News Companies Topic - Sentiment Analysis

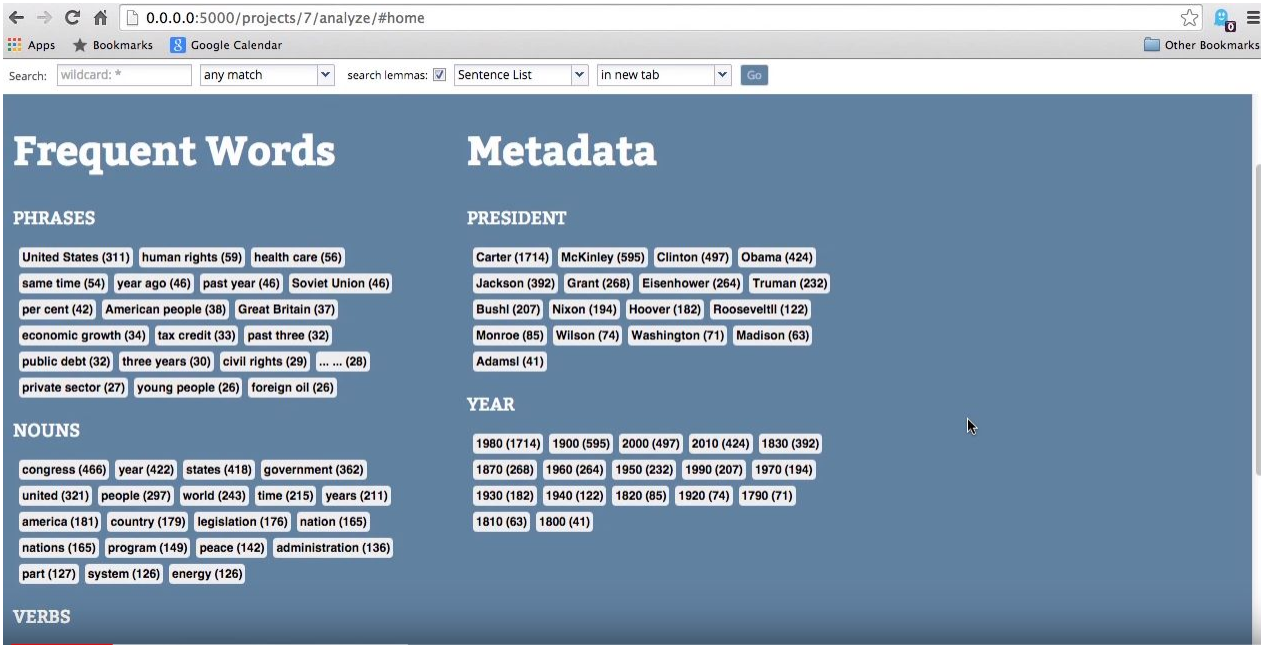
Thesis Idea

Generate topics such as (Brexit, political parties, political leaders) and then graph overtime how sentiment changes to these topics. For example how the sentiment of brexit, popular sports teams and political leaders changes over the span of months or years. Also a useful measurement is how sentiment towards these topics changes based on the news articles. As an example, it may be possible to determine what political parties certain news article support by adding up all these sentiments, or businesses, etc.

After researching these topics the most interesting use case where there would be a lot of useful information to be obtained is the final one. Being able to determine the interests, disinterests and biases of the news companies themselves provides a lot of useful information for the consumer, as well as how this sentiment lines up with other companies and has been altered over time. This can also be used for more general topics rather than political ones. For example, an understanding of how new technologies are perceived by the media is also another use case if say the topic is VAR.

Technologies

WORDSEER



Project Page: http://wordseer.berkeley.edu/
Github: https://github.com/Wordseer/wordseer/
Learned about wordseer in this academic paper:
https://academic.oup.com/dsh/article/30/suppl_1/i130/365257

News Companies Topic - Sentiment Analysis

Hadoop / Amazon Cloud / Google Cloud

There seems to be a need for a large scale data processing system which can be used to:

- Clean raw articles files. The only information needed from each file is the all of the context within the article, the name of the news company, the date it was published and the country of origin. This is the data that will be aggregated. Doing this processing locally may be challenging, whereas with a third party cloud processing provider, it may be more feasible.
- Build the themes using LDA topic modelling within all of the articles.
- Build sentiment values on each mention of topics.
- Correlate and aggregate the information appropriately (Newspaper companies sentiment to a topic over time).

Project which uses hadoop is as processing platform for detecting trends:
https://github.com/datawrangling/trendingtopics

Work online in the domain of LDA Topic analysis

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371067/ - Using LDA and thematic analysis to reflect how the media in China communicates information on the problems caused by third hand smoke.
- https://scholar.princeton.edu/sites/default/files/bstewart/files/nipshpml2014.pdf - Computer assisted content analysis using topic modelling and machine learning and disocobering their impacts on the outcome.
- https://academic.oup.com/dsh/article/30/suppl_1/i130/365257 - Resarch on TOME (Interactive Topic Model and Metadata Visualization) to support exploratory analysis.