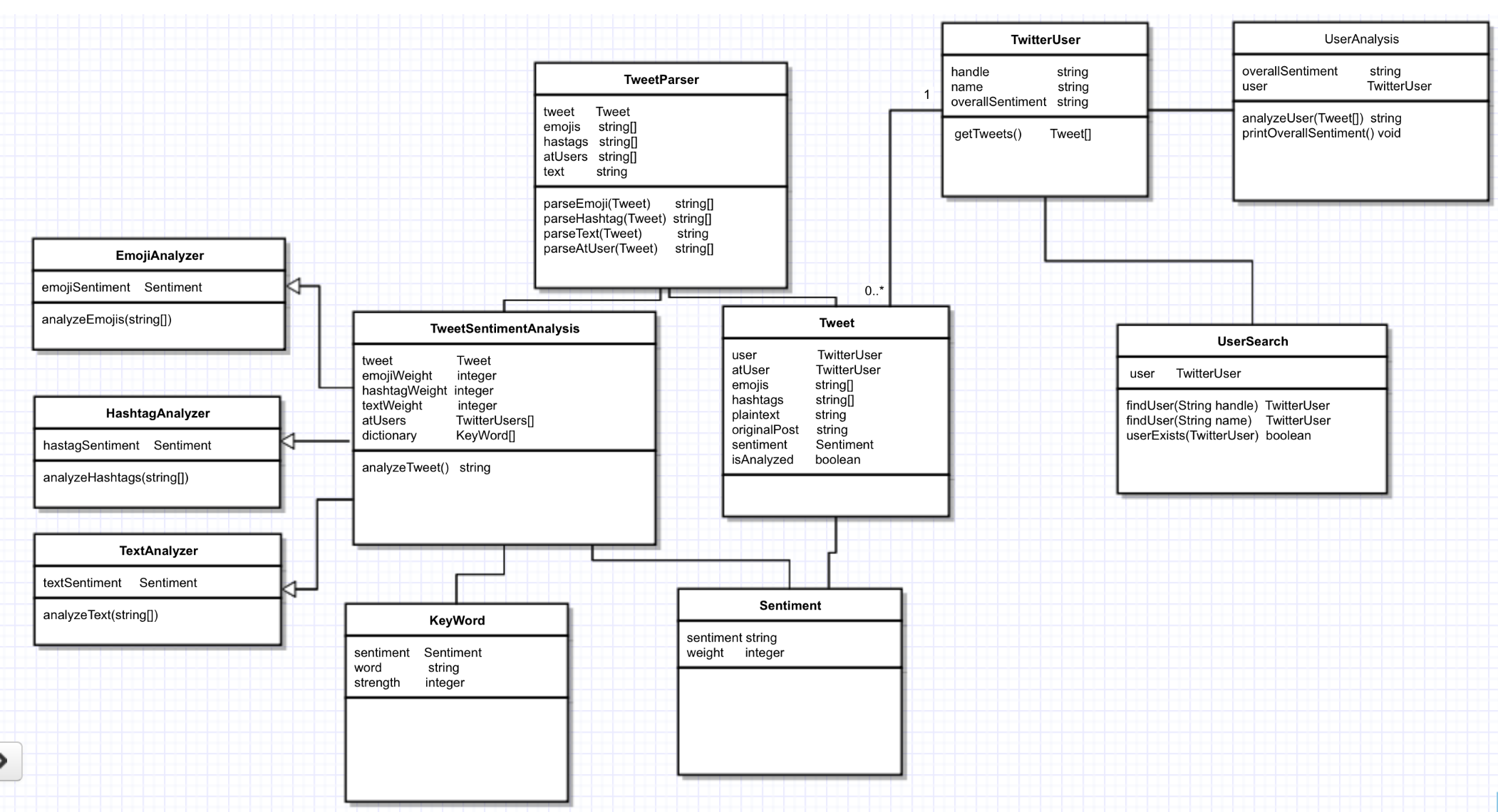
* 1. Software Interfaces

There will not be any software interfaces designed for this project.

