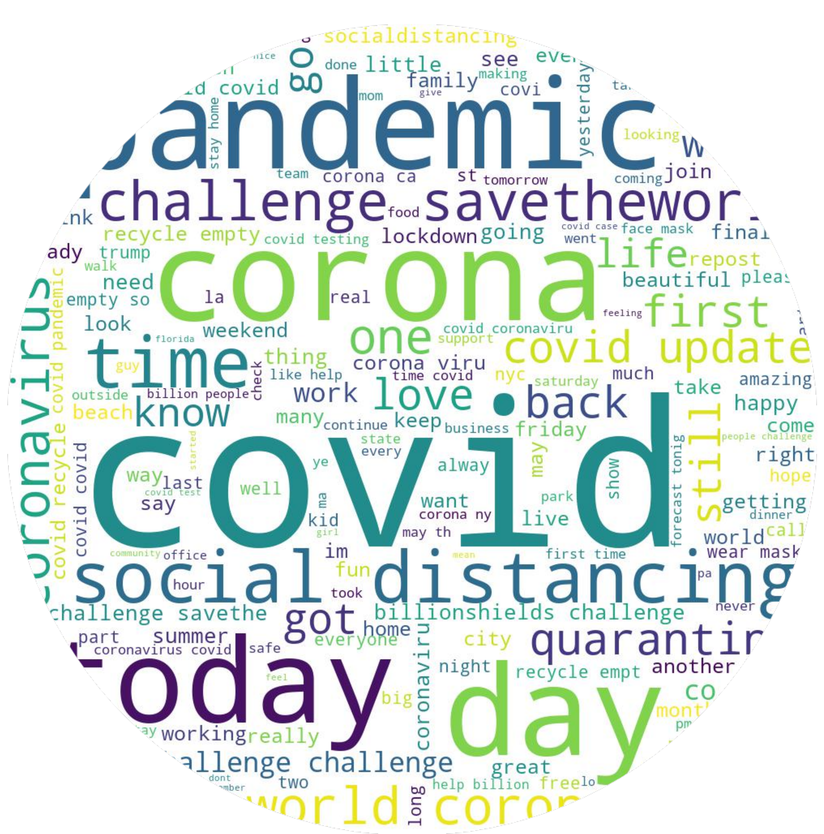
Using Machine Learning to Measure Sentiment During the COVID-19 Pandemic

Truong (Jack) P. Luu | Mentors: Prof. Rosangela Follmann and Prof. James Wolf
School of Information Technology | Illinois State University, Normal, IL 61790

