Britney Johnson

Bricksquad

Introduction

Project Overview

* Construct a general-purpose Tweet Analyzer – meaning technical or non-technical customers can use custom or written classes to link to the tweet analyzer to perform simple or complex analysis of tweets
* Minimum of three classes –
* Tweet Analyzer class (uses methods in class to analyze tweets)
* Regular Expressions class (produces a list of tweets that matches specified keywords)
* and weather map class (analyzes real-time tweets across US to produce weather map)
* After developing analyzer, a Software Development Kit (which includes resources, design plan, milestones, etc.) must be created to assist customers who want to write own Tweet Analyzer class.

Scope

* Tweet Analyzer program must have a class must link to regular expressions to find matching tweets and another class that links to the weather class to produce a weather map based on tweets in US
* Target audience is the general population, so its to be fairly simple

Problem Statement

* Bricksquad will create web application to
* Retrieve tweets from Twitter API to perform a regular expression search
* Retrieve tweets from Twitter API to create a weather map based on tweets about weather
* Store received date in MySQL database
* User enter expressions for searching twitter to get results

Requirements

* Develop a program that allows users to accurately search through tweets on Twitter using keywords (regular expressions)
* List of requirements include technical requirements for the hardware/software platform and functional requirements are for the developed software
* \*Requirements – Marcellus

Interaction Models

User Analysis

* Tweet analyzer application will support users who are familiar with Twitter and those who have never registered an account. This is why it does not require users to register for a Twitter account
* Application will have clearly marked interface to allow users to navigate from a list of streaming tweets depending on the users’ regular expression entered and weather map
* Anyone can use the application, but knowledge of the English language as well as basic computer skills are need
* Knowledge about particular subjects depending on the regular expression is needed. Knowledge of weather by region in the US is also expected

Component Analysis

* Streaming Tweets
* Tweepy is a Twitter API, which provides access to the entire Twitter RESTful API methods, as well as grants the HTTP request after proper authorization
* Each method in the API can accept various parameters and return responses
* This API will use the Twitter API, which uses the REST format, to provide relevant tweets in order to search the past for info such as terrorism attacks
* Can also use the Streaming API that produces real time tweets in order to provide current info
* The endpoints for the HTTP method, known as POST statuses/filters, returns public statuses in JSON format.

Filter Twitter

* Filter the stream of tweets to provide only those including a regular expression
* The POST request is used because the GET request are often rejected due to excessive URL length
* One of the predicate parameters in the POST request is the Track – keywords/phrases (regular expressions) that are specified by a comma separated list

Parse Tweets

* Info received from tweets
* When the main class receives Tweet from Twitter using API commands, the tweet is returned in a JSON format, which is interchangeable between languages.
* To parse through JSON format in python, Bricksquad used the SimpleJSON implementation. It is a simple, fast, complete, correct and extensible JSON encoder and decoder for python.
* The decoder will handle incoming JSON strings of the tweets being fetched and will help gather only the useful data (location, index of reg expressions, etc)
* The number of retweets is also a measure of engagement. This shows how valuable the tweet is.

Tweet Processing Algorithm

* Algorithm used to handle the heavy stream of tweets.
* Responsible for fetching and processing tweets.
* Each tweet process will recheck the API’s filter in order to ensure the exact regular expressions match and it will store the proper information that needs to be obtained by each tweet in the correct place.

Store Tweets

* MySQL will be used in order to store the tweets being streamed real-time from Twitter
* After each tweet is arranged accordingly, the tweet will then be stored in MySQL

Parse regular expressions

* After the regular expressions class reads in a file containing the regular expressions, each line of the file that contains that regular expression will be parsed to an array.
* This array will be used to identify exactly how many times the keyword has been used, as well as for indexing purposes. The list will then be transferred to the analyzer class in order to define the keywords before the twitter search process begins.
* After each tweet is found, an alert will be sent to display the list of found regular expression tweets. This list is constantly updated.

Produce list of relevant (Tweet matching keywords) Tweets

* The regular expressions class for non-sophisticated users will produce a list of tweets that match any of the regular expressions.
* Flask, a free, microframework written in Python, will be used to display said Tweets.

Display map of real-time US weather map based on predefined regular expressions

* Google Maps provides a JavaScript API for Bricksquad to use. It gives the team the ability to design and implement a custom map that displays weather conditions on the US map for specified regions/cities/states. It already provides a weather layer with weather updates, but it can be customized.
* The JavaScript API allows the team to implement an algorithm that fetches JSON text for different weather conditions in the Continental US.
* The Tweet Analysis class will perform an analysis for tweets about the weather related regular expressions and store the info in JSON format for the Google Maps API to fetch. Once it fetches the info, it will iterate through the info and populate the different weather conditions on its own weather layer that Google allows to be customized.
* Google weather layer uses info from weather.com

Developing an SDK

* Bricksquad’s Software Development Kit (SDK) will assist users who want to write their own class to implement the Tweet Analyzer class
* SDK serves as a set of software dev tools that allow for the creation of application for a certain software package (Tweet Analyzer class)
  1. *Task Analysis*

**Regular Expression Task list:**

1. Add a text file for the regular expression in specified directory
2. Start the application
   1. Program verifies text file to user
   2. Program displays list of regular expressions before proceeding to analyze tweets
   3. Display to user that tweets are currently being analyzed
3. View of currently processed tweets that match regular expressions
   1. View will include location of tweets
   2. Index of matched regular expressions
   3. Character string of corresponding tweet
4. Constant update of matching tweets

**Weather Map Task list**

1. Start application
2. Process predetermined list of regular expression relating to weather
3. Display to user that tweets are currently being analyzed
4. View of weather map produced by currently processed tweets
5. Constantly update of weather map according to real-time tweets

Content Analysis

* The content of Bricksquad Tweet Analysis will be determined by the stakeholders and the dev team.
* Any data retrieved from the Tweet Analyzer will be pre-approved with the requirements of the stakeholder.
* Content will contain the tweets pulled from the search of the user.

Environment (System) Analysis

* Tweet Analyzer application will be created for the Windows Operating System and development is being handled on the Windows 7 platform

Interface Analysis

General Purpose Tweet Analyzer

* Most important feature of the application
* Used for the non-technical customer that will read a file containing regular expressions and applies the regular expressions to tweets producing a list of tweets that matches

Weather Map

* Displays a map of the US
* There will be a weather layer on top of the map, which populates the weather forecast from each individual state
* The forecast will provide an approximate temp for each state, along with the image of the state
* Temperatures may change in diff states, depending on what data it reads from Twitter
* The weather component will work directly with the tweet analyzer class

Configuration

* The project’s configuration depends on the team’s stakeholder’s operating system. Therefore, it was designed on a window’s environment.

Behavioral Diagrams

* Sequence Diagram: displays how the processes operate with one another and in what order they will be performed
* Weather Map Sequence Diagram: displays how the processes operate with one another and in what order they will be performed; describes that the user is responsible for opening the application and viewing the different pages.
* Activity Diagram: shows graphical representation of user scenarios with the Tweet Analyzer
* Weather Map Activity Diagram: shows graphical representation of user scenarios with the Tweet Analyzer
* State Diagram: illustrate the behavior of the Tweet Analyzer and displays the system composed of a finite number of states