

NEWS BIAS SCORING SYSTEM

LIM CHONG SENG HERMANN (A0195392U)   
YEO WHYE CHUNG NELSON (A0195405A)

KOH SOOK BING (A0195413E)

Table of Contents

[1 EXECUTIVE SUMMARY 3](#_Toc39516187)

[2 PROBLEM OVERVIEW 4](#_Toc39516188)

[2.1 PROBLEM STATEMENT 4](#_Toc39516189)

[2.2 PROJECT OBJECTIVES 5](#_Toc39516190)

[2.3 PRODUCT OVERVIEW 5](#_Toc39516191)

[3 TECHNICAL DISCUSSION 6](#_Toc39516192)

[3.1 INTRODUCTION 6](#_Toc39516193)

[3.2 TECHNICAL SOLUTIONS (SYSTEM ARCHITECTURE) 6](#_Toc39516194)

[3.3 TECHNICAL SOLUTIONS (TECHNOLOGIES) 6](#_Toc39516195)

[3.3 LIMITATIONS 7](#_Toc39516196)

[4 CONCLUSION & REFERENCES 8](#_Toc39516197)

[4 IMPROVEMENTS AND FUTURE ENHANCEMENTS 8](#_Toc39516198)

[5 BIBLIOGRAPHY 9](#_Toc39516199)

[6 APPENDICES 10](#_Toc39516200)

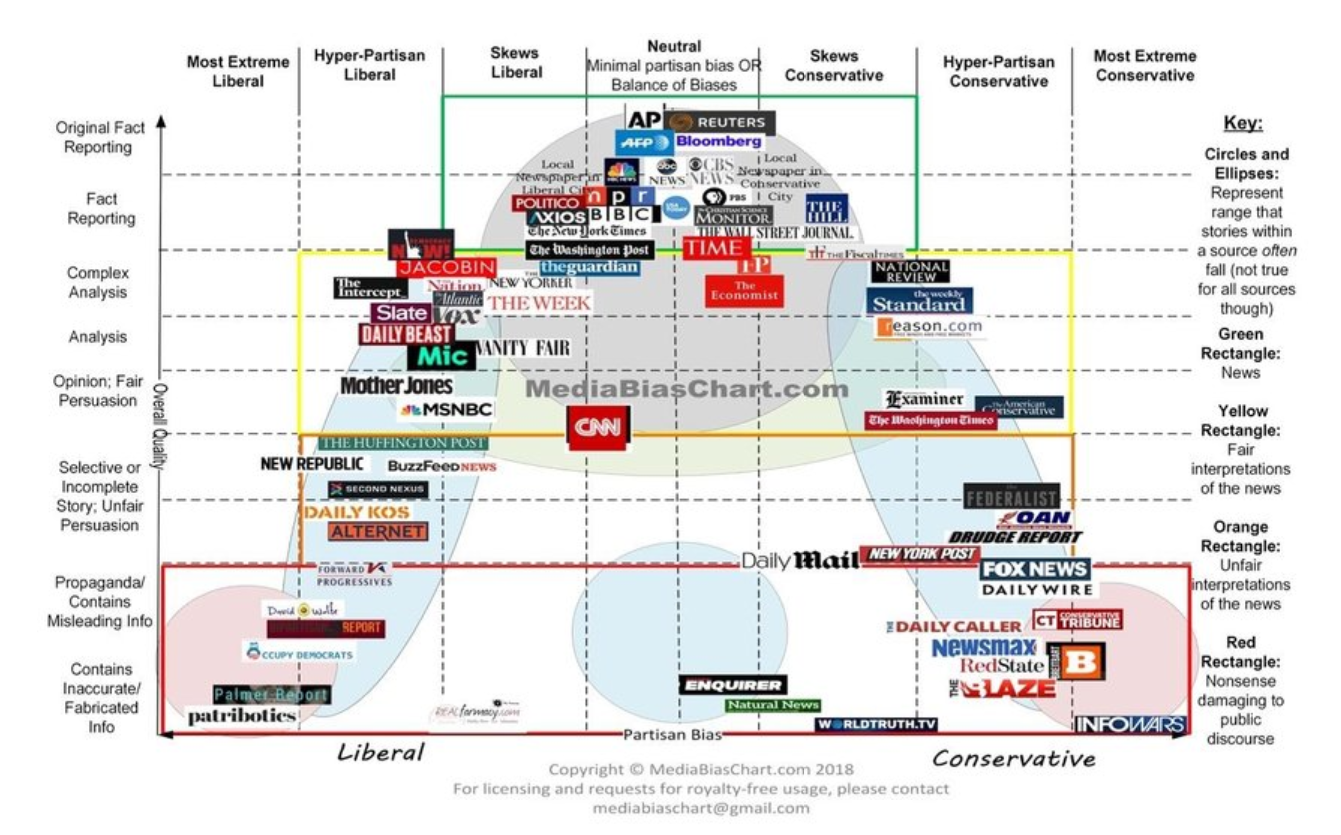
[6.1 INSTALLATION AND USER GUIDE 10](#_Toc39516201)

