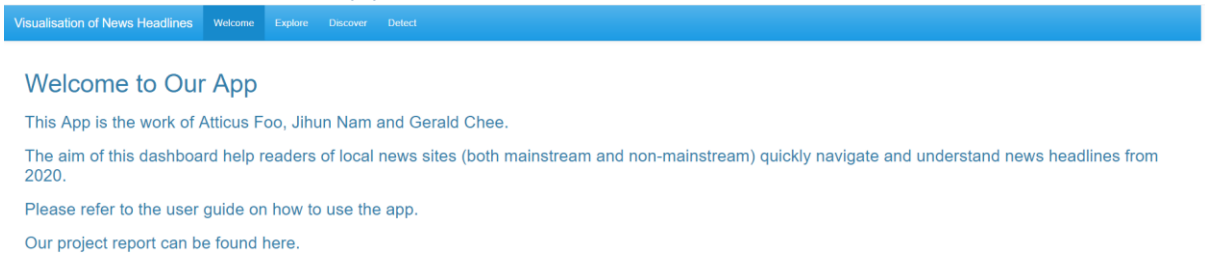
Visualisation of News Headlines

User Guide

Contents

1.Welcome to our App

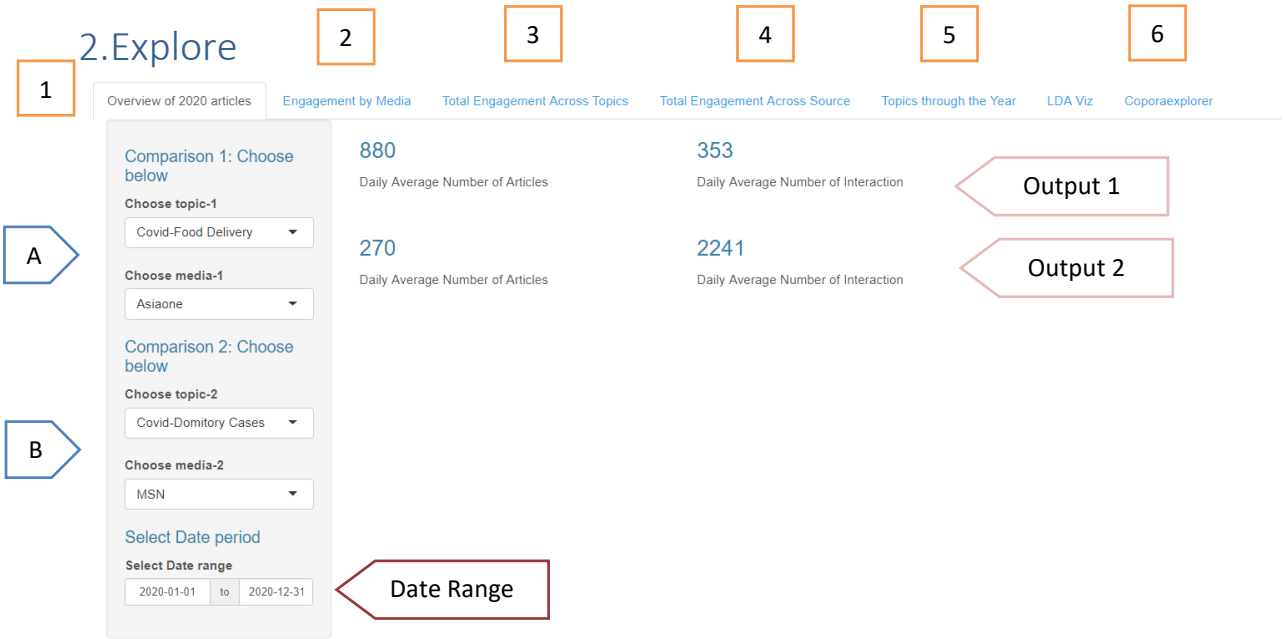


The Welcome page displays a greeting message to the users, lists a brief overview and contains links to our Research paper and User Guide.

The app is mainly divided into three broad sections which will allow the user to perform these functions

- Explore the news corpus by time period, media source, topic, or key word.
- Discover in greater detail the words used in news reporting
- Detect anomalous events and articles based on social media engagement

2.Explore



After clicking on the explore tab, the app will display further tabsets, while the default page is the ***Overview of 2020 articles*** tab.

2.1 Overview of 2020 Articles

This tab displays The Daily average of articles and interactions by the user selected from the sidebar panels option A and B.

