

Python Script with Facebook API and Selenium

S. No	Component	Feasibility (Tested)
1	Posts	<ol style="list-style-type: none">1. Exact content of all posts.2. Unique Post ids3. Timestamps for each post.4. Unique user ids for each user who posts.5. List of user ids of users who have seen the post along with seen count.6. All reactions (Like, Love, Haha, Wow, Sad, Angry, any special reactions that come around sometimes like Pride reaction etc.) along with the user ids of users who reacted to any post. (reactions + user id pairs).

Python Script with Facebook API and Selenium

S. No	Component	Feasibility (Tested)
2	Comments or Replies	<ol style="list-style-type: none">1. Exact content of all comments and replies.2. Unique Comment ids and Reply ids.3. Timestamps for each comment and reply.4. Unique user ids for each user who comments or replies.5. All reactions (Like, Love, Haha, Wow, Sad, Angry, any special reactions that come around sometimes like Pride reaction etc.) along with the user ids of users who reacted to any comment or reply. (reactions + user id pairs).

References:

1. [https://en.wikipedia.org/wiki/Selenium_\(software\)](https://en.wikipedia.org/wiki/Selenium_(software))
2. <https://selenium-python.readthedocs.io/>
3. <https://developers.facebook.com/docs/graph-api/>

Functional set up for the user

Setup Overview

- The steps in the next two slides are to set up the group as a “closed” group on Facebook and set up the *developer options* that will give us access to the information we need. Essentially we need to pose as “developers” to get access.
- The developer options include setting up an “**app id**” (a 10-digit number) and “**user access token**” (a long alphanumeric string), which the administrator of the group will use to extract the required usage information
- The administrator needs to be able to set up a group on FB, and follow these instructions to set up the “user access token”, which the script we provide requires to operate on the group and extract the usage information for the researchers.

Setup Steps

1. Create a "Closed" Facebook group and edit its settings, add all group details.
2. Register on <https://developers.facebook.com> (With the Facebook id of the Admin of the closed group)
3. Click on "Add the app/Create new app" while hovering over "My Apps" on <https://developers.facebook.com/tools/explorer>
4. Choose a display name and a category and then "Create App ID"
5. Again open <https://developers.facebook.com/tools/explorer>, choose application from drop down "Application:" and then click on "Get Token" dropdown and further choose "get user access token"

Setup Steps

6. Select permissions from the menu in the popup windows as shown in the screenshot in next few slides, followed by clicking ok until the popup windows goes away and then get the token.
7. Go to link <https://developers.facebook.com/tools/accesstoken>
8. Select “Debug” corresponding to “User Token”.
9. Go to “Extend Token Access”. This will ensure that your token does not expire every two hours. Otherwise it will expire in 90 days, and this process needs to be repeated (steps 7 to 9).

Setup Steps

10. Create privacy policy document www.freeprivacypolicy.com
 - a. Step 1 - Choose "Website".
 - b. Step 2 - Choose "Individual", enter country name and province.
 - c. Step 3 - Enter Website URL as Facebook Group URL and Website Name as Group Name.
 - d. Steps 4, 5, 6 - Choose all "No" and enter basic information until privacy policy is available.
11. Download the privacy policy from that website as .txt file.
12. Upload privacypolicy.txt on google drive and create a shareable link.
13. Add this google drive URL in Privacy Policy URL in group.
14. There might be a need to extend token access one more time using the same process as above, if it expires.

Step 1 of 14

neurologytest - Google Chrome

Secure | <https://www.facebook.com/groups/neurologytest/edit/>

neurologytest

Joined ▾ ✓ Notifications Share ... More

Your changes have been saved.

Group Name

Group Type **Project** Change

Group types help people see what the group is about.

Description

Potential members see the description if privacy is set to public or closed.

