

Brief Documentation to Run Code Base

Dependencies

- 1) Install Selenium library. *pip install selenium*
- 2) Install WebDriver binaries. Ensure it matches your chrome version. Refer to the Selenium documentation:
https://www.selenium.dev/documentation/en/selenium_installation/installing_webdriver_binaries/

Other Libraries used: requests, BeautifulSoup, pymongo, and more. See code base.

To scrape the data for a single target individual, from terminal run:

```
python main.py "First_Name" "Last_Name"
```

This will begin the scraping script, looping through all 7 websites scraped for this individual's full name. The scraped data is populated into mongoDB and can be viewed on the front-end at <https://blooming-dusk-68777.herokuapp.com/> when the script is finished.

Important Note: In the *.env* file, I used my personal account usernames and passwords for mongoDB, LinkedIn, Facebook, and API key from Twitter. Since these information are private, I removed the details. If you want to run the script, you will have to input your login details to these sites and a Twitter API key.

To scrape the data for a group of individuals, input the list of individuals into the "people" list in *batch_scrape.py*, then from terminal run:

```
python batch_scrape.py
```

This will begin the scraping script, looping through all 7 websites scraped for all individuals. There are some limitations due to the data broker sites, for example thatsthem only allows 8 requests per day.

The names that have already been scraped, which are in the "people" list in the file *batch_scrape.py*, are already populated into mongoDB. You can search their names at <https://blooming-dusk-68777.herokuapp.com/>, which will show the details of the scraping. Many of their details have already been verified, and you can see their verifications as well.

Note: when packaging for submission, the following are modified:

- Various passwords were redacted
- Node.JS npm modules were removed