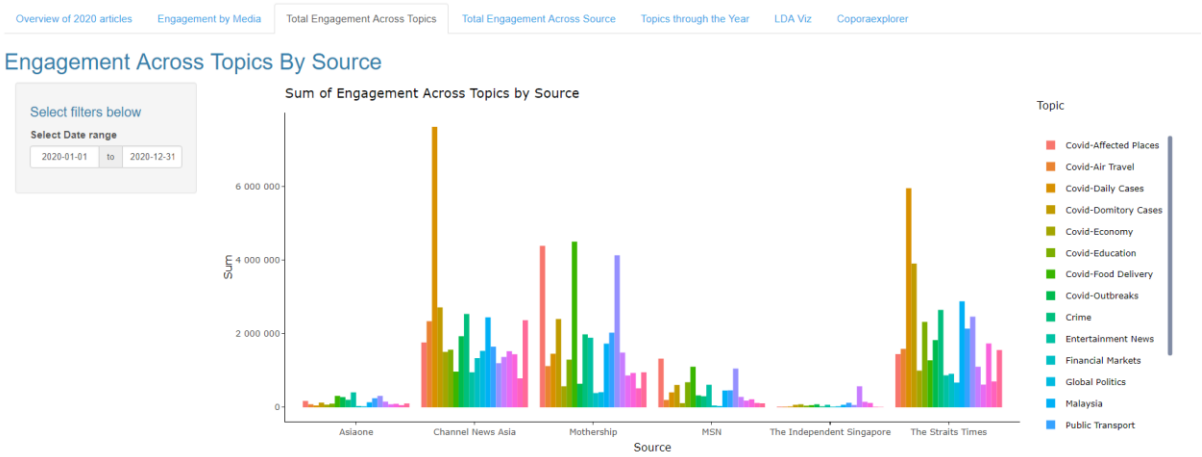
Within Options A and B, the user can determine the topic and media source he or she would like to evaluate.

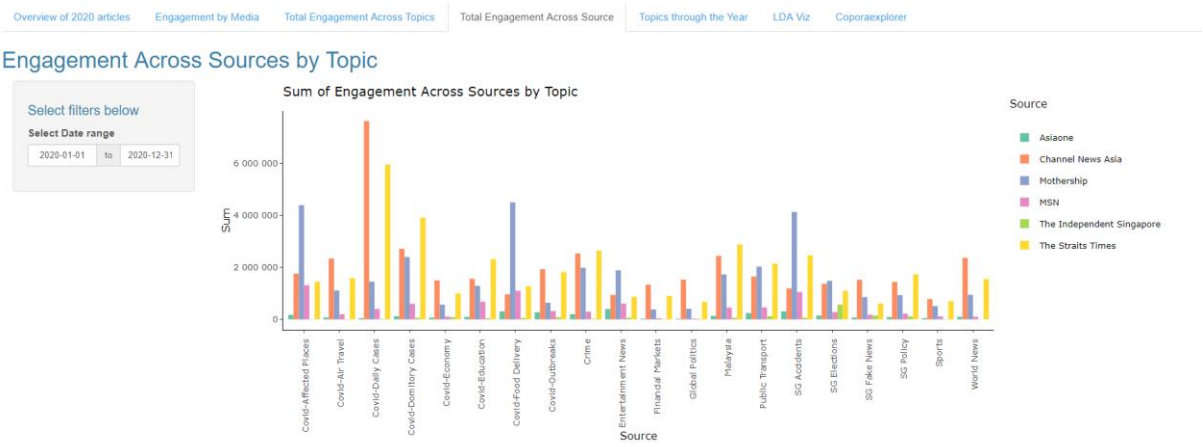
The Date Range filter allows the user to interactively filter the date range for evaluation.

2.3 Total Engagement Across Topics/Source

This function allows readers to visualise how

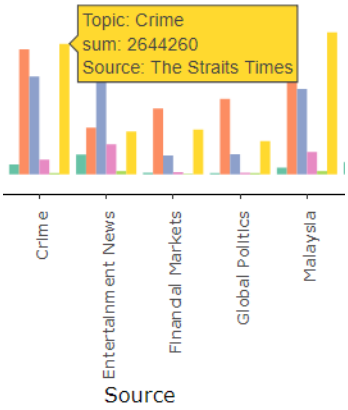
- Different news topics were covered by different media
- Readers of different media sources engaged with topics published





