Lab 1 – Option A (Reddit Comments)

Alex Adrian Avila

The University of Texas at El Paso

CS 2302 – Data Structures

Introduction:

For this lab we must traverse through the comments of a reddit post a using reddit’s API to get all the comments on a certain post. We were given three methods that calculate if the comment is either positive or negative. So, we only need to traverse through all the comments and replies of those comments in order for us to decide if the comments is positive or negative then they be stored in different lists depending on their status

Proposed Solution:

For my solution I decided to traverse all the comments as if they were store in a three, and I would use pre-order traversal which in terms of recursion will be represented as: Visit Tree, go to the right, go to the left. However, in this case I analyze the comment, then go to the next comment, then go to the replies. These recursive call should analyze all the comments in any given post

At first, I tried to traverse through the main comments only to see if I could do it recursively. I used an index as a parameter to keep track of what comment I am at incrementing the index every time and as a base case I the index needed to be smaller than the length of the list, and the method worked as expected. Next, I needed to traverse through the comments and I did that by accessing comment[index].replies and also resetting the index so that we get to the first reply.

Experimental Results:

* When I use the post that was given by the instructor, the program divides all the comments into three arrays and I calculated the number of comments on each array:

Number of positive comments: 11

Number of negative comments: 2

Number of neutral comments: 186

* An example of each comment would be:

Neutral comment: [here's a good reddit post about recursion](<https://www.reddit.com/r/learnprogramming/comments/5w50g5/eli5_what_is_recursion/>)

Positive comment: very clever indeed

Negative comment: You cheeky bastard!

* I also tested the following post “<https://www.reddit.com/r/learnprogramming/comments/d1fbfy/google_summer_of_code_2019_turned_out_to_be_a/>” when the post was new and it did not have any comments. And it returned arrays with no comments
  + Number of positive comments: 0
  + Number of negative comments: 0
  + Number of neutral comments: 0
* After some hours the post received some comments my program updated the comments so I got some comments
  + Neutral comment: This sounds like an ad
  + Positive comment: This is a great post !
  + Negative comment: No, it doesn’t

Test Cases:

1. By using the given link, the comments in the reddit post are traversed and separated into three arrays that are global variable, which store positive negative and neutral comments as expected
2. process\_comments stores the comment and the responses that other people have made to that comment in form of arrays. Therefore process\_comments was tested with an array that is empty and the three arrays that the method returns are empty as well just as expected.
3. “<https://www.reddit.com/r/learnprogramming/comments/d1fbfy/google_summer_of_code_2019_turned_out_to_be_a/>” was another reddit post that has fewer comments and the program works perfectly it returns only a few comments for any of the three categories.
4. For this test case instead of analyzing another post process\_comments will use one of the responses of the first post as an input for example comments[0].replies will analyze the comments and the arrays will have fewer comments only including the comments in that particular response.

Extra Credit Test Cases:

1. get\_oldest\_comment\_any gets the very first comment that appears on the post, and the other two methods to get negative and positive comments appear to be working as well
2. if the program uses the same method it twice it will get the second method the second time just as it was expected
3. If the method is used more than there are comments in the post then the oldest comment will always be the last one.
4. Using the method with no comments as input will leave the comment arrays empty as well.

Conclusion

This lab is the first project that I have done on Python, so it was a good experience learning all the syntax and how to work with global variable and methods. Also, the installation of nltk and praw were good experiences as it was also my first-time installing python libraries. And one of the most important things that we learn is that we can access information from the web by using Application Programing Interfaces (APIs). These allow our apps to update its information if the information of the API changes.

**Appendix**

*def* process\_comments(comments, index):  
 *# if the array of comments is less than the list then return  
 if len*(comments) <= index:  
 *return  
  
 # assign a number for the probability of each comment* curr\_comment = comments[index].body  
 pos = get\_text\_positive\_proba(curr\_comment)  
 neg = get\_text\_negative\_proba(curr\_comment)  
 neu = get\_text\_neutral\_proba(curr\_comment)  
  
 *# Add comment to list depending on its category  
 global* positive\_comments  
 *global* negative\_comments  
 *global* positive\_comments  
  
 *if* pos >= neg *and* pos >= neu:  
 positive\_comments.append(comments[index])  
 *elif* neg >= neu *and* neg >= pos:  
 negative\_comments.append(comments[index])  
 *else*:  
 neutral\_comments.append(comments[index])  
  
 *# Go to the next comment until there no more* process\_comments(comments, index + 1)  
  
 *# Go to the first reply of the comment* process\_comments(comments[index].replies, 0)