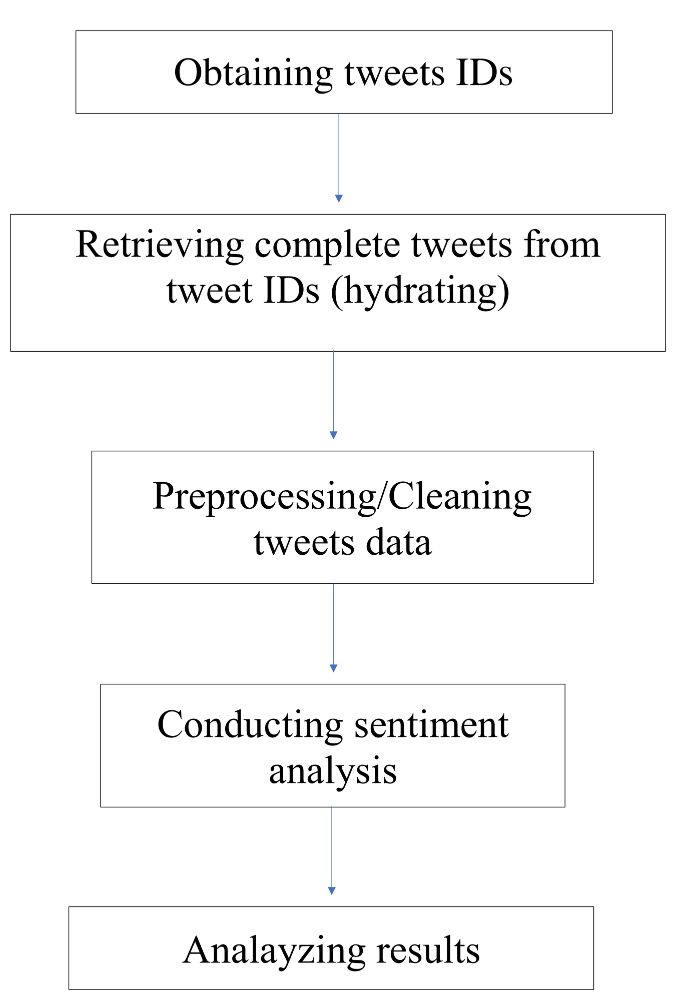
Introduction

During the COVID-19, many people have used Twitter to share their thoughts and viewpoints on various facets of their lives. In this project, we examine COVID-19-related tweets produced in the United States between April and August 2020. We analyze the relationship between user sentiment and COVID-19 cases across the United States and the impact of particular COVID-19 milestones on sentiment scores.



Word cloud of tweet messages.