Location [Edit Location](#)

Tags

Tags help people find groups they might be interested in joining. You can change these at any time. [Learn more](#)

Apps Tools and features offered by other businesses. [Learn More](#)

Chat

Step 4 of 14

Graph API Explorer - Facebook for Developers - Google Chrome

Secure | <https://developers.facebook.com/tools/explorer?method=GET&path=me%3Ffields%3Did%2Cname&version=v3.2>

facebook for developers Docs **Tools** Support My Apps Search developers.facebook.com

A beta version of the new Graph Explorer is available. [Try it out now!](#)

Graph API Explorer Application: [?] neuro developmental research

Access Token: [Get Token](#)

[Learn more about the Graph API syntax](#)

Node: me

- ☒ id
- ☒ name

```
{
  "id": "1159971707485055",
  "name": "Simon McDonald"
}
```

Step 4 of 14

neuro developmental research - + Add Product - Facebook for Developers - Google Chrome

Secure | https://developers.facebook.com/apps/ /add/

facebook for developers Docs Tools Support My Apps Search developers.facebook.com

neuro developmental... APP ID: [REDACTED] OFF Status: In Development View Analytics Help

Add a Product

Account Kit

Seamless account creation. No more passwords.

Read Docs Set Up

Facebook Login

The world's number one social login product.

Read Docs Set Up

Analytics

Messenger

Welcome, Simon!
Links, tools and resources

Get Started
Development Overview
Learn how to set up and configure your integration.

Learn About User Permissions
See the usage guidelines and permissions for Facebook products.

Learn About Available Features
See the usage and requirement guidelines for working with Features.

Tools for Testing
Create a Test User

Need more help?

[Platform Policy](#) [Privacy Policy](#) [Cookies](#) [Terms](#)
Facebook © 2018

Step 5 of 10

Access Token Tool - Facebook for Developers - Google Chrome

Secure | <https://developers.facebook.com/tools/accesstoken/>

facebook for developers Docs Tools Support My Apps Search developers.facebook.com

Access Token Tool

The user tokens listed here are provided for convenience to test your apps. They expire like any other user access token and should not be hard coded into your apps. App tokens do not expire and should be kept secret as they are related to your app secret. For more information on how access tokens work and should be used, see the documentation. If you want to debug an access token issue, try using the access token debugger.

neuro developmental research

User Token	<div><div></div></div>	Debug
App Token	<div><div></div></div>	Debug

LANGUAGES

[English \(US\)](#) Español 中文(简体) 한국어 日本語 Português (Brasil) Deutsch Italiano العربية ...

Products	Programs	Support	News
Artificial Intelligence	Developer Circles	Developer Support	Blog
Augmented Reality	F8	Bugs	Success Stories
Business Tools	FbStart	Platform Status	Videos
Gaming	ThreatExchange	Facebook Developers Community Group	Facebook for Developers Page
Open Source			
Publishing			
Social Integrations			

