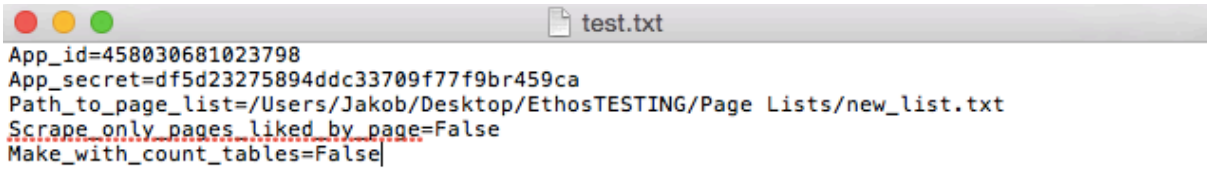
The Ethos Facebook Collector BETA

Setting up a project

- Unzip the package and put “TheEthosFacebookCollector 0.9.1 Beta” wherever you want the program. Avoid special folders like “Applications” or any root folders.
- Run the program one time to get create the folder structure. Open command line or terminal, write “python 3 ”. Then drag and drop the “TheEthosFacebookCollector 0.9.1 Beta.py” into the window. Press enter to run the program. Quit the program.
- The Projects folder contains your projects and the Page lists folder contains the lists of pages you want to scrape. You must tie your project to a list. > Start by going to the Page lists folder. Duplicate the “list_example.txt” file and give it a new name. Fill the new file with all the pages you want to scrape. Write one page per line.
  The page can be denoted either by name or id. The name of a given facebook page can be found by going to the page and copy the name from the URL. For example the name of Ethos Lab is: “EthosITU”.
  Add the pages you want and save the file.
- Go to the projects folder. Duplicate the “project_example.txt“ file. Change the filename of the copy to whatever you want your project to be called.
- Open your custom project file. The following settings must now be set:
  - ***App_id:*** Go to https://developers.facebook.com. Either create a new app or use and old one. Go to your apps page and find your App_id. Copy and paste it into your project file.
  - ***App_secret:*** Find your App_secret where you found your App_id.
  - ***Path_to_page_list:*** This should show the full path to your page_list file, which tells the project which pages to scrape. If the folder was correctly created, you should only have to change the last part of the path. For example you would change “list_example.txt” to “new_list.txt”

- ***Scrape_only_pages_liked_by_pages:*** This should be set to either "True" or "False". If set to True then you will only scrape information regarding the pages that the pages in your page_list have liked. You want get any info on activity like posts and comments. This is good if you have many pages and want to find out how organizations are connected without having to scrape everything.
- ***Make_with_count_tables:*** This should be set to either "True" or "False". If set to True, it automatically creates aggregated tables in your database, which can more quickly provide and overview of your data. It should only be set to True if your likes_info table has less than a million rows. Or else it might take a long time to make.

➢ When you are done, your project file and page_list file should look something like the pictures below (depending on what settings you chose). App_id and App_secret not real:



Scraping data:

- With the project set up, it is time to run the program again. Open command line or terminal, write “python 3 ”. Then drag and drop the “TheEthosFacebookCollector 0.9.1 Beta.py” into the window. Press enter to run the program.
- The first thing you should do is creating a new data collection. You can have as many data collections as you want. And you can always delete them again.
- Follow the instructions of the program.
- Make sure that your data collection gets attached to your newly created project file. Also, choose a short name for your data collection, as you will have to type it in every time you run the program.
- When the data collection has been made you can now go to “use existing data collection”. Follow instructions.
- “Update collection with newest data” will search for activity that has happened on your pages since the last time you scraped. “Collect older data” will collect data back in time. If you haven’t collected any data yet, you can only choose to collect older.

- Follow instructions and you should be collecting your first data!
- If you have many pages in your list, it is recommended to try and only go back 1 week in time on your first run. You can always run the collection again and go further back in time.

You can always add new pages to your page_list, even if you have already made a data collection from your project. Just add them at the bottom, and the collection will make sure that they will be synchronized.
New CSV-files are mirrored from the database every time you push "Make CSV-files". If you update your data collection you should create new CSV-files.

You can also create Gephi-files, but it requires that you install the Networkx python module. To install module go to your terminal and type in "pip3 install networkx".

Creating extended page_list:

- If you want to quickly get a lot of pages, you have the opportunity to create one from the network of one or a few pages.
- After you have created a data collection you can "create extended page list from page likes". This will take the pages that have been liked by the pages in your page_list file and make a new list. This list is created next to your old list in the Page lists folder.
- Important! You can create the extended list by 1 or 2 degrees distance from your original page(s). If you choose *2* degrees you have to be prepared that it might take a very long time.
- You can now create a new project that uses your new page_list file. (If you want to of course.)

The program is still in Beta and some things are still being worked on, so there might be glitches.