CS594: Python Programming Lab

Take Home Assignment - 4 (2 Questions, 100 Points)
Submission Dead Line: 03-Oct-2019 23:59 Hours Pages: 3

IIT Guwahati

12 Sept 2019 (Thu)

Question 1: (0 points)
Reading assignment

Twitter API `https://developer.twitter.com/en/docs.html`

Twitter Open Authentication API `https://dev.twitter.com/oauth`

Tweepy documentation `https://tweepy.readthedocs.io/en/v3.1.0/`

Question 2: (100 points)
Implement in Python Programming Language the following problems:

Q1. 100 Marks Read the above twitter API interface

Your task is to build a crawler that collects

(a) A user's profile information from Twitter given the user's Twitter ID. User profiles describe about several attributes of the user. The following is a snapshot of github profile information (the user ID is 13334762).

Attribute	Value
Twitter User ID	13334762
Screen Name	github
User Name	GitHub
User Location	San Francisco, CA
User Description	Social coding
Total Followers	1,849,804
Total Statuses	4,790
User URL	`https://t.co/FoKGHcCyJJ`

Given a user ID, your task is to obtain the above information

(b) A user's social network information given the user's ID. User's social network is a directed graph. There are two types of connections between users: *follower* and *friend*. Let v_i and v_j be two nodes in the graph. v_j *follows* v_i if there is a directed edge between v_j to v_i. v_j is *friend* of v_i if there is directed edge between v_j to v_i *and* directed edge between v_i to v_j.
Following is the snapshot of the user's network information (User ID: 13334762)

(i) The friends list

- GitHubCommunity
- GitHubEducation
- mcolyer
- benogle

- lildude
- ...

(ii) The followers list

- lavanyarisa
- asucan_rick
- Rodoshi12
- sarj_cdr
- ...

Given a list of *user's IDs*, write a crawler to collect users' screen names of user's friends and followers.

(c) The tweets using a specified keyword or a geolocation based criteria. Search on twitter is through keywords or geographic regions. Twitter provides APIs to collect tweets that contain specified keyword or originated from a given geographic region. Returned tweets of the search are in JavaScript Object Notation (JSON) format.

(i) Given a keyword, collect the tweets that are associated with the keyword.

(ii) Given a geographic region in the form of $[longitude\_1, latitude\_1, longitude\_2, latitude\_2]$ collect the tweets that are associated with the geolocation.

Instructions File Naming Convention Create a directory with your roll number. Inside this directory, place all the above python programs and input files. Prefix the file name with your roll number followed by "_" followed by question number followed by ".py". Example: `194161000_q1a.py, 194161000_q1b.py, 194161000_q1c.py`.

README.txt Write a short notes on sequence of steps involved to run the your programs. Include what is the input for the program (with an example) and what will be the output from the program (with an example).

tar gzip Create (roll number).tar.gz file using the above directory. This directory must contain the above three python source files

Submission Email the above tar gzip file to the CS594 TA `vaibhav18@iitg.ac.in` as per the above given dead line

Copying You should avoid indulging in copying. Every submission will be subject to software similarity using the tool `Measure of Software Similarity` available at `https://theory.stanford.edu/~aiken/moss/`. Two submissions having similarity score equal to or more than 40.0% will be declared copied. If you are found involved in copying act, your name will be referred to disciplinary committee. Therefore you are requested to place individual efforts and avoid copying.

Marking Scheme Your implementation will be evaluated as described below.

Q1a 25 Marks Correct logic and correct output

Q1b 25 Marks Correct logic and correct output

Q1c(i) 25 Marks Correct logic and correct output

Q1c(ii) 25 Marks Correct logic and correct output