Facebook Profile Sociality Analyzer

TECHNICAL DOCUMENT

AUTHORS:

1. **INTRODUCTION**

This application aims to collect and analyze the reactions of any user’s friends…

1. **FACEBOOK DATA STRUCTURE AND RETREIVING MECHANISM**
   1. **Data retrieving mechanism**

In this application, we use Graph API (version 2.11) supported by Facebook to retrieve data. Facebook (FB) allows applications to retrieve a personal profile or a page with limited data fields. However, applications first must get the user token from the user profile. We obtained this by design a login webpage, registered as FB App. By this way, every Facebook user could be invited to login through this webpage and accept our requested data fields. After all, we would have that user token and their permissions to retrieve necessary data, in our case, are posts, comments, and reactions.

We will request directly to FB Graph API (https://developers.facebook.com/tools/explorer/) with necessary fields to retrieve the data. However, FB API does not allow developers to retrieve in an all-in-one data. Therefore, we will retrieve data in two (2) phases. (1) All the posts will be retrieved and saved to a file. (2) Using the retrieve posts data to extract all post-id and use that post-id to retrieve other fields (comments and reactions). Fields that are retrieved in this phase will then append together and to each of the post.

* 1. **Data structure**

The user token obtained from the user login and acceptance would be transferred to our Main class to retrieve data. The would-be-retrieved FB data is in JSON format (Fig. 1). There are 2 levels of data, we only take care of the 2nd level that is “data”. The Fig. 1 shows the JSON data retrieve from a posts field.

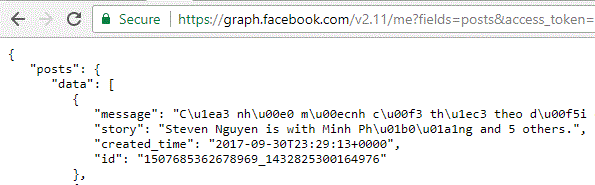


Figure . Example of a return of JSON (fields=posts)

As mentioned above, in phase 2, those data would be extracted to have the id and get other fields. Other fields would be appended continue to the “id” in this figure. In short, we will have a list of contents [list] (fig. 2), with each content have a list of message, storey, created\_time, id, comments [list], reactions [list].

Figure . Data structure created by the application

* 1. **Data storing and encryption**

After the first phase, data will be stored under a file with the name as format {filename}.{ext}

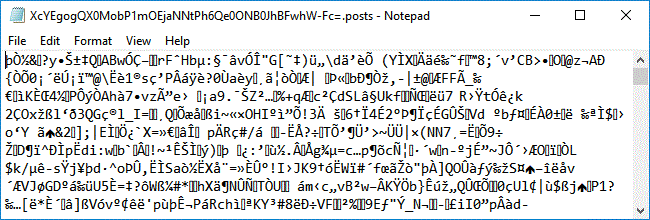
The filename is a string obtained from a hash of SHA-256 followed by a Base64 URL-safe-encoding. The hash function uses FB user-id as password and FB user-name as salt. Therefore, this will help to secure the file information that even the developer could not know the end user information. This is a one-way filename encoding which allows an end-user must login into the system with their permission so that the app could get the user-name and user-id to hash

For every phase, data will be saved to different file under different extensions (fig. 3). The scope of this is to store separately the posts file and the posts-comments-reactions file, which helps the app to save and read easier, avoid data losing and also for further development (scalable).



Figure . Data storing on file system

The extensions are set as “.posts” for posts only file and “.postsNreactz” for posts-comments-reactions file. Those extensions could be set whatever as we want. For privacy and security, we implemented an encryption and decryption (fig. 4) third-party module and applied inside the functions that take the responsibilities on reading and writing the data. Therefore, data would be encrypted and saved to ensure no one except the end-user could load it. Besides, we also implemented a delete method, which will delete even the encrypted save data on the file system whenever the end-users finish their application uses sessions.

  
Figure . An encrypted file (.posts) opened in Notepad

The module that is used for this is PyCryptodome ( <http://pycryptodome.readthedocs.io>). File would be encrypted and decrypted based on AES 128bit key. The key is the first 16 characters of a combination of FB user-id and user-name. This 16 characters form the max 16 bytes key used in encryption. By applying two different filename and data encryption, even us could not know which one to decrypted without kn

1. **PROGRAMING LANGUAGE**

Python was used to program this. This application was tested under Python version 3.x

Python flask was used to program the web connection to Facebook.

1. **MODULES AND CLASSES**
   1. Built-in modules
   2. Third-party modules
   3. Self-built modules
      1. Core

This module is used to retrieve the JSON data from Facebook, handling errors and retries

* + 1. FileIO

This module is used to handle file reading and writing. This file also ensures that data would be saved and open in UTF-8 encoding.

* + 1. Main

This module implements other modules (Core, FileIO) to retrieve, save and load data from the internet or from the saved data files.