* 1. Communications Interface

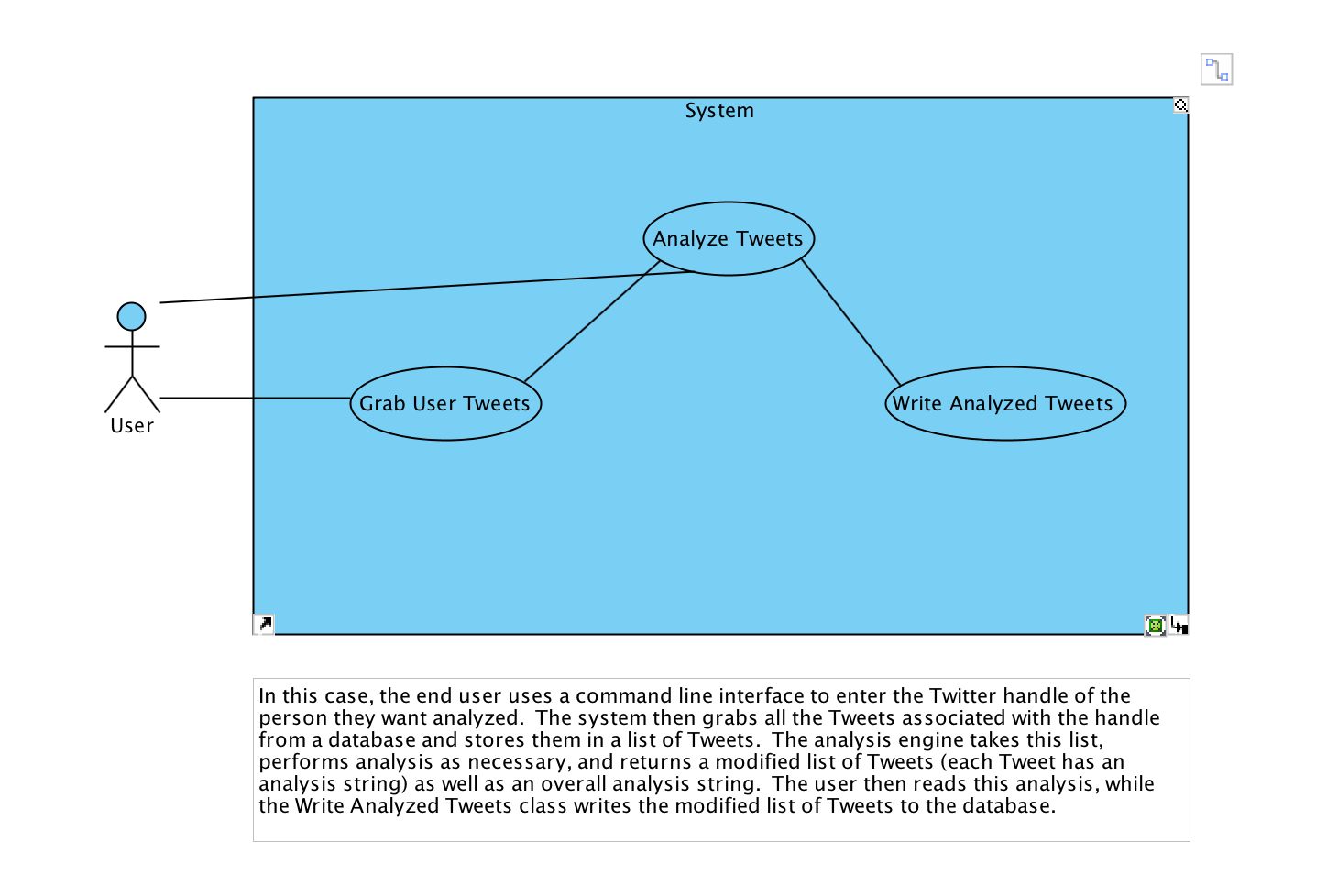
There will not be any communications interfaces implemented.

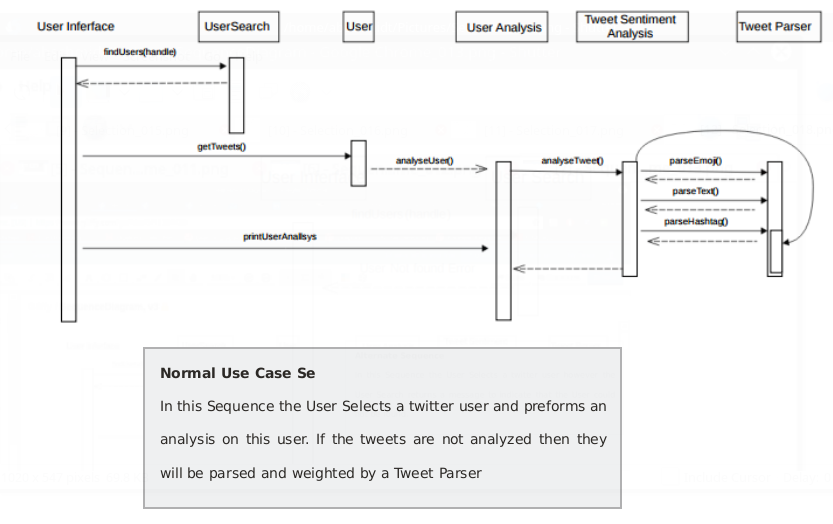
**Appendix A:** Class Diagram with Normal Use Case and Sequence Diagram



The idea of this structure is to balance the cost of string analysis with database space. Before any analysis report is generated for a specific user, each of their Tweets must be given a Sentiment. Through this structure, the system can analyze Tweets when the system load is low, potentially producing faster results. Each Tweet gets parsed into useful parts, then each of these are analyzed and combined to return a Sentiment on the Tweet. The UserAnalysis class then compiles a report based on all of a Twitter User’s Tweets and prints it to the standard output.

**Appendix B:** Normal Use Case and Sequence Diagram





**Appendix C:** Alternative Sequence Diagram and Use Case