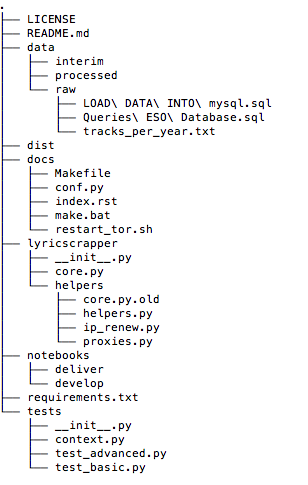
**Guide of Installation of Lyrics Scraping Program**

The module that I uploaded to the repository have the following structure:



To use it please carefully follow these instructions:

1. Clone the repository to your Mac
2. Install these libraries into your Python 3.6 with pip install:
   * re
   * urllib
   * bs4
   * user\_agent
   * pymysql
   * random
   * time
3. Install **tor** in your mac with this command: **brew install tor**
4. Follow the instructions to set it up as a service
5. Install **privoxy** in your mac with this command: **brew install privoxy**
6. Follow the instructions to set it up as a service
7. In the folder **/usr/local/etc/privoxy** edit the “**config”** file and uncomment the line that contains something like **forward-socks5t** in order to look like this (the dot at the end is important, don’t deleted):

**forward-socks5t / 127.0.0.1:9050 .**

1. Use the command crontab –e to create a cronjob with the following line:

**\*/1 \* \* \* \* /georgetown/ds/Music-Mood/docs/restart\_tor.sh**

1. Take into account that the previous script uses the folder **/georgetown/ds/Music-Mood/docs/** please change it to the folder where you clone the repository
2. You can verify if the Tor – Privoxy is working by running the file proxies.py that is included in the **helpers** folder within **lyricscrapper** module

The idea behind a part of the program is protect your external IPs from being blocked. Therefore, we will use Privoxy as a Proxy server to forward the traffic through the circuits of the Tor network. The cron job is restarting the tor service every minute so a new external IP is used.

1. To use the scrapper, change the value of the MAX\_LIMIT and the MIN\_LIMIT and the run the core.py file like: **python core.py**

The program is already connecting to the AWS Server configured by Ivan, and I’ve migrated the table that my program use. The current efficiency is 30%, any ideas or changes are welcome to improve and obtain lyrics from more songs. In the following table I’m proposing the limits for each one of us:

|  |  |  |
| --- | --- | --- |
| **NAME** | **MIN** | **MAX** |
| Ivan Narvaez | 1 | 128895 |
| Katie Sebby | 128896 | 257790 |
| Sebastian Oliva | 257791 | 386685 |
| Marcelo Carrillo | 386686 | 515580 |