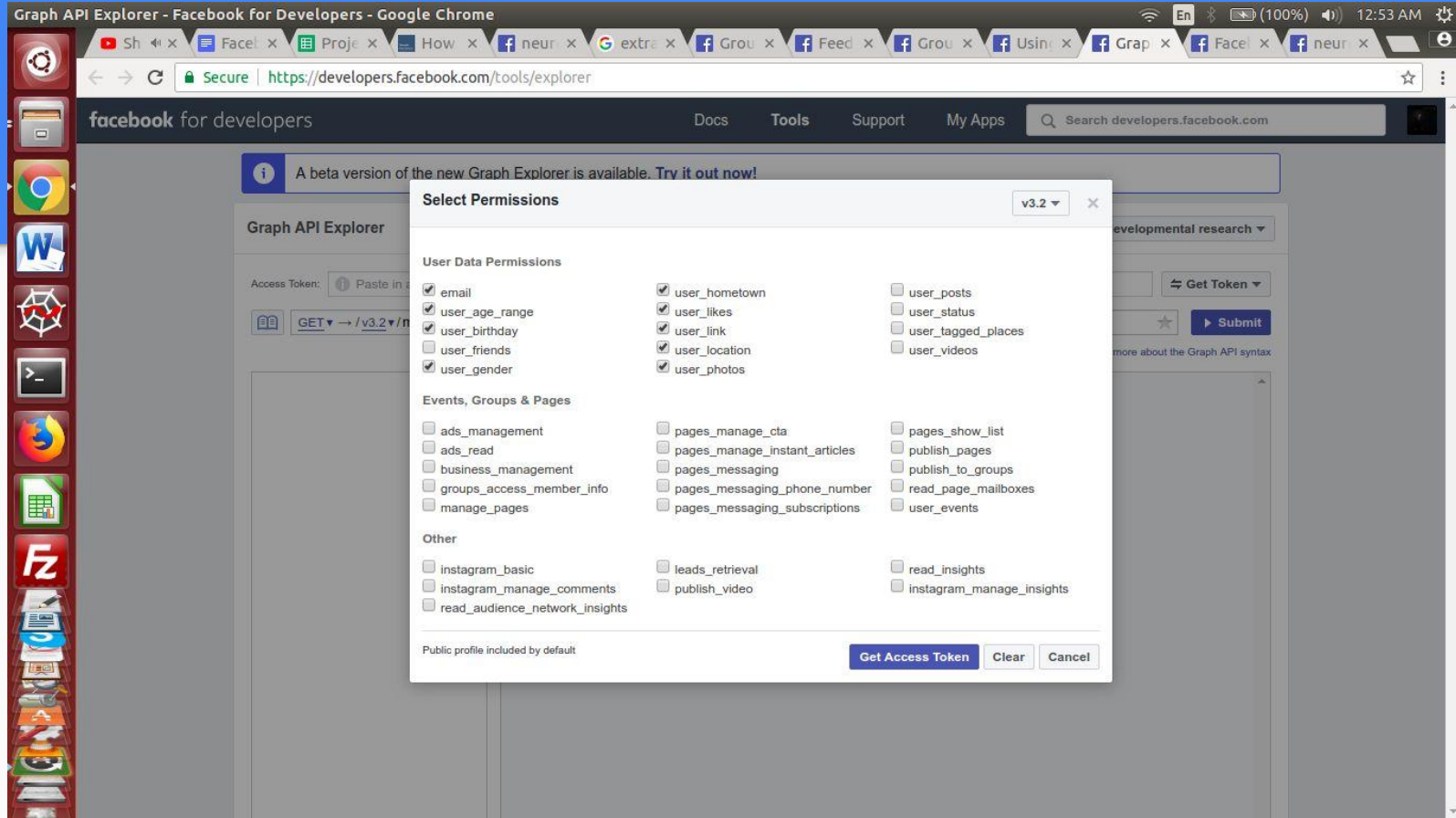
Step 6 of 10

The screenshot shows the Facebook for developers interface. At the top, there's a navigation bar with 'facebook for developers', 'Docs', 'Tools', 'Support', and 'My Apps'. A search bar is on the right. Below this, a 'Get Token' button is visible. The main content area is partially obscured by a modal dialog titled 'Access Token Info'. This dialog contains a table with the following information:

App	[Redacted]
User	[Redacted]
Valid	True
Expiration Time	[Redacted]
Scopes	user_birthday, user_hometown, user_location, user_likes, user_events, user_photos, user_videos, user_friends, user_status, user_tagged_places, user_posts, user_gender, user_link, user_age_range, email, read_insights, read_audience_network_insights, publish_video, publish_to_groups, groups_access_member_info, leads_retrieval, public_profile, basic_info

At the bottom right of the dialog is a button labeled 'Open in Access Token Tool'.