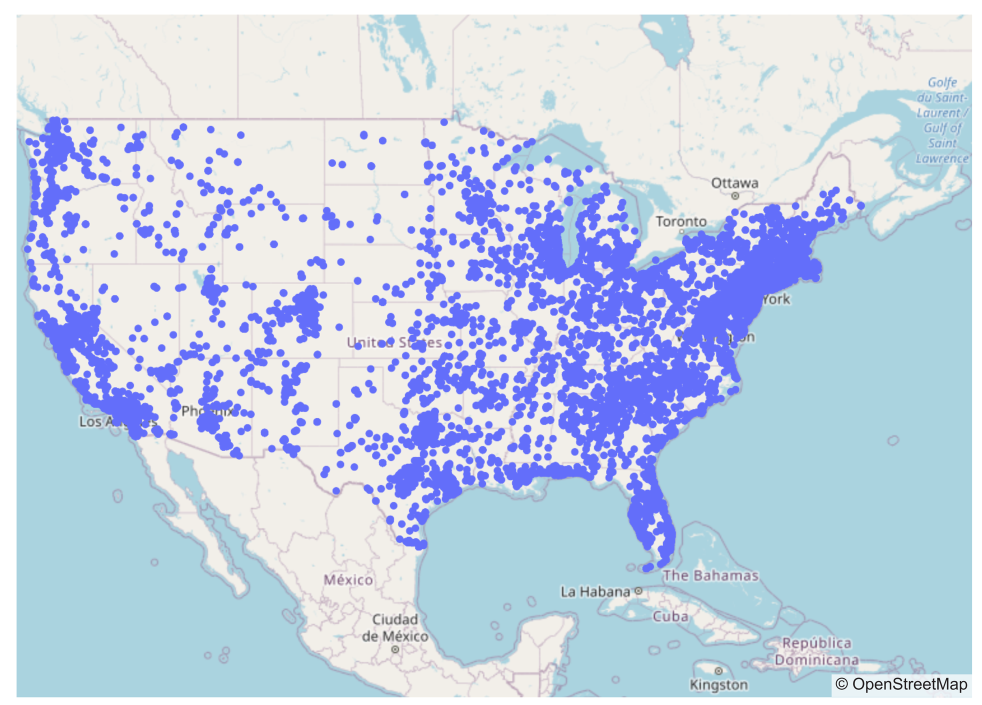
Methodology

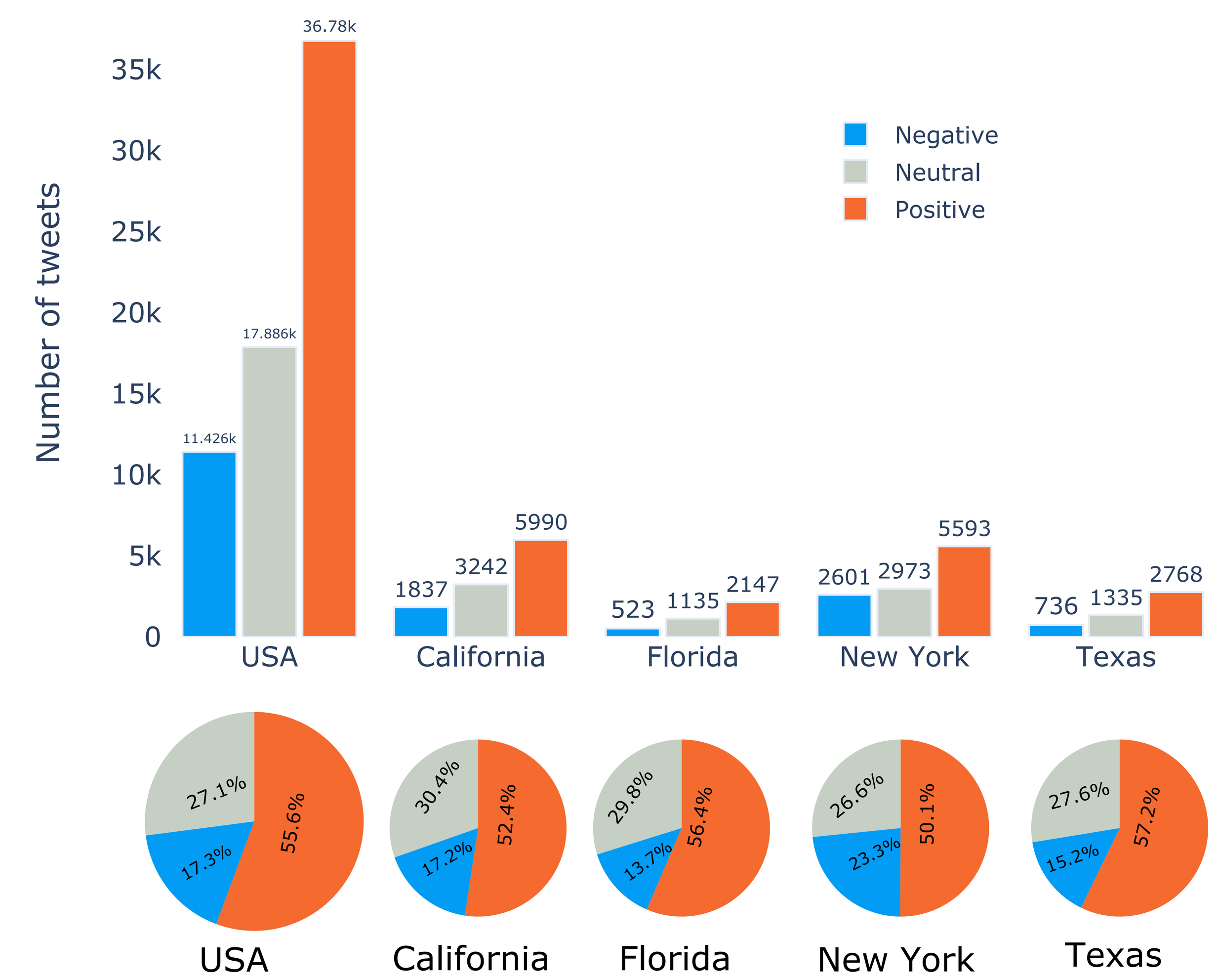


TextBlob - Open source python library for sentiment analysis. It can be used to determine the polarity (positive or negative) of a text along with its subjectivity [1].