[6.2 INDIVIDUAL PROJECT REPORT 11](#_Toc39516202)

# EXECUTIVE SUMMARY

30 years ago, evening newscasts on Channel 8 and Channel 5 gave straightforward accounts of the day’s events, and the morning newspapers provide the hardcopy version of the news which told us what happened while we slept. Fast forward to today, we are constantly bombarded by information from multiple social media (Facebook, Twitter, Instagram) which most would read them with a pinch of salt. Instead, we often rely on the mainstream news channels (e.g. CNA, BBC ABC websites) as our source of truth to understand what is really happening around us.

The mainstream news media often agree on the basic facts of a story. But they may still disagree on what people should think about them. This is because while the facts are clear, their implications are murky — hence it can be entirely legitimate for two articles to take utterly opposing views on the same story. A recent media bias chart in 2018 by Pew Research Centre revealed that even mainstream media could be highly biased.



Our project team would like to build a system that could compute the Neutrality Score for the coverage of that story as an additional dimension for the user. Using the techniques imparted to us in lectures, our group first use the RPA tool UIpath to automate the scrapping of news from the mainstream news websites. We then develop an intelligent system using NLP and machine learning techniques to group them into related stories, then further analyse those groups to attempt to produce a metric for how well the set of articles covers multiple perspectives on the story.

Our team had an exciting time working on this project and hope to share our insights with everyone. Sentimental score analysis is more of an art than science and we only wish there was more time to work on the scope and scale of the project.

# PROBLEM OVERVIEW

## PROBLEM STATEMENT

We’ve identified 2 applications which the neutrality score could be applied in local context:

On the fateful evening of April 9th, 2017, United Airlines forcibly removed a passenger from the overbooked flight was filmed by other passengers on their smartphones and posted to the social media platforms immediately. Within hours, it was picked up by news sites and spread like wildfire across the US. News then spread to China and Vietnam, as the passenger was reported to be an American of Chinese-Vietnamese descent and people accused the perpetrators of racial profiling. In China, the incident became the number one trending topic on Weibo, a microblogging site with almost 500 million users. The fiasco was magnified horrifically by the company’s dismissive response. This is exactly the kind of PR catastrophe you can avoid with neutrality score. The PR could use our news neutrality scoring tool to gain insight of the sources of the news and the sentiments attached to them to identify potential PR crisis and work out the immediate action needed.

 Mainstream media report the same events with different sentiment. While there will be readers who are content to have their bias and prejudice reinforced, but there will also be readers who want a more balanced understanding of the world. On 29 June 2016, a single news story in two major publications on the release of the Benghazi panel’s report as follows:



The reader will either come away thinking that the panel’s report was good news for Clinton, or that it was bad news for Clinton depending on which paper was read. The reader could potential rely on the neutrality score of the news as an added dimension on the reliability of the news.

## PROJECT OBJECTIVES

To build a system that could compute the Neutrality Score for the coverage of that story as an additional dimension for the user. By using the techniques imparted to us in lectures, our group will first use RPA tools to automate the scrapping of news from the mainstream news websites. We will then develop an intelligent system using NLP and machine learning techniques to group them into related stories, then further analyse those groups to attempt to produce a metric for how well the set of articles covers multiple perspectives on the story.

## PRODUCT OVERVIEW

* The system to be developed is a web application presenting a summarisation of recent news articles related to the search terms/topics selected by the user.
* The backend will be deployed within a VM. At launch, it will starting scraping all articles from a predefined list of news websites, either through RPA or some other technique to be explored.
* At the frontend web page, the user will select the topic phrase(s) to search, and the time window to look at. For example, news related to "COVID-19" from the past 24 hours.
* The scraped news articles make up the dataset to be searched. Articles from the past 24 hours are filtered, and NLP is then performed to identify articles contextually relevant to the "COVID-19" topic - ideally, relevant articles need not necessarily contain the term "COVID-19" explicitly (i.e. could alternatively mention "pandemic", "COVID virus", "Wuhan virus", "novel coronavirus" etc.).
* Text summarisation is then performed on the identified articles, which is then returned to the user as a singular text passage as output.
* The system will also return a biasness score of the article to the user.

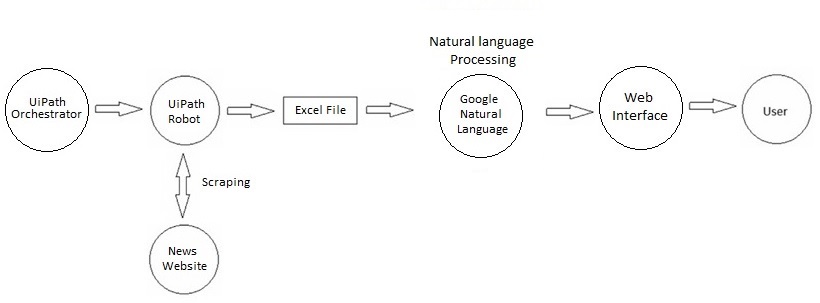
# TECHNICAL DISCUSSION

## INTRODUCTION

To implement our News Bias Scoring System, our team made use of several technologies.