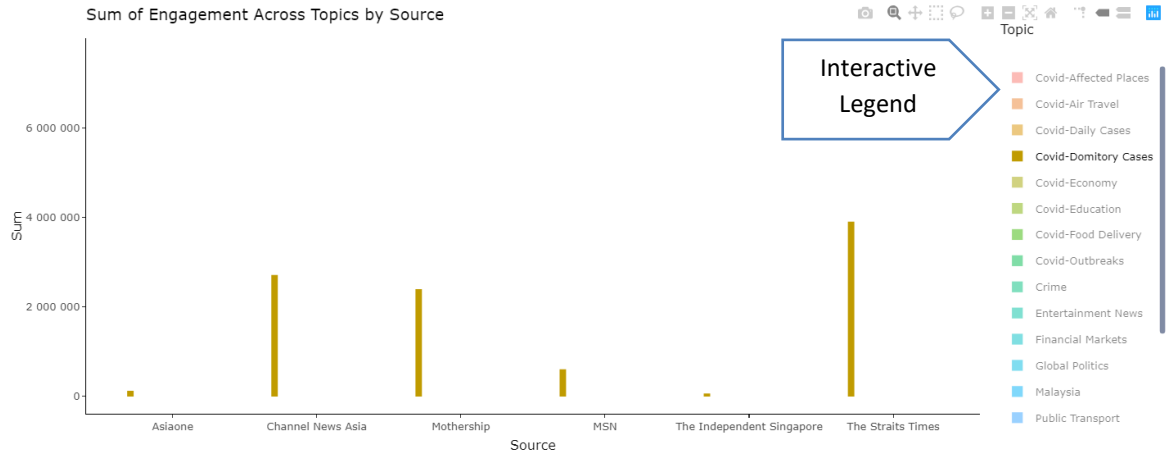
The next 2 tabs displays the sum of total engagements by the 20 topics (***Total Engagement Across Topics***) or 6 local media sources (***Total Engagement Across Sources***).

The Date Range filter allows the user to interactively filter the date range for evaluation.



Because the plot has wrapped by the plotly function, this allows the user to view valuable information on hovering over the plot.

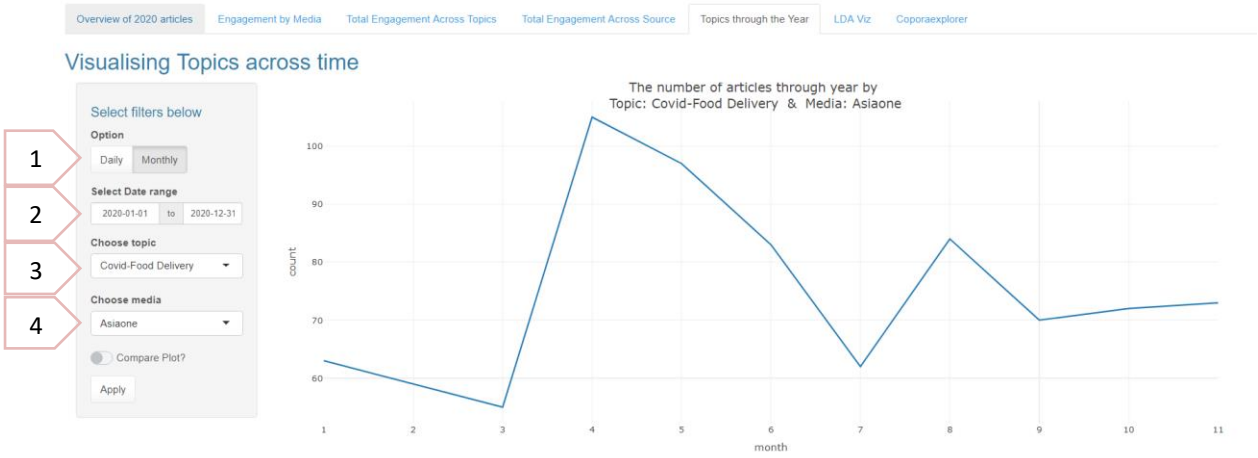
The example shown here is from the ***Total Engagement Across Sources*** tab.



Another useful feature of the plotly output is that it allows for interactive filtering by clicking on the legend. Clicking on the legend will cause the name to be greyed-out and will hide the select bar from the plot for further analysis.

The example shown here is from the ***Total Engagement Across Sources*** tab. This allows us to derive an insight that 'Domitory Cases' Headlines posted by 'Asiaone', 'Must Share News' and 'TISG' garner lesser interactions than 'CNA', 'Mothership' and 'The Straits Times'.

