Twint 2.1.21 Cheatsheet (ingenieriainformatica.uniovi.es)

An Advanced Twitter Scraping Tool

https://github.com/twintproject/twint

Install the latest version:
```
git clone --depth=1 https://github.com/twintproject/twint.git
cd twint
pip3 install . -r requirements.txt
```

GENERAL USAGE

```
python3 twint [options]
```

NOTES

You don't need a Twitter API Key to use this tool. If is recommended to install the latest version to avoid problems with scraping the Twitter UI

OPTIONS

Typical arguments

-g GEO, --geo GEO: Search for geocoded Tweets.

-h, --help: show this help message and exit

-l LANG, --lang LANG: Search for Tweets in a specific language.

--location: Show user's location (Experimental).

--near NEAR: Near a specified city.

-s SEARCH, --search SEARCH: Search for Tweets containing this word or phrase.

-u USERNAME, --username USERNAME: User's Tweets you want to scrape.

EXAMPLES

To get only follower usernames/following usernames
```
twint -u username --followers
twint -u username --following
```
To get user info of followers/following users
```
twint -u username --followers --user-full
twint -u username --following --user-full
```
Userlist: To get only user info of user
```
twint -u username --user-full
```
To get user info of users from a userlist
```
twint --userlist inputlist --user-full
```
Tweet translation (experimental): To get 100 english tweets and translate them to italian
```
twint -u noneprivacy --csv --output none.csv --lang en --translate --translate-dest it --limit 100
```

Miscellaneous examples
```
twint -u username - Scrape all the Tweets of a user (doesn't include retweets but includes replies).
twint -u username -s pineapple - Scrape all Tweets from the user's timeline containing pineapple.
twint -s pineapple - Collect every Tweet containing pineapple from everyone's Tweets.
twint -u username --year 2014 - Collect Tweets that were tweeted before 2014.
twint -u username --since "2015-12-20 20:30:15" - Collect Tweets that were tweeted since 2015-12-20 20:30:15.
twint -u username --since 2015-12-20 - Collect Tweets that were tweeted since 2015-12-20 00:00:00.
twint -u username -o file.txt - Scrape Tweets and save to file.txt.
twint -u username -o file.csv --csv - Scrape Tweets and save as a csv file.
twint -u username --email --phone - Show Tweets that might have phone numbers or email addresses.
twint -s "Donald Trump" --verified - Display Tweets by verified users that Tweeted about Donald Trump.
```

Tweet filtering

--email: Filter Tweets that might have email addresses

-fr, --filter-retweets: Exclude retweets from the results.

--images: Display only Tweets with images.

--links LINKS: Include or exclude tweets containing one o more links. If not specified you will get both tweets that might contain links or not.

--media: Display Tweets with only images or videos.

--members-list MEMBERS_LIST: Filter the tweets sent by users in a given list.

--min-likes MIN_LIKES: Filter the tweets by minimum number of likes.

--min-replies MIN_REPLIES: Filter the tweets by minimum number of replies.

--min-retweets MIN_RETWEETS: Filter the tweets by minimum number of retweets.

--phone: Filter Tweets that might have phone numbers

--since DATE: Filter Tweets sent since date (Example: "2017-12-27 20:30:15" or 2017-12-27).

--source SOURCE: Filter the tweets for specific source client.

--until DATE: Filter Tweets sent until date (Example: "2017-12-27 20:30:15" or 2017-12-27).

--verified: Display Tweets only from verified users (Use with -s).

--year YEAR: Filter Tweets before specified year.

Search modifiers

--all USERNAME: Search all Tweets associated with a user.

-cq CUSTOM_QUERY, --custom-query CUSTOM_QUERY: Custom search query.

--favorites: Scrape Tweets a user has liked.

--followers: Scrape a person's followers.

--following: Scrape a person's follows

-nr, --native-retweets: Filter the results for retweets only.

--profile-full: Slow, but effective method of collecting a user's Tweets and RT.

-pt, --popular-tweets: Scrape popular tweets instead of recent ones.

--replies: Display replies to a subject.

--resume TWEET_ID: Resume from Tweet ID.

--retweets: Include user's Retweets (Warning: limited).

--to USERNAME: Search Tweets to a user.

--user-full: Collect all user information (Use with followers or following only).

--userlist USERLIST: Userlist from list or file.

--videos: Display only Tweets with videos.

Output Options

--cashtags: Output cashtags in seperate column.

--count: Display number of Tweets scraped at the end of session.

--csv: Write as .csv file.

-db DATABASE, --database DATABASE: Store Tweets in a sqlite3 database.

--debug: Store information in debug logs

-es ELASTICSEARCH, --elasticsearch ELASTICSEARCH: Index to Elasticsearch.

--essid [ESSID]: Elasticsearch Session ID, use this to differentiate scraping sessions.

```
twint -g="48.880048,2.385939,1km" -o file.csv --csv - Scrape Tweets from a radius of 1km around a place
in Paris and export them to a csv file.
twint -u username -es localhost:9200 - Output Tweets to Elasticsearch
twint -u username -o file.json --json - Scrape Tweets and save as a json file.
twint -u username --database tweets.db - Save Tweets to a SQLite database.
twint -u username --followers - Scrape a Twitter user's followers.
twint -u username --following - Scrape who a Twitter user follows.
twint -u username --favorites - Collect all the Tweets a user has favorited (gathers ~3200 tweet).
twint -u username --following --user-full - Collect full user information a person follows
twint -u username --timeline - Use an effective method to gather Tweets from a user's profile (Gathers
~3200 Tweets, including retweets & replies).
twint -u username --retweets - Use a quick method to gather the last 900 Tweets (that includes retweets)
from a user's profile.
twint -u username --resume resume_file.txt - Resume a search starting from the last saved scroll-id.
```

Other Options

--backoff-exponent BACKOFF_EXPONENT: Specify a exponent for the polynomial backoff in case of errors.

-ho, --hide-output: Hide output, no tweets will be displayed.

--min-wait-time MIN_WAIT_TIME: specifiy a minimum wait time in case of scraping limit error. This value will be adjusted by twint if the value provided does not satisfy the limits constraints

-sc, --skip-certs: Skip certs verification, useful for SSC.

Translation

--translate: Get tweets translated by Google Translate.

--translate-dest TRANSLATE_DEST: Translate tweet to language (ISO2).

by José Manuel Redondo López

--format FORMAT: Custom output format (See wiki for details).

--hashtags: Output hashtags in seperate column.

-if [INDEX_FOLLOW], --index-follow [INDEX_FOLLOW]: Custom Elasticsearch Index name for Follows.

-it [INDEX_TWEETS], --index-tweets [INDEX_TWEETS]: Custom Elasticsearch Index name for Tweets.

-iu [INDEX_USERS], --index-users [INDEX_USERS]: Custom Elasticsearch Index name for Users.

--json: Write as .json file

--limit LIMIT: Number of Tweets to pull (Increments of 20).

-o OUTPUT, --output OUTPUT: Save output to a file.

--pandas-type [PANDAS_TYPE]: Specify HDF5 or Pickle (HDF5 as default)

-pc PANDAS_CLEAN, --pandas-clean PANDAS_CLEAN: Automatically clean Pandas dataframe at every scrape.

--stats: Show number of replies, retweets, and likes.

--store-pandas STORE_PANDAS: Save Tweets in a DataFrame (Pandas) file.

--userid USERID: Twitter user id.

HTTP Proxies

--proxy-host PROXY_HOST: Proxy hostname or IP.

--proxy-port PROXY_PORT: The port of the proxy server.

--proxy-type PROXY_TYPE: Socks5, HTTP, etc.