Facebook Scraper

Dependencies:

Python (> 3.0) (https://www.python.org)

Getting Facebook Graph API permission:

The scraper needs an access token as an identifier to make Graph API calls. The specific type of access token should be App Access Token. To obtain an App Access Token, a Facebook App must be created first. You only need to do this once unless the conf file is cleared.

1. Visit https://developers.facebook.com.
2. Select create a new app, from the 'App' drop down menu in the navbar. The category of the app won't affect the scraper's behavior.
3. Once a new app is created, a pair of App ID and App Secret can be found in the dashboard page. That's what we need.
4. Open the 'app_id_secret.conf' file in directory 'conf'. Paste the App ID and App Secret in the file in two separate lines, App ID being the first line and App Secret being the second. Make sure there is no extra space or other characters.
5. Save the file and it's all set.

Running the scraper:

1. Linux / Mac OSX: open terminal. Window: open command prompt (`Win-R,` type in `cmd`).
2. Navigate to the scraper's directory, where search.py is located.
3. Run

   `python search.py -q <your search query> -f <the output csv file> -v`

Command line parameters:

There are 3 options: -q, -f and -v.

-q : should be followed by your search query. The query must be inside quotation marks if it contains spaces.

-f : should be followed by a file name as the output csv file. If the file already exists, it will be overwritten.

-v : verbose mode. It's an optional parameter to turn on verbose mode. If -v is set, in addition to saving to the file, every line will also be printed in the command line window to show you the progress. By default, the scraper runs in silent mode.

Press `ctrl-c` if you want to terminate the scraping before it finishes. It might result in incomplete or broken csv file.

Example:

```
python search.py -q bakeries -f bakeries.csv -v
python search.py -q "restaurants in new york" -f restaurants.csv
```

Configuration files:

All conf files are located in the directory conf. It's important not to move them or change the file names.

There are 3 configuration files:

1. app_id_secret.conf
2. access_token.conf
3. fields.conf

app_id_secret.conf contains the app ID and secret of your Facebook app, the first line being the app ID and the second being the secret.

access_token.conf contains the access token assigned by Facebook. If missing, app_id_secret.conf will be used to generate a new access_token. Therefore, at least one of the two files must be present for the scraper to run properly.

fields.conf covers all the fields that will be scraped from Facebook and saved. Each field should be one separate line.

Available fields for fields.conf:

```
id
about
email
category
checkins
description
likes
link
city
country
latitude
longitude
```

```
street
zip
name
phone
public_transit
username
website
founded
```

The order of columns in the final CSV file is also determined by the order in fields.conf.

*Please make sure all the configuration files do not contain extra spaces or empty lines.