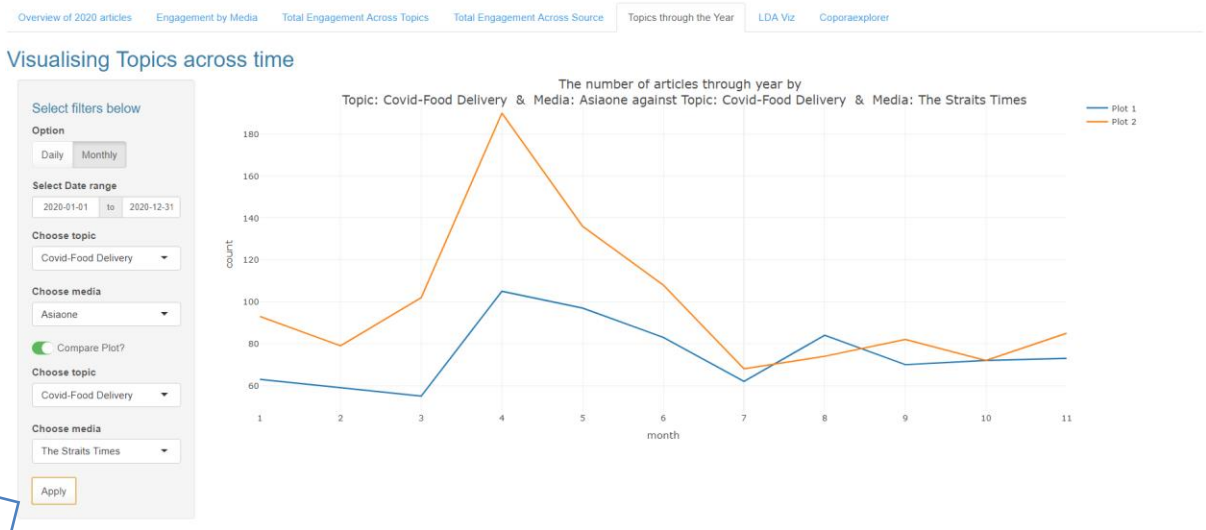
2.4 Visualising Topics across time



The ***Topics through the Year*** tab allows us the user to visualise how a news topic unfolds over time. They can do this by charting the count of articles covered by respective media. They will be able to explore in greater detail by adjusting various filters such as:

1. Daily or Monthly View
2. Date range
3. Topic
4. Media source

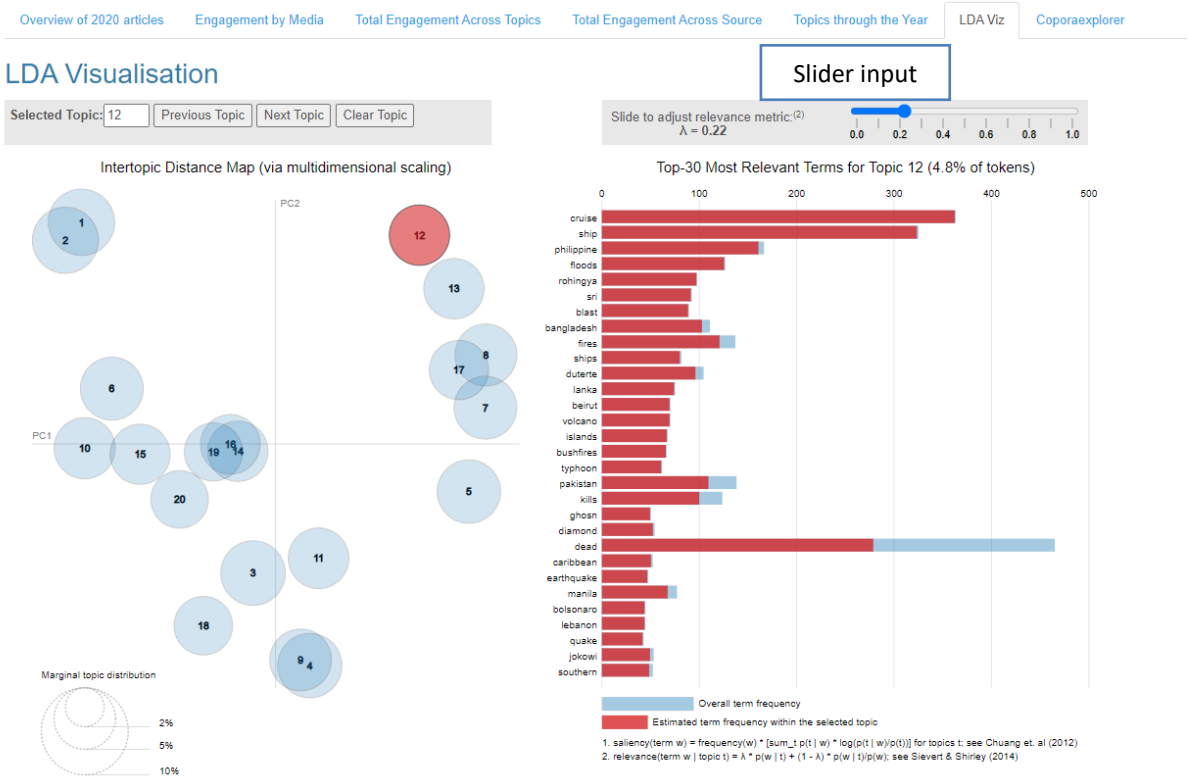
The Sidebar panel also contains a radio button to allow the user to compare plots.



