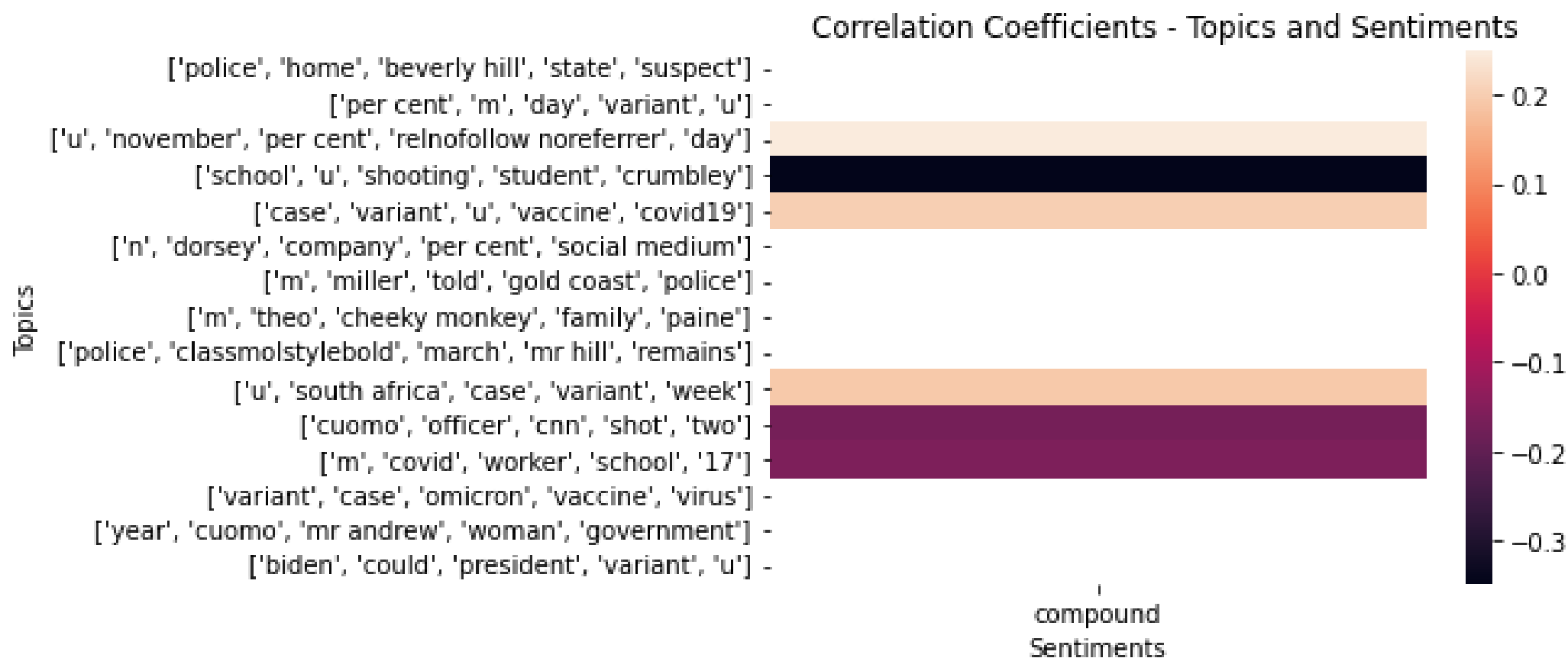
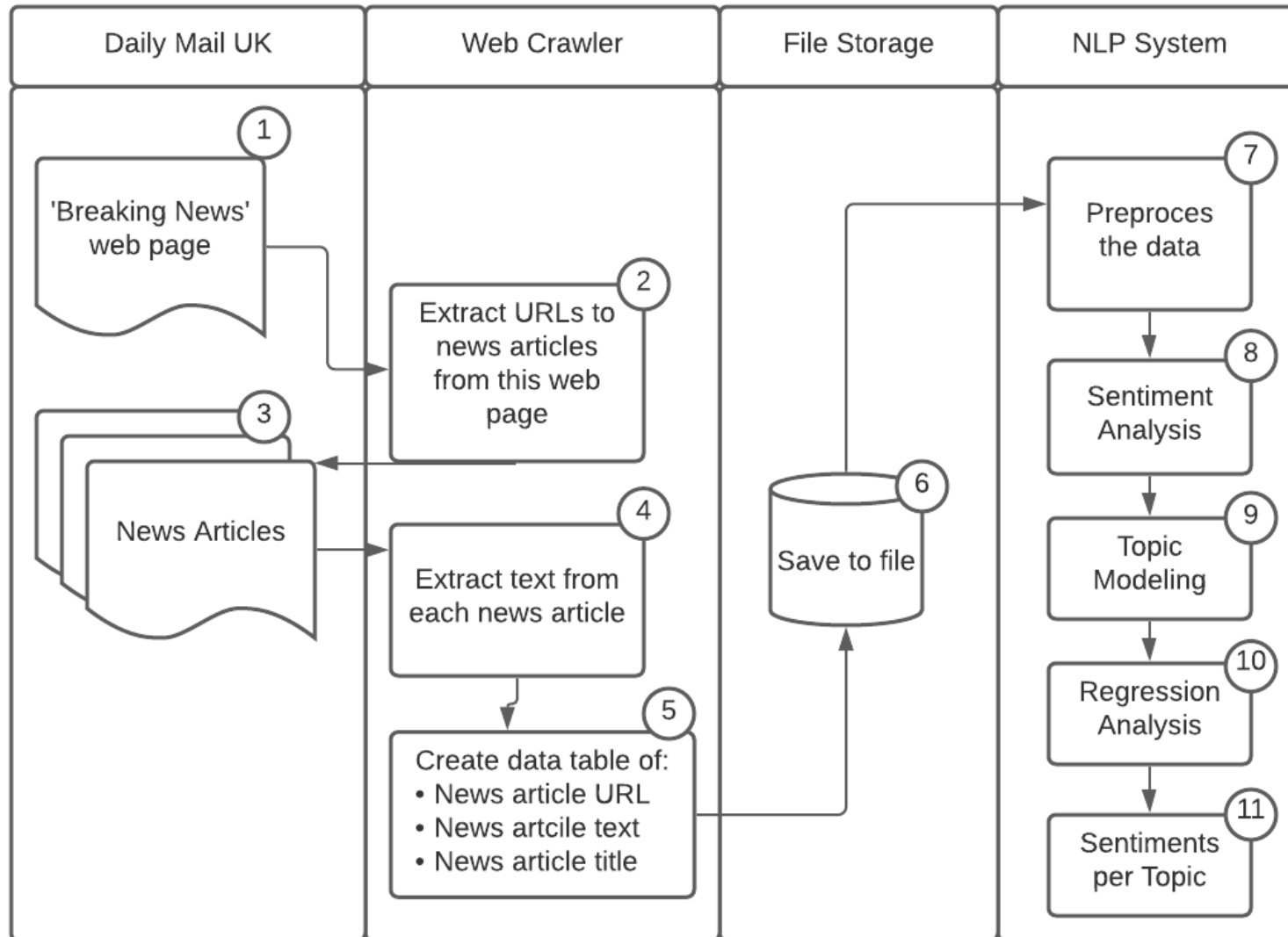


MA5851-A3: Webscraping and NLP Systems

By: Jack Collins



Application Architecture



- Repository:

<https://github.com/JackCollins1991/Topic-Modelling-and-Sentiment-Analysis-on-News-Media>

Limitations

- WebCrawler
 - Some residual html matter which creates noise in the data
 - Only one news outlet in the prototype
- Harvested data
 - Small n
 - Not scalable processing time
- The NLP tasks
 - Some counter-intuitive results