Tweets Distribution and Sentiment Score

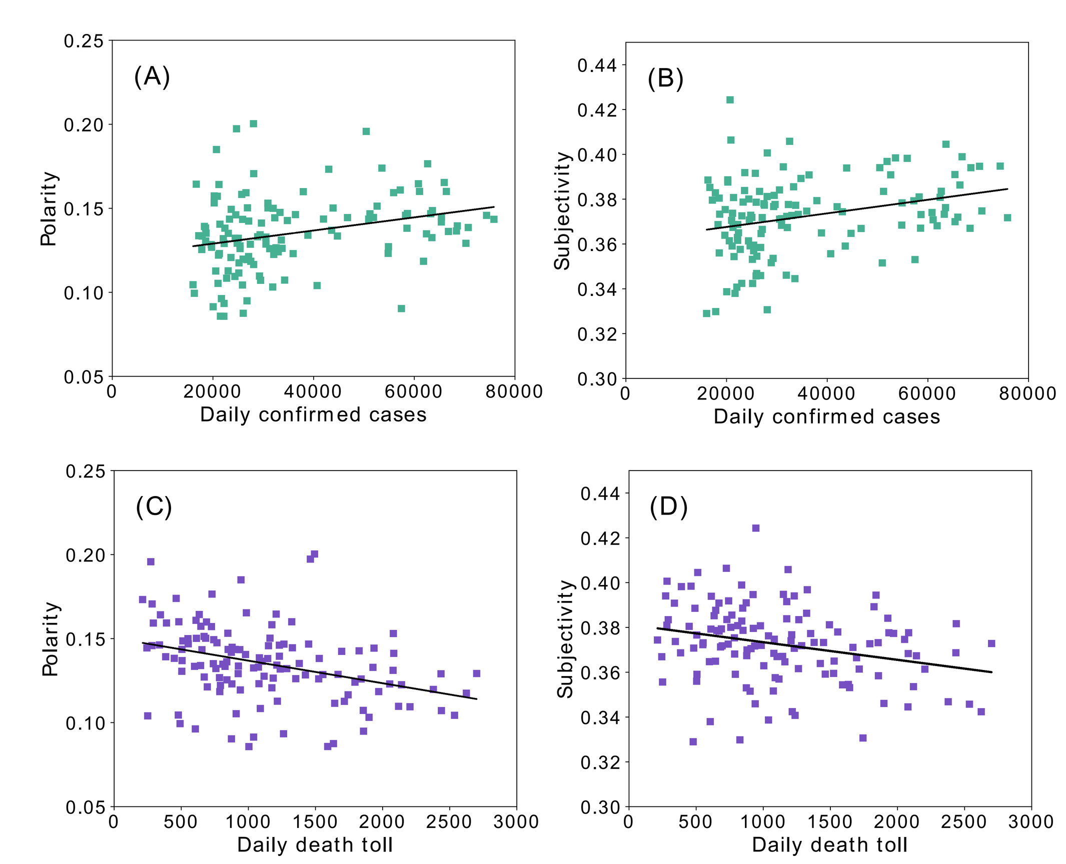
Out of 66,094 geo-tagged tweets in the United States, we discovered that 55% of them expressed a positive sentiment score, 27% expressed a neutral sentiment score, and 17% expressed a negative sentiment score. The top four most populous states California, Florida, New York, and Texas all showed a similar pattern in sentiment distribution, accounting for almost half of COVID-19-related tweets.





Sentiment Score and COVID-19 Cases

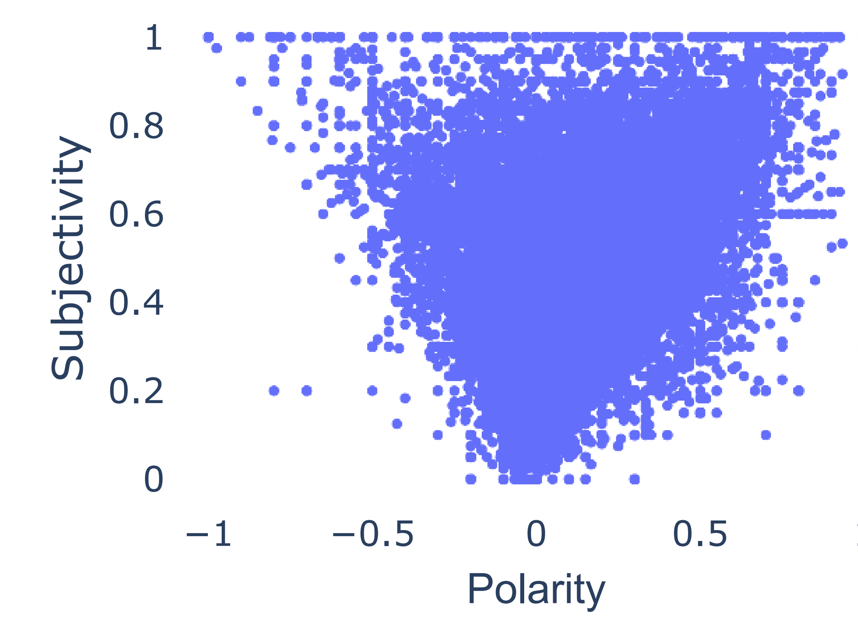
Correlations results indicate that an increase in the number of deaths poses a more threatening challenge compared with an increase in the number of new cases which, even though threatening, still has the door open for a possible recovery.