Apply button

To view changes based on the selected options, click on the ***Apply*** button to update the plot.

2.5 LDA Viz



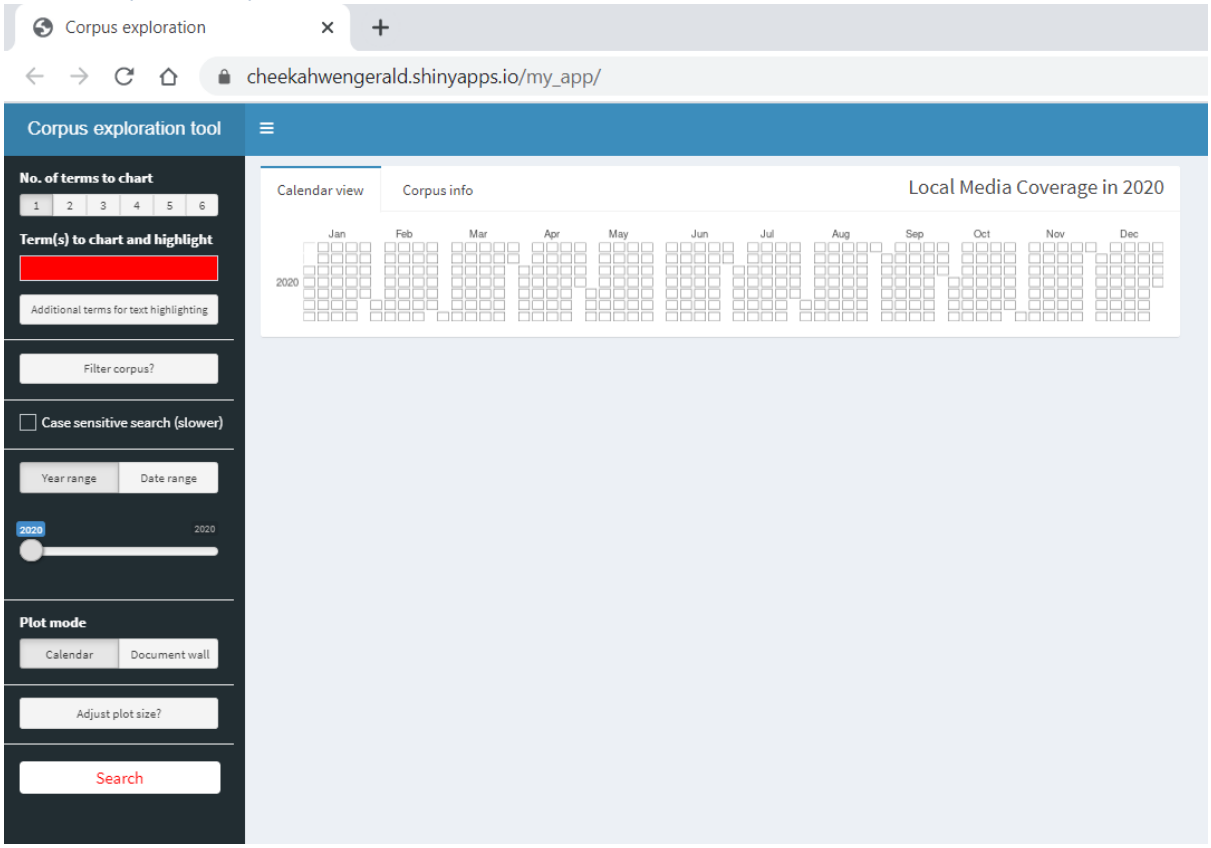
This Tab visualises the topics identified from the LDA model ran through the use of the 'textminer' package and converted to a 'JSON' output.

The plot on the *left* shows the topics clustered by the LDA function based on their posterior probabilities. The closer the bubble plots are, the closer the relation they are to each other. Users can interact with the bubble pot to identify the corresponding word probabilities for each topic.

This can be found in the plot on the *right.* Each topic will shows the top 30 words and it's likelihood of appearing in that topic.