Step 6 of 10



Step 9 of 10

Access Token Debugger - Facebook for Developers - Google Chrome

Secure | https://developers.facebook.com/tools/debug/accesstoken?access_token=

facebook for developers Products Docs More ▾ My Apps ▾

Access Token Info

App ID	: neuro developmental research
Type	User
App-Scoped User ID Learn More	User last installed this app via API N/A
Issued	Unknown
Expires	1540533600 (in about an hour)
Data Access Expires	1548303644 (in about 3 months)
Valid	True
Origin	Unknown
Scopes	user_hometown, user_location, user_likes, user_events, user_posts, user_gender, user_link, user_age_range, email, publish_to_groups, groups_access_member_info, public_profile

[Extend Access Token](#)

This new long-lived access token will expire on [redacted]

[redacted] [Debug](#)

Documentation and Code for the Developer

Developer information

These slides are intended for a developer (someone with computer programming skills) who may want to update the script either to add new functionalities, or in response to changes to the Facebook API or Facebook Website Structure.

Facebook API and Selenium role in data extraction

- Content of posts/comments/replies, a unique id for each post/comment/reply, their corresponding timestamps is being extracted by the [Facebook Graph API](#). If the Facebook Graph API changes, the script will need to be updated.
- All user ids of each post, comment and reply, Seen count of each post and exact user ids of users who have seen that post (seen count + user ids of users who have seen the post), All reactions along with the user ids of users who reacted is being extracted for each post, comment and reply (reactions + user ids pairs for each post/comment/reply) using Selenium Python library. This script will need to change if Facebook changes the format/Structure of the group pages.

[https://en.wikipedia.org/wiki/Selenium_\(software\)](https://en.wikipedia.org/wiki/Selenium_(software))

<https://selenium-python.readthedocs.io/>

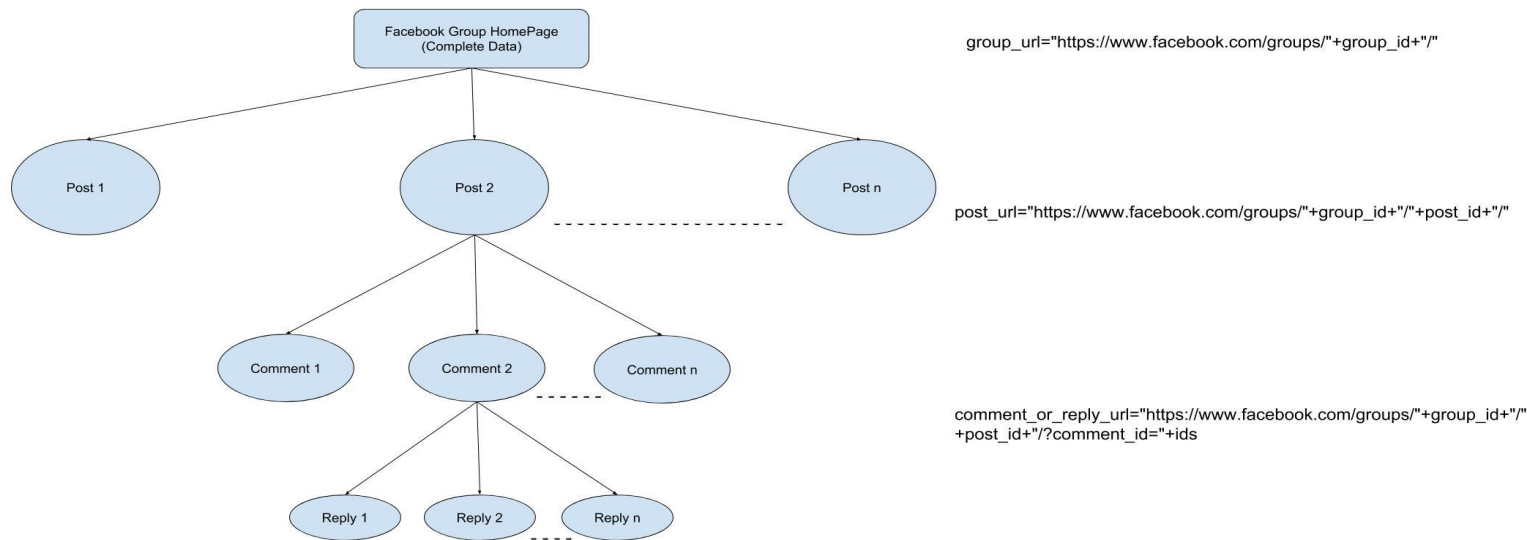
Computer Requirements

- Computer Requirements
 - Python 2.7 needs to be installed on computer to run this script. Ref - <https://www.python.org/download/releases/2.7/>
 - Mozilla Firefox browser needs to be present on computer. Ref- <https://www.mozilla.org/en-CA/firefox/new/>