A-B: r = 0.29, $p < 0.001$ and r = 0.30, $p < 0.0008$, respectively. C-D: r = -0.35, $p < 0.00001$ and r = -0.27, $p < 0.01$, respectively. n = 127

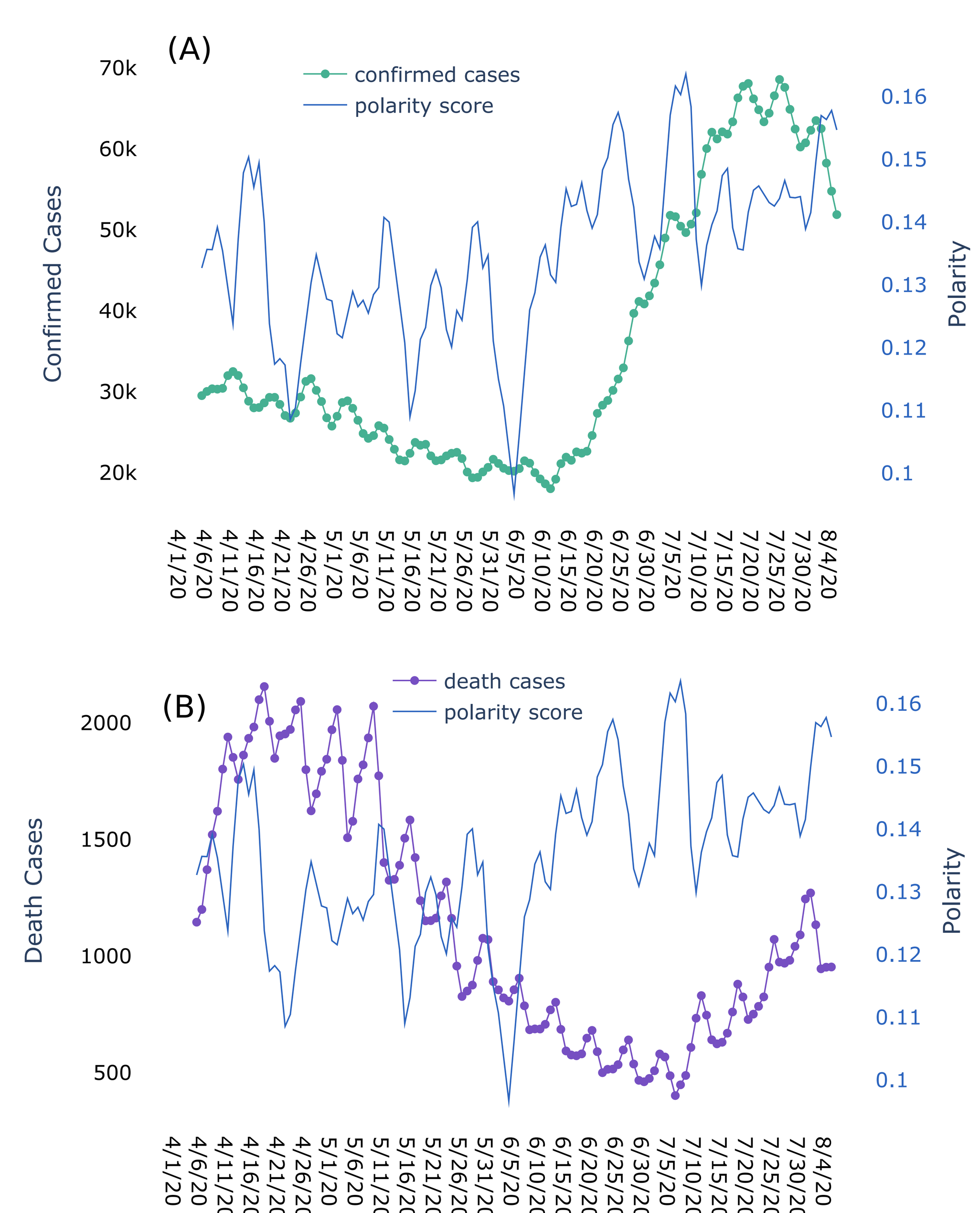
Polarity and Subjectivity

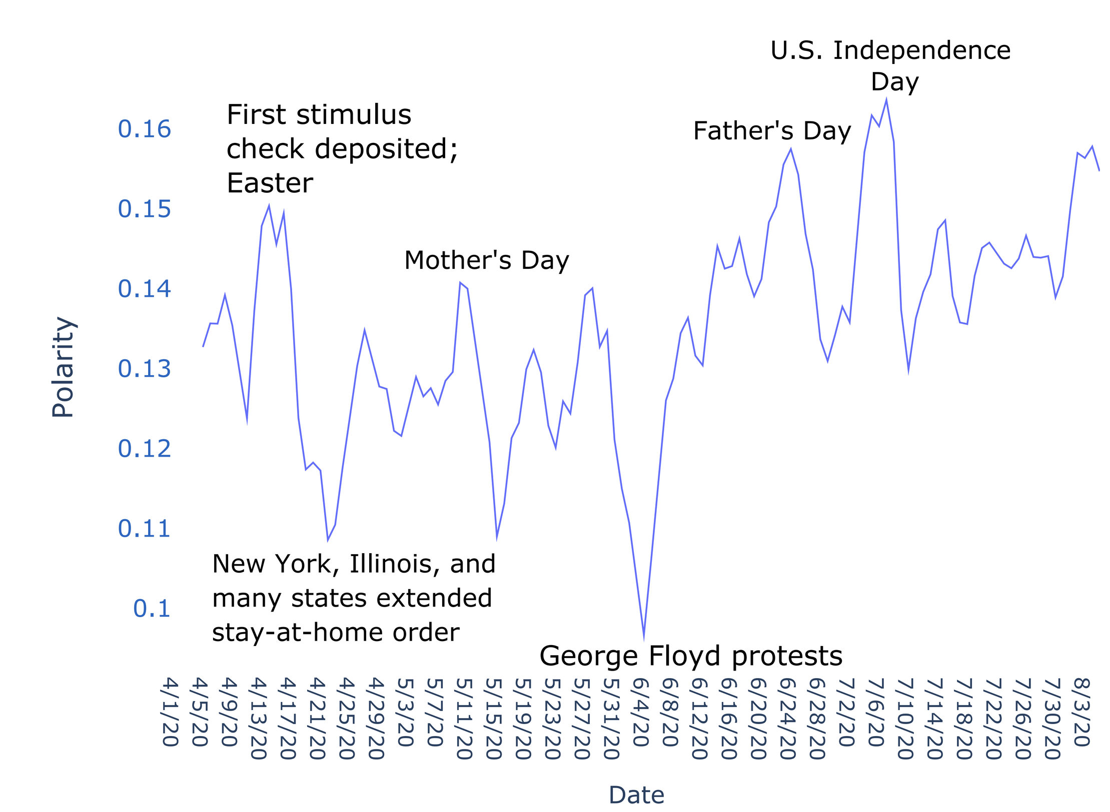
Skewed distribution showing more points towards the positive polarity and higher subjectivity (>0.5), suggesting that the more positive-oriented a tweet is, the more opinion-oriented its meaning will be.



Subjectivity >0.5	Tweet counts	Percent
Polarity >0	15,640	78.3%
Polarity <0	3,612	18.1%
Polarity = 0	732	3.6%

Reasoning Between Polarity Score and the COVID-19 Milestones





Predominantly the oscillations in polarity show a direct connection with the pandemic itself. However, other factors connected or not with the pandemic may have a punctual influence on the polarity score, as illustrated by the selected significant events.

Conclusion and Future Directions

In the United States, there is a connection between sentiment scores, confirmed COVID-19 cases, and death toll. Significant events, such as new legislative laws, major holiday celebrations, and social tensions, may directly impact public sentiment.

Future directions include collecting up to date Twitter data, and develop a sentiment classifier applicable to Emoji.

References and Acknowledgments

[1] Lamsal R. Coronavirus (covid-19) geo-tagged tweets dataset, 2020.
[2] Textblob: Simplified text processing — textblob 0.16.0 documentation.https://textblob.readthedocs.io/en/dev/.

Contact: pluu@ilstu.edu