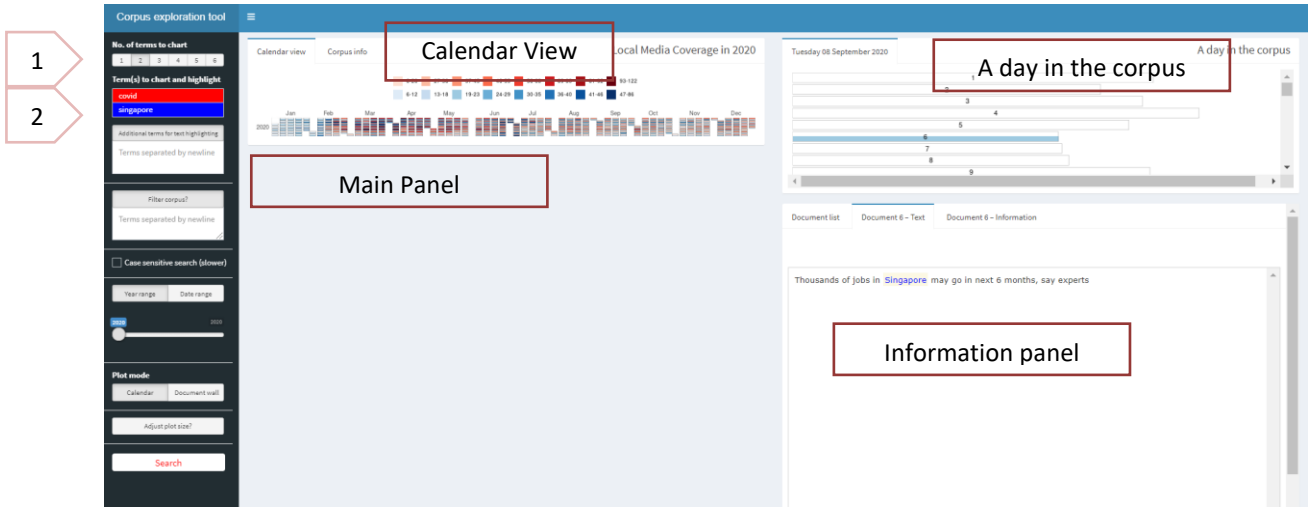
The slider input allows adjustment of the relevance parameter (relevance of a term to a topic that allows users to flexibly rank terms in order of usefulness for interpreting topics).

2.6 Corporaexplorer



The button on the Corporaexplore tab opens a new window to the Corporaexplorer app that was published separately. CorporaExplorer is a very useful function that allows users to view all the documents in the corpus.

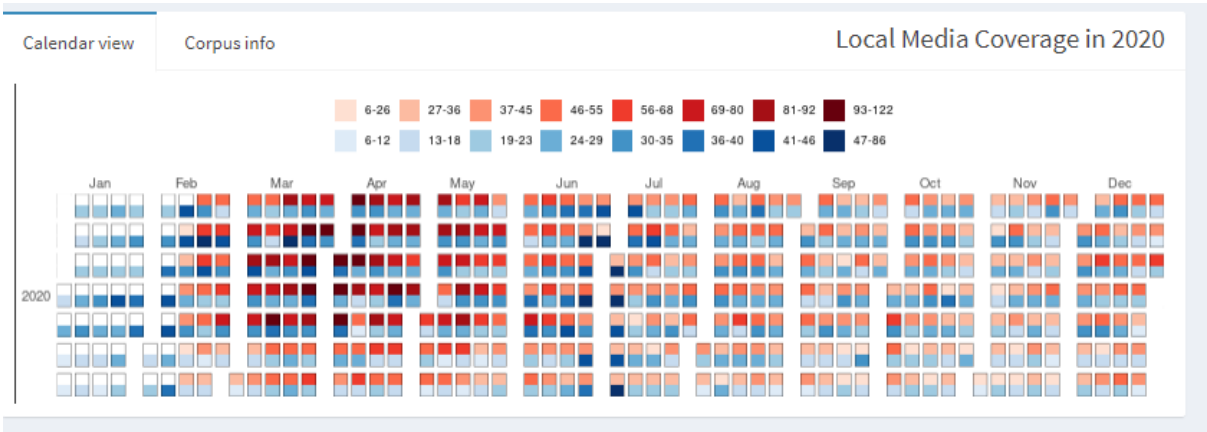
They are able to search based on (up to 5) keywords and retrieve the corresponding meta data for engagement and article link using the Information Panel. Instructions on how to do this will be listed below.



From options 1 and 2, the user can determine the number of words and the words to chart and highlight within the corpus of our dataset.

