WE1S TweetSuite

The WE1S TweetSuite is a set of tools used to collect data from Twitter and prepare it for topic modeling. TweetSuite consists of three components: a scraper to download tweets directly from Twitter's website, a preprocessor to clean the tweet text and prepare it for topic modeling, and an aggregator to combine multiple tweets into single documents based on user-specified criteria.

Scraper: Scraping Twitter can be challenging because Twitter requires setting up a developer account and using its Application Programming Interface (API), which places restrictions on the number of tweets that can be downloaded. The WE1S Twitter scraper avoids these complications by providing an easy-to-use wrapper for the [Python Twint library](), which, according to its website, "doesn't use Twitter's API, allowing you to scrape a user's followers, following, Tweets and more while evading most API limitations." (See example of scraped tweet at right.)

Preprocessor: The WE1S Twitter preprocessor is based on WE1S's standard preprocessing tool with modifications specific to working with Twitter data. Modifications include methods for removing emojis and an expanded list of stop words. More information on the preprocessing methods can be found on our card describing our TweetSuite Preprocessor (S-32).

Aggregator: The WE1S tweet aggregator allows users to combine multiple tweets into a single document based on criteria such as the tweeter's username or the date of the tweet. This can be useful for some kinds of analysis. WE1S has found that combining tweets into larger documents before topic modeling yields more coherent and interpretable topic models. More discussion of this observation can be found on Key Finding Card (TBD).

```
{
    "id": 1148301102321930242,
    "date": "2019-07-08",
    "username": "mshumanities",
    "place": "",
    "tweet": "absolutely! It's free
and open to the public. We hope you
can join us.",
    "mentions": ["melissaagarriga",
"coffeeprosejxn", "ritabrentcomedy",
"imanikhayyam", "millsapscollege"],
    "urls": [],
    "hashtags": [],
    "link":
"https://twitter.com/MSHumanities/st
atus/1148301102321930242",
    "retweets_count": 0,
    "likes_count": 1,
    "name":
"2019-07-08__/1234_something"
}
```

An example tweet captured by WE1S's Twitter scraper (part of WE1S TweetSuite).

Further Information:

* [Twint on GitHub]()

WE1S resources:

* [WE1S TweetSuite on GitHub]()
* WE1S Preprocessor [TBD]