Long Nguyen
Thien Binh Dinh

Documentation

Overview of Functions

- Utils
    - Uses the webdriver object to execute javascript code and get dynamically loaded web content
- Scrape Function
    - Begin scraping the loaded HTML
- Clean Function
    - Clean up unnecessary attributes from the result of scraping
- Tokenizer
    - Begin tokenizing every word
    - Tokenize positive and negative dataset that was manually created
- Results function
    - Parse through each word to determine if each review is a positive or negative

Implementation

We began by creating a Chrome web driver and manually input a Yelp link to the python script. From there, we were able to execute javascript on the web page and load the elements of the web page and parse through it. However, the result of the scrape function resulted in lots of unnecessary information. Therefore, we went on to remove all html tags and only keep the valid words from reviews. Then, we began tokenizing each word from the results. We also create our own text files of words that represent positive and negative sentiment. We tokenized these files as well. After tokenization, we simply parse through the words and count the number of positive and negative words for each review. We take the max(pos, neg) and use that to mark whether a review is positive or negative. In the end result, we output a json that lets us know the number of positive and negative reviews the specified yelp link conainted.

Usage

- Paste in the Yelp link of the restaurant review
- Run Python script
- The program will print of the total number of positive and negative reviews