1. UiPath RPA platform
2. Google Natural Language API

## TECHNICAL SOLUTIONS (SYSTEM ARCHITECTURE)



The system consists of several parts. A web scraping component fulfilled by the UiPath RPA platform, Google Natural Language API to provide sentiment analysis and identification of topic phrases.

When the web server is started up, UiPath will be automatically launched to instruct the robots to start scraping on the respective websites on a schedule. The results will be output to an excel file, which the server will then read and process using Natural Language Processing, the results of which will be presented to the user alongside the pulled articles through the web interface.

## 3.3 TECHNICAL SOLUTIONS (TECHNOLOGIES)

1. UiPath RPA Platform
   1. Web Scraping

Using UiPath Studio, we create a robot to do web scraping from a news site, Channel News Asia (CNA). While different news sites have different layout and interfaces, the process of scraping the relevant data is generally similar. The robot will first open the news website, and get the name and URL of the news articles. Next, the robot will open the URL of the individual news articles, and begin scraping for the relevant data, such as the article source, published date, and story content. Finally, the robot will then store the scraped data in to an excel file where it will then be used for processing by natural language tools.

* 1. Robot Management

The UiPath RPA Platform provides Orchestrator management dashboard to manage the RPA Robots.

1. Google Natural Language API

We used Google Natural Language API to provide sentiment analysis and identification of topic phrases, which will help provide the neutrality score.

## LIMITATIONS

1. Due to the limitation on the number of bots being able to run on the Community version of UiPath RPA Platform, for this project we decided to only ran a robot for doing web scraping.

# CONCLUSION & REFERENCES

## IMPROVEMENTS AND FUTURE ENHANCEMENTS

Given that the application is a proof of concept project, the team had brainstormed and are looking forward to incorporating the following enhancements subject to availability of resources:

1. Multiple news websites

The current version of our application provides web scraping on one news website. There is potential to expand this to multiple new websites around the world. This would provide a bigger variety of categories of articles, as well as possibly more variation in calculated biasness scores.

# BIBLIOGRAPHY

<https://towardsdatascience.com/reading-the-markets-machine-learning-versus-the-financial-news-bcd3704f37b8>

<https://towardsdatascience.com/machine-learning-versus-the-news-3b5b479d8e6a>

<https://en.wikipedia.org/wiki/List_of_news_media_APIs>

# APPENDICES

## 6.1 INDIVIDUAL PROJECT REPORT

Lim Chong Seng Hermann

(1) Personal contribution to group project

* Business idea generation
* Web scraping UiPath robot
* Powerpoint presentation
* Project report
* Project video

(2) What you have learnt

This project has given me practice on using multiple intelligent systems to create a product, as well as trying to add value to it.

Technical wise, my main role was to create the web scraping functionality to extract the require data for language processing. So my most significant takeaway was becoming more proficient with the RPA tool UiPath, as well as understanding its strengths and weaknesses better. I also learned more about scraping data from the web, and have a better understanding of how to use several techniques and also potential challengers and constraints, especially when the make-up of each website can differ greatly.

As we are building a whole system, I have also learned how I can apply the different technologies to accomplish the tasks required, such as integating RPA with NLP.

(3) How you can apply the knowledge and skills in other situations

In the areas of business application, being exposed to these technologies will help in the workplace, such as process automation, especially when doing tedious tasks like data extraction. Having knowledge on how to use these tools will help me with the actual implementation of the required tasks.

YEO WHYE CHUNG NELSON

(1) Personal contribution to group project

* Business idea generation
* RPA-NLP system integration

(2) What you have learnt

(3) How you can apply the knowledge and skills in other situations

KOH SOOK BING

(1) Personal contribution to group project

* Business idea generation
* Code generation for NLP
* Project report
* Project video

(2) What you have learnt

By developing the python scripts for NLP, I have a better exposure to the different functions used in NLP (such as lemma, removal of stop words, ngrams, and spacy library) . This has improved my understanding of the different functions and how they all come together to perform NLP. In addition, this is also my first time developing the video where I understand how PowerPoint can be used to make a simple video. I also contributed to the business plan generation where I used the concepts learned in developing a MVP learned during the course.

(3) How you can apply the knowledge and skills in other situations

In the areas of business application, we could use NLP to help to do topic modelling on documents which are too wordy to categorize them. In addition, the tool we have developed could be used to highlight certain key areas of the news that the business unit are keen in and provide timely update.