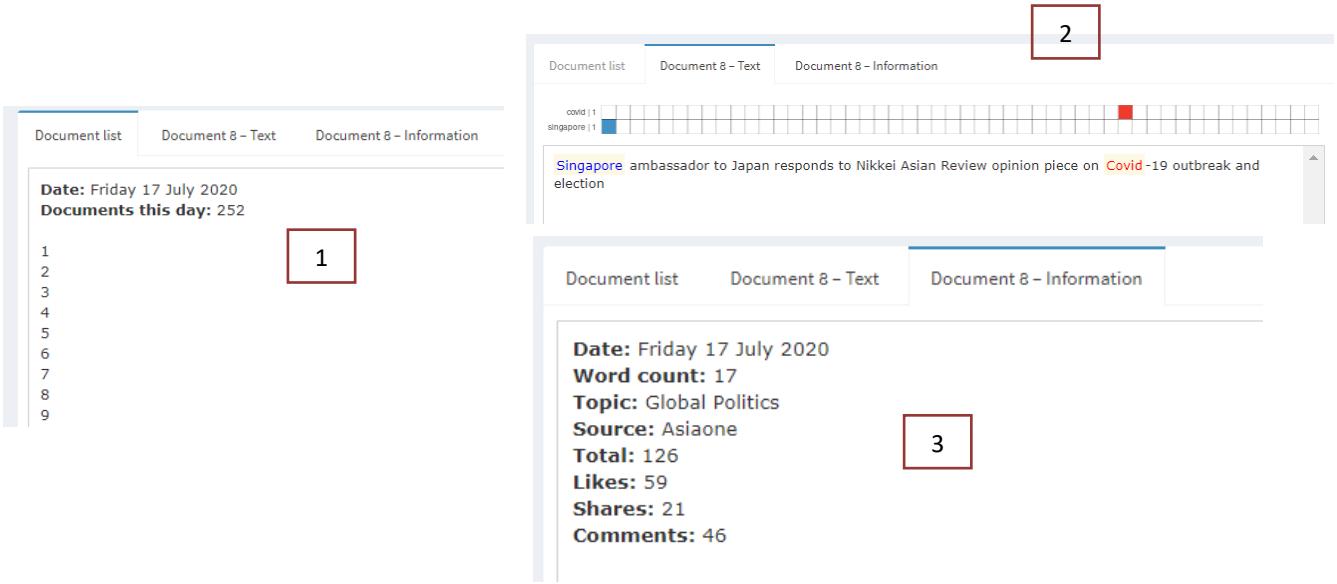
There is a date range filter to filter out dates as determined by the user.

Click on the ***search*** button at the bottom to display the results



The ***Calendar view*** shows a heatmap distribution of the words selected in option 2 by the months of the year. Clicking on the selected day brings out the '***A day in the corpus***' tab which allows the user to see the list of documents on that day and the documents containing the words from option 2.

Selecting the corresponding document with the highlighted color brings out the ***information panel***.