Script Guidelines

- Input required for script:
 - Facebook Group Admin's "Email address"
 - Facebook Group Admin's "Password"
 - User Access Token (Admin access token. Steps:
 - Group id - An integer which shows up in the homepage URL of group e.g. "https://www.facebook.com/groups/1662317732110904" has group id as "1662317732110904".
- There is an API call limit of 200 calls/hour. The script has been set up to issue 200 calls, stop for a while, and resume to bypass this limit.
- The scripts will be posted on Github, with documentation.

Facebook Group Sitemap Structure Analysis



Algorithm/Code Highlights

- Functions `extract_posts()`, `extract_comments()` and `extract_replies()` have been written to extract all data using Facebook API. These functions call each other to extract -
 - Combined content of posts/comments/replies.
 - A unique id for each post/comment/reply.
 - Their corresponding timestamps.
- An HTTPS GET request is sent to the Facebook for this data extraction.
- Function `user_values()` is the core function that extract all other data which is not feasible with Facebook API. It uses Selenium to extract -
 - All user ids of each post, comment and reply.

Algorithm/Code Highlights

- Seen count of each post and exact user ids of users who have seen that post (seen count + user ids of users who have seen the post).
- c. All reactions (Like, Love, Haha, Wow, Sad, Angry, any special reactions that come around sometimes like Pride reaction etc.) along with the user ids of users who reacted is being extracted for each post, comment and reply. (reactions + user ids pairs for each post/comment/reply).
- Facebook has a complete Sitemap hierarchy for group, post, comment/reply, post seen list, reactions for post, reactions for comments/replies URLs -
 - `group_url="https://www.facebook.com/groups/"+group_id+"/"`
 - `post_url="https://www.facebook.com/groups/"+group_id+"/"+post_id+"/"`

Algorithm/Code Highlights

- `comment_or_reply_url="https://www.facebook.com/groups/"+group_id+"/"+post_id+"/?comment_id="+ids`
- `seen_url='https://www.facebook.com/ufi/group/seenby/profile/browser/?id='+post_id+'&av='+unique_id`
- `reactions_url='https://www.facebook.com/ufi/reaction/profile/browser/?ft_ent_idenfier='+post_id+'&av='+unique_id`
- `reactions_url='https://www.facebook.com/ufi/reaction/profile/browser/?ft_ent_idenfier='+ids+'&av='+unique_id`
- `"https://www.facebook.com/profile.php?id=" + "any user id"` gives us profile of the particular person/user.

Algorithm/Code Highlights

- These variables names have been used as:
 - group_id: A unique group id.
 - post_id: A unique id for a particular post.
 - comment_id: A unique id for a particular comment or reply.
 - unique_id: A unique id for each Admin which creates the group (which is the user id/profile id of the Admin.)
 - ids: A unique id for a particular comment or reply.
- Each URL is opened with Selenium making it possible to extract data one at a time from each page. This helps in overcoming the hidden "comment/reply/seen user ids/reaction user ids" issue.
- This code also handles edge cases like no reactions, 0 seen counts etc and it also does exception handling if in case browser freezes, it will restart itself and resume from the same place.

Possible Future Improvements (for developers)

- Data Extraction Algorithm can be made to run faster by extracting as much data as possible from each page, instead of opening each page.
- Pros and cons of this approach are - It will make the script open less pages and hence speed up the process. However, we need to keep a track of all the data that we have already captured at each page.
- Data Structures used can be improved on. Eg. Hash Table can be used to speed up the process, instead of list to store unique ids in the above case.