**Tracing of Information Flow in Social Network**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Title** | **Summary** | **Document** |
| 1. | Ushare: user controlled social media based on blockchain | <https://docs.google.com/document/d/1WYInMowsgxBICH-muYEqHWAP-eQdgv0n_0sCrCIcbkg/edit?usp=sharing> | <https://drive.google.com/file/d/0BxdBfELcvvx5U25ueHE1NUQ4OEU/view?usp=sharing> |
| 2. | Step 1: I have only text part of tweets in CSV files. The programs are divided as   1. task 1a.py 2. task1b.py | For search API, the output is saved in result.csv.  For streaming API, the output is in tweets stream1.csv | [task1a.py](https://drive.google.com/open?id=0BxdBfELcvvx5Nm9NX2V1YUNtMU0)  [Result.csv](https://drive.google.com/open?id=0BxdBfELcvvx5SXJHRllqdHo1bFE)  [Task1b.py](https://drive.google.com/open?id=0BxdBfELcvvx5Y3JNQ2k4MXFBaHc)  [tweet stream1.csv](https://drive.google.com/open?id=0BxdBfELcvvx5XzhhV3dBaUFvVms) |
| 3. | Step 2: Tweets as json object using streaming API are stored in MongoDB | Collection used : search  Database used : twitterdb | [task 2a.py](https://drive.google.com/open?id=0BxdBfELcvvx5ZlBITkg2ME1yTzA)  [Task 2b.py](https://drive.google.com/open?id=0BxdBfELcvvx5ZmFoeF95U3k4bVU) |
| 4. | Step 3: |  |  |

@himani 28.9.2017

Title: Data Provenance on Social Networks

Steps:

1. Write python program to collect tweets containing a particular keyword say ‘xyz’. One program to use search API and another program to use streaming API.
2. Save all the collected tweets in MongoDB.
3. Perform following analysis (write python code for each of them):
   1. How many times tweet with same (or almost similar i.e. more than 80% similarity) content gets posted ? (Hint: For similarity, use jaccard coefficient)
   2. What are the timelines of these similar tweets ? Can we trace when the earliest tweet was posted ?
   3. Who are the users posting such similar tweets ? Are they related to each other through followee-follower relationship ?

Base Paper:

<http://mediamining.univ-lyon2.fr/people/guille/publications/survey.pdf>

Rishabh Sir and Ankita Ma’am,

Motivation for project:

Video can be found [here](https://drive.google.com/open?id=0BxdBfELcvvx5QXJlTjJfTTdUTDQ).

Presentation 1(09.01.2017) :

[Tracing of Information Flow in Social Network](https://drive.google.com/open?id=0BxdBfELcvvx5MWwzVWh6SG9ZYVk)

Find the source of tweet : By comparing **created\_at** entity

Find the relationship of source and retweeters and the relationship of retweeters

: using **show\_friendship**

For analysis

<http://www.sciencedirect.com/science/article/pii/S0020025516308453?via%3Dihub#sec0002>

01.11.2017 Tasks assigned

1. Find the Jaccard coefficient of all the pair of tweets and also plot distribution graph

2. Identify similar content paths

3. Describe the follower-folllowee relationship

3333

**InfoTracer-Tool for tracing information flow**

1. Collect data from B.Tech Group(Srishti one).
2. Search for information from different domains and handles.
3. Defining parameters for