**Intelligent Web Assignment**

**2a:**

**Introduction:**

This part of function will extract tweets form specific users and analyze the high frequent keywords from it in the last few days. The keywords will be sorted depending on the total amount. Moreover, the most frequent Y keywords used in the last Z days must be displayed (with associated total count, as well as count for each users).

**Issues:**

1.The tweets from users will contain a lot of punctuations and blank. This part must be considered otherwise there would be a plenty of illegal keywords come out.

2.The high frequency keyword may not exist in a specific user's tweets, which should also be considered.

**Design choices:**

The design of this part is not complex, but the algorithm of implementation is quite sophisticated.

First of all, we need extract tweets from a specific users. The days and users' screen\_name can be controlled by the search/tweets API. Following by token the tweets and eliminate the punctuations and blanks, we store them into arrays.

After we repeat this step for a few times(which are limited by the number of users entered), we can get a series of arrays which are contained all text of tweets we get from the API. We can combine them all, and divide it into individual words. Next, we could calculate the frequency of each words and sort them in the order. According from the keywords number we entered, we can put these specified high frequency keywords into an array. At last, we compare this array with the arrays of each users tweets, and get the date of each keywords for specific users.

Advantages:

we can decide the details of data we need by written algorithm, which can make our system stronger and more flexible.

Disadvantages:

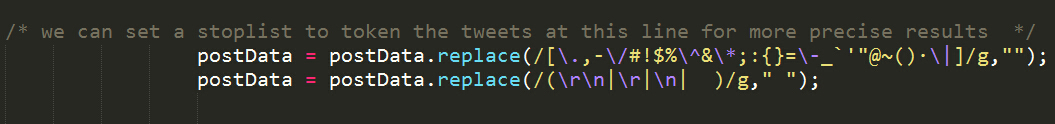
The process of getting and handling data is quite complex. All parts of implement are designed in the same function, there is lacking in several interface and encapsulation to control these functions. So it is quite difficult to debug the code and identify which part comes out the problem in a short time.

**Requirements:**

we can return the correct information to the frontpage, including the screen\_name and the keywords we entered. (which are based on the number and days). However, the data is not shown into a perfectly correct format. We can send the correct array which store correct data to the frontpage, but failed to show it into a table. The reason is that the frontpage concern the other part of this project and we have limited time to fix it.

**Limitations:**

1.The principle of token tweets is quite loosen due to there is not specific requirement in assignment. If it does, it is quite easy to add a stop-list to do the token work.



2. The way of extract and cope with data is complex. We should divide the whole requirement into smaller parts and make it happened into separated functions. So that we can debug our program more effective and more readable.

**2b:**

The aim of this question is to find venues the user visited in last several days, the main strategy to solve this problem is shown below: Firstly, I used the search/tweets API to get the user’s tweets with Swarmapp check-in messages; then, find use split() to find the check-in id from user’s tweets; after that,use foursquare checkins resolve API(https://api.foursquare.com/v2/checkins/resolve) to get venue name through the check-in id we got.

**Issues:**

There are so many Issues I met when I was doing this problem; firstly, I need to look for a correctly API from both twitter and foursquare API documentations to get desire date. Secondly，I had to analysis data I got to retrieve the desire information(ie. Check-in Id),,after that, the largest issue is that some users cannot find through Twitter and foursquare API, so that it makes a large challenge during testing the system.

**Design choice:**

For overcoming the problem of searching user’s tweets with check-in information, I used search/tweets API swarmapp/com/c/ as query and set parameter “from ”be a given user, and I used setDate() to get the days we need.

For finding out all the user’s tweets with check-in message, I split ‘expanded\_url’ properties of tweets to get check-in ids because it is easy to find that all the check in messages are shown by the format swarmapp/com/c/check-in ID, After that, I used foursquare API 'https://api.foursquare.com/v2/checkins/resolve' to get venues name.

For storing the information, I built two tables in the database which are named “users Information”(which used to store “user name”, ”user id”, ”user location” and “user’s description”) and ”locationVisited”(which used to store user id, venue id and venue name) ,I design them with the concept of relational database, I used “user id” in the table “usersInformation” as the foreign key in the table “locationVisited”.

**Advantage:**

The advantage of this system is that it can basically implement all the functionalities of assignment demand.

**Disadvantage:**

Firstly the system is not so robust and always suffers from all kinds of issues, such as some users cannot find through this system.

Secondly, the system is not so efficiency and always cannot retrieve some user when processing.

Finally, the database is not so efficiency when accessing data.