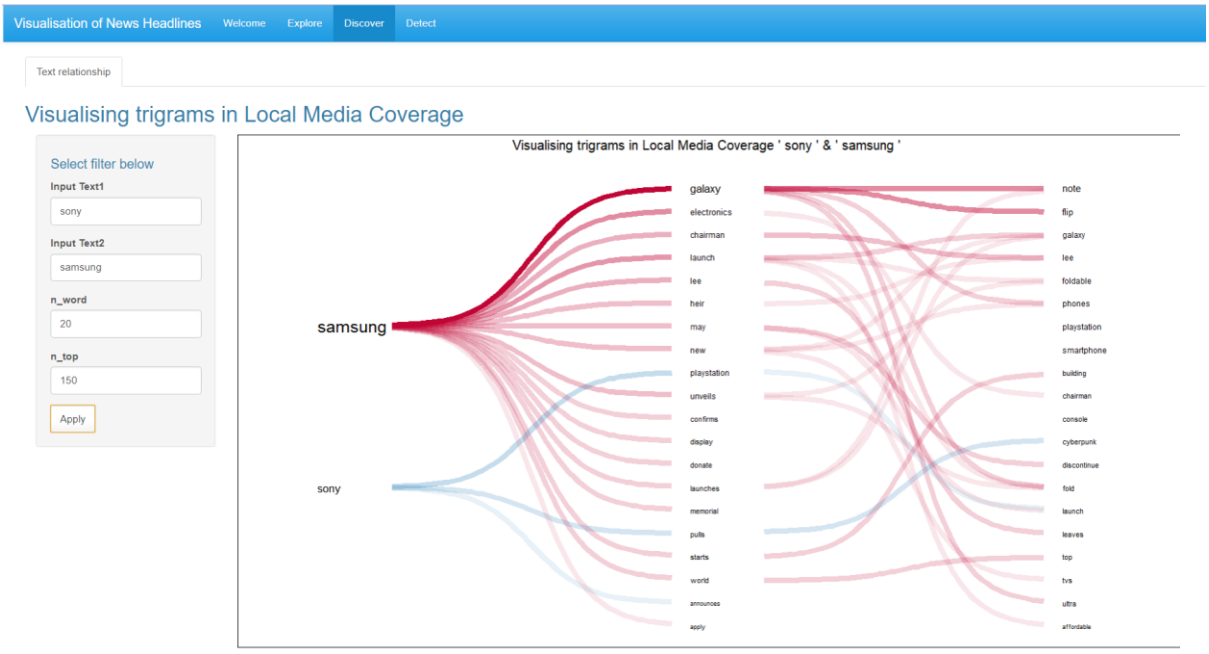


The Information panel shows

1. List of documents posted on that day
2. the text of the document (in the case of our dataset, the Headline)
3. Metadata specific to the document.

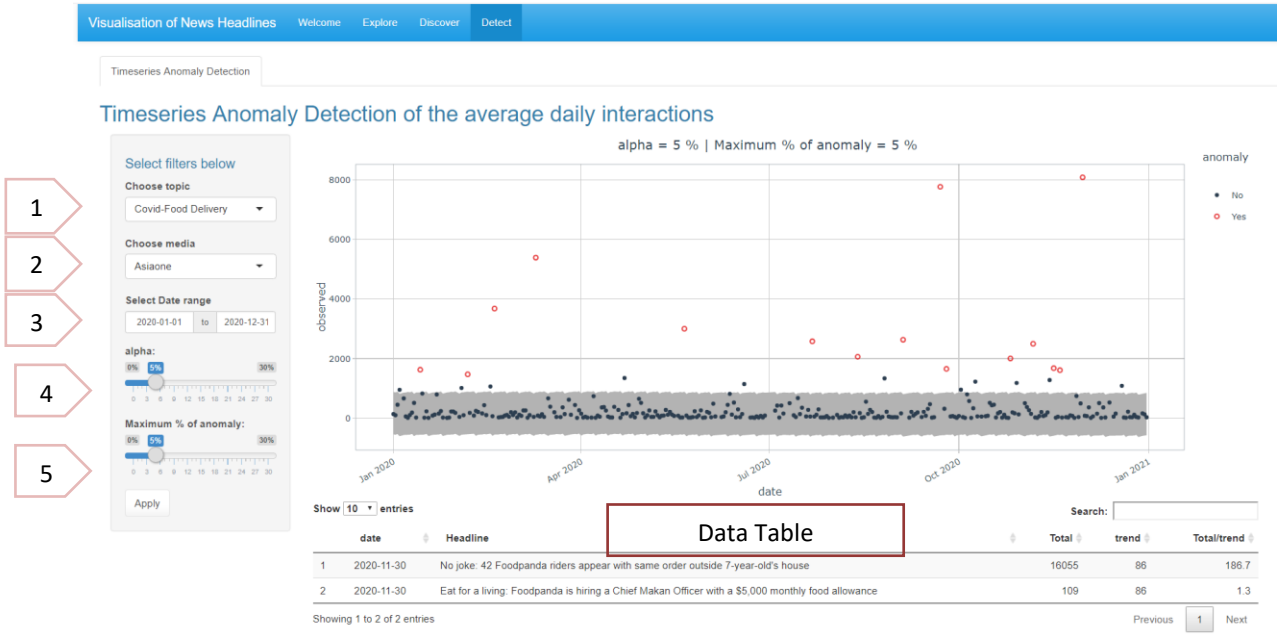
The ***main panel,Corpus info*** tab also displays key statistics based on the words searched by the user.

3.Discover



From the ***Discover*** tab, the user can visualise tri-grams from our corpus based on the choice of input (**Do note that the input is case sensitive, in that it require user input to be in lowercase due prior preprocessing of the corpus)

4.Detect



The ***detect tab*** contains the Timeseries Anomaly plot output wrapped by plotly by making use of the 'anomalize' package.

The Sidebar contains filters to

1. Choose Topic
2. Choose Media Source
3. Filter Date Range
4. Adjust the alpha (determines the width of critical values) from 0% - 30%
5. Adjust Maximum percentage of anomaly (man_anoms) from 0%-30%%

Points that are detected as anomalous in the news corpus will be indicated as red as show in the legend.

To better understand the potential cause for the anomaly, you can interact with the plot to select the articles that were most likely to have caused it.

The articles, the corresponding headlines, total and trend scores will be found in the data table below.

- Total refers to the Engagement metrics taken from Facebook Likes, Shares and Comments
- Trend is the moving average of Engagement score before and after the date selected

5.Ending

You have reached the end of the user guide. We hope that this has helped you in navigating our application and that the app itself has provided you with useful information in understanding headlines in the Singapore context in the year of 2020.