John Schulz

6/21/2020

DIGHUM 101

**Black Lives Matter in the News: Data Collection and Sentiment Analysis**

Compared to recent years, 2020 has been an extremely interesting year filled with many significant events and societal changes. I couldn’t help but notice how these significant events immediately lead to the publication of many news headlines and articles. Reading and understanding each and every one of these articles is a near-impossible task that no ordinary human could complete. However, with the tools taught in class, I realized that I could create a program to virtually read and interpret the perspectives each article shared. Thus, in an effort to best capture the reality of the news, I narrowed my topic to compare the perspective of specific news events relating to the Black Lives Matter movement. Through my research and thorough examination of thousands of news headlines and summaries, I hope to discover how the news has portrayed the Black Lives Matter and other conflicting movements in the past month.

In this report, I will research into answering the following question: How does the news portray the Black Lives Matter and other counter-movements including All Lives Matter and Blue Lives Matter in the headlines they choose to share and words they use to describe events? Additionally, through my analysis and research process, I hope to better understand if the news encourages societal change and reflects the general public's perspective towards these specific movements. Similar research has been done into understanding the sentiment of the news. Compared to other research of this type, I will focus my study on the last month of news and only the three movements I previously mentioned. This specific topic and timeframe of interest will allow me to spotlight the reality of how the news has shared these three movements that have been extremely active recently.

Before beginning my research, I brainstormed for hours on end on how best to study this unique research question. After generating many complex approaches that didn’t seem feasible given our time constraint, I decided to simplify my solution by breaking down my study method into three parts: data collection, analyses, and visualization. For each of these three parts, I created Python programs on Jupyter Notebook to adequately complete their respective portion of the larger project. When put together, my collection of programs achieve the much more complex level of analyses I initially sought out to do.

For the data collection portion, I had to figure out a way to get up to date news every day since I would be focusing my research on this past month. I realized that I would need to scrape web data from multiple news sources with a program that would automatically run everyday. Ultimately, I decided to scrape Google News, Yahoo News, and Bing News with search queries relating to the three movements I chose. I used Beautiful Soup to identify and separate each aspect of the HTML and a Raspberry Pi with Crontab to automatically run the programs I created. The final result of this data collection is tabulated data updated and backed up on a Google Sheets document featuring tabs for each news source.

After collecting data for a number of days, I needed to revisit my Google Sheets document to analyze the hundreds of articles I had scraped. To gain a better understanding of these articles and the respective movements they discussed, I needed to implement natural language processing in the form of sentiment analysis to quantify the text I had collected. For my sentiment analysis, I decided to use NLTK Vader which we learned in class as well as TextBlob to guarantee my analysis was adequate. Together, these two Python libraries helped prepare the data I had collected to be ready to be analyzed and visualized.

Finally, after sufficiently cleaning and preparing my data, I needed to visualize my data to begin answering my research question. With the help of Matplotlib, I visualized the sentiment values I had generated in relation to the date the articles were published in. Moreover, I compared how often news articles were published for each movement by plotting respective article counts each day. For both plots, I mapped important news events relating to each movement to see if there was any relation to both sentiment and article count when significant events occurred. This collection of plots is evidence to support my analysis and further develop my understanding of how the news portrays the Black Lives Matter movement.

From the analyses and visual representations I have created, I expect to find a correlation between the news articles released and major events including Black Lives Matter protests and police brutality. More specifically, I believe there will be spikes in the number of Black Lives Matter news articles and sentiment values that reflect whether the major event was positive or negative. I expect to make assumptions on how delayed the news publishes articles for specific events, and I will have to speculate on the general public's sentiment and perception of these specific events. These assumptions will not hinder my ability to support my findings and use my best judgment to create an effective claim. Regardless, my research will provide further insight into publicity the Black Lives Matter movement receives.

Extrapolating from my research question, I expect to gain a better understanding of the skills we learned in class and how to approach research problems in the digital humanities. Following the completion of this project, I hope to use my codebase to support further research into topics related to news sentiment and how current events are portrayed online. Ultimately, the research process I have completed and the discoveries I make will help guide further explorations into the digital humanities.