**Instruction Manual: News Aggregator Application**

*Windows-Based Interactive News Extraction and Analysis Tool*

|  |  |
| --- | --- |
| **GROUP 17** |  |
| **Team member** | **Student ID** |
| Dan Hansen | 24718999 |
| Brian Shimmer Bino Deva Kumar |  |
| David Pawley |  |
| Maximus Chandrasekaran | 25614189 |

[1. Introduction 3](#_Toc197894999)

[2. How to run the Program 3](#_Toc197895000)

[3. Getting Started 3](#_Toc197895001)

[4. Interface Overview 4](#_Toc197895002)

[5. API Options Panel 5](#_Toc197895003)

[6. Article Extraction Buttons 5](#_Toc197895004)

[7. API Key Handling 6](#_Toc197895005)

[8. Extract Information Panel 6](#_Toc197895006)

[9. Extracted News Stories Display 6](#_Toc197895007)

[10. Keyword Frequency Graph 7](#_Toc197895008)

[11. Summary of Controls 7](#_Toc197895009)

[12. Additional Information 7](#_Toc197895010)

# 1. Introduction

The News Aggregator is a Windows-based python interactive desktop application that allows users to extract, combine, and analyse news articles from two primary sources:

* **NewsAPI.org** (via Web API)
* **ABC News Australia** - **abc.net.au/news** (via web scraping)

The extracted articles are compiled into a single dataset, displayed in a user-friendly interface, and visualised through keyword frequency graphs. The application supports real-time interaction, customisable keyword filters, and source selection options.

# 2. How to run the Program

The following Python packages must be installed:

* Tkinter
* requests
* pycountry
* pandas
* datetime
* bs4 (BeautifulSoup)
* matplotlib
* collections
* string
* random
* sys

It may be necessary to obtain a NewsAPI key from [newsapi.org](https://newsapi.org) and insert it into the application (the app will handle key cycling automatically).

Launch the application from with VS Code (or similar) or via a command line

* python -u Assignment\_36122\_Group\_17.py

# 3. Getting Started

Upon startup, the application will go to the NewsAPI.org service and automatically retrieves the latest available:

* Sources
* Categories
* Countries

# 4. Interface Overview

The main window is divided into the following components:

* API Options selection
* News Extraction Controls
* Extracted News Display
* Common Title Words Graph
* Status Indicators

A screenshot of a computer

AI-generated content may be incorrect.

# 5. API Options Panel

**The Radio Button Choices**

At the top right of the screen, choose the option on how to retrieve NewsAPI articles:

**Source:**

* Displays a list of all available news sources.

Select one to extract articles from that specific source.

**Category / Country:**

* Displays two list boxes:
  + Category List (e.g., business, health, technology, etc.)
  + Country List (e.g., United States, Australia, etc.)

Both must be selected in order to perform an extraction.

Nb: A special entry titled "Chinese Language" appears for articles from China, Hong Kong, and Taiwan.

# 6. Article Extraction Buttons

**API News Button**

**Label:** "Get API News" / "Clear API News"

**Functionality:**

When "Get API News" is clicked:

* + Extracts articles from the selected Source or Category/Country.
  + Displays articles in the Extracted News Stories panel.
  + Generates a keyword frequency graph from article titles.

Status Label: Indicates the total number of articles extracted or notifies if no articles were found.

When "Clear API News" is clicked:

* + Removes all previously extracted API articles.
  + ABC articles, if any, remain unaffected.

Status Label: Displays the current state of the API extraction.

**ABC News Australia Button**

**Label:** "Get ABC News" / "Clear ABC News"

**Functionality:**

When "Get ABC News" is clicked:

* + Scrapes and imports the latest articles from ABC News Australia.
  + Articles are added to the existing dataset (if any).

When "Clear ABC News" is clicked:

* + Removes only the ABC articles from the dataset.
  + API articles remain untouched.

Status Label: Displays the status of ABC extraction.

# 7. API Key Handling

The application includes a built-in list of API keys. Upon startup, it randomly selects and tests these keys.

The first valid key is used until:

* The application closes, or
* The API quota is exceeded, whereupon the application must be restarted to ensure the selection of the next valid key.

# 8. Extract Information Panel

Displays information regarding the currently extracted articles from both NewsAPI and ABC sources.

• The total number of articles.

• Displays the percentage of the source contributing the most articles

• The minimum and maximum publication dates of the article set.

# 9. Extracted News Stories Display

Each article is numbered and shown in a structured format:

* Title
* Source
* Publication Date
* Author (if available)
* Description
* Extraction Method (API or ABC Scrape)
* URL (can be copied into a browser).

# 

# 10. Keyword Frequency Graph

This feature visualises the 20 most frequent words found in article titles.

Graph Features:

* Displays words in descending frequency
* Omits words from the Omitted Words list

Omitted Words Textbox:

* Default: Contains 20 common English stopwords (e.g., "the", "and", "is")
* Customisation:

Words can be added or removed using semi-colon (;) separators

Click the Replot button to update the graph.

# 11. Summary of Controls

|  |  |
| --- | --- |
| **Control** | **Action** |
| API Options | Choose between Source or Category/Country |
| Get API News | Extract articles using selected Source/Category-Country |
| Clear API News | Remove all API articles from the dataset |
| Get ABC News | Scrape and add articles from ABC News Australia |
| Clear ABC News | Remove ABC articles from the dataset |
| Replot Button | Refresh keyword graph after editing omitted word list |
| Extract Information | View count of articles by source |
| Status Labels | Track API and ABC extraction progress |

# 12. Additional Information

Internet access is required to fetch and scrape news.

Ensure your system clock is accurate to prevent API key issues.

If article fetching fails, try switching between sources or updating the omitted words list.