

A Mini Project Report on

**NEWSIN - A NEWS SUMMARIZER**  
**AND ANALYZER**

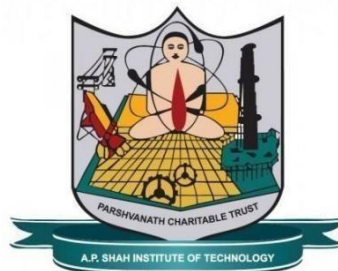
Submitted in partial fulfillment of the requirement of  
University of Mumbai for the Course

**Natural Language Processing**  
**In**  
**Computer Engineering (VIII SEM)**

Submitted by

**Pooja Tumma (20102126)**  
**Kunal Saini (20102154)**  
**Pragya Tripathi (20102047)**

Under the Guidance of  
**Prof Suchita Dange**



Department of Computer Engineering

A.P. Shah Institute of Technology  
G.B. Road, Kasarvadavli, Thane(W), Mumbai-400615 UNIVERSITY OF  
MUMBAI  
**Academic Year 2023-2024**

**Department of Computer Engineering**  
**A. P. Shah Institute of Technology**  
**G.B. Road, Kasarvadavli, Thane(W), Mumbai-400615 UNIVERSITY OF**  
**MUMBAI**

**CERTIFICATE**

This is to certify that the requirements for the project report entitled ‘NewsIn - A news summarizer and Analyzer’ have been successfully completed by the following students:

<b>Name</b>	<b>Moodle ID</b>	<b>Roll No:</b>
Pooja Tumma	(20102126)	190
Kunal Saini	(20102154)	152
Pragya Tripathi	(20102047)	189

in partial fulfillment of the course Natural Language Processing in Computer Engineering (VIII SEM) of Mumbai University in the Department of Computer Engineering, A P Shah Institute of Technology during academic year 2023-24.

**Subject In-charge**  
**Prof Suchita Dange**

## **Abstract**

NewsIn introduced a system for summarizing news articles from various online newspapers automatically. It uses a custom web crawler to find relevant articles, employs computational linguistic techniques like Triplet Extraction and Semantic Similarity, and utilizes clustering methods for coherent summaries. The goal is to help readers quickly understand the essence of news articles without reading the entire text. With the increasing volume of digital news data and the lack of summaries, this approach aims to condense and present essential information in a way that aligns with human reading habits, using knowledge graphs for accurate abstractive news summarization.

## **Introduction**

News articles are valuable sources of information, but extracting meaningful comparisons from them can be challenging due to the diverse perspectives they present. This complexity arises from various factors, including the timing of events, the individuals involved, and the sentiments expressed. For instance, words like "storm" and "rain" both relate to "weather" and can create potential comparisons. Addressing these challenges, the task of news summarization aims to concisely capture both the similarities and differences between two related news topics using human-readable sentences. This involves analyzing two sets of news articles, identifying commonalities and distinctions, and presenting them pairwise. For example, when comparing earthquakes in Haiti and Chile, the summary should encompass factors like magnitude, damages, and government responses. This process utilizes computational linguistic methods to streamline and organize information efficiently, making it easier for readers to grasp key points from news articles.

## **Problem Statement**

In the current world situation when there is a rapid increase in technology, the data on the World Wide Web is increasing at a tremendous rate. However due to the hectic schedule of the people and an intense amount of news available on various different websites it becomes difficult for people to be daily updated with the knowledge of the surroundings. Also as the web is getting developed on a daily basis the news which might get surfaced on the web may not provide an overview of the news . We are not able to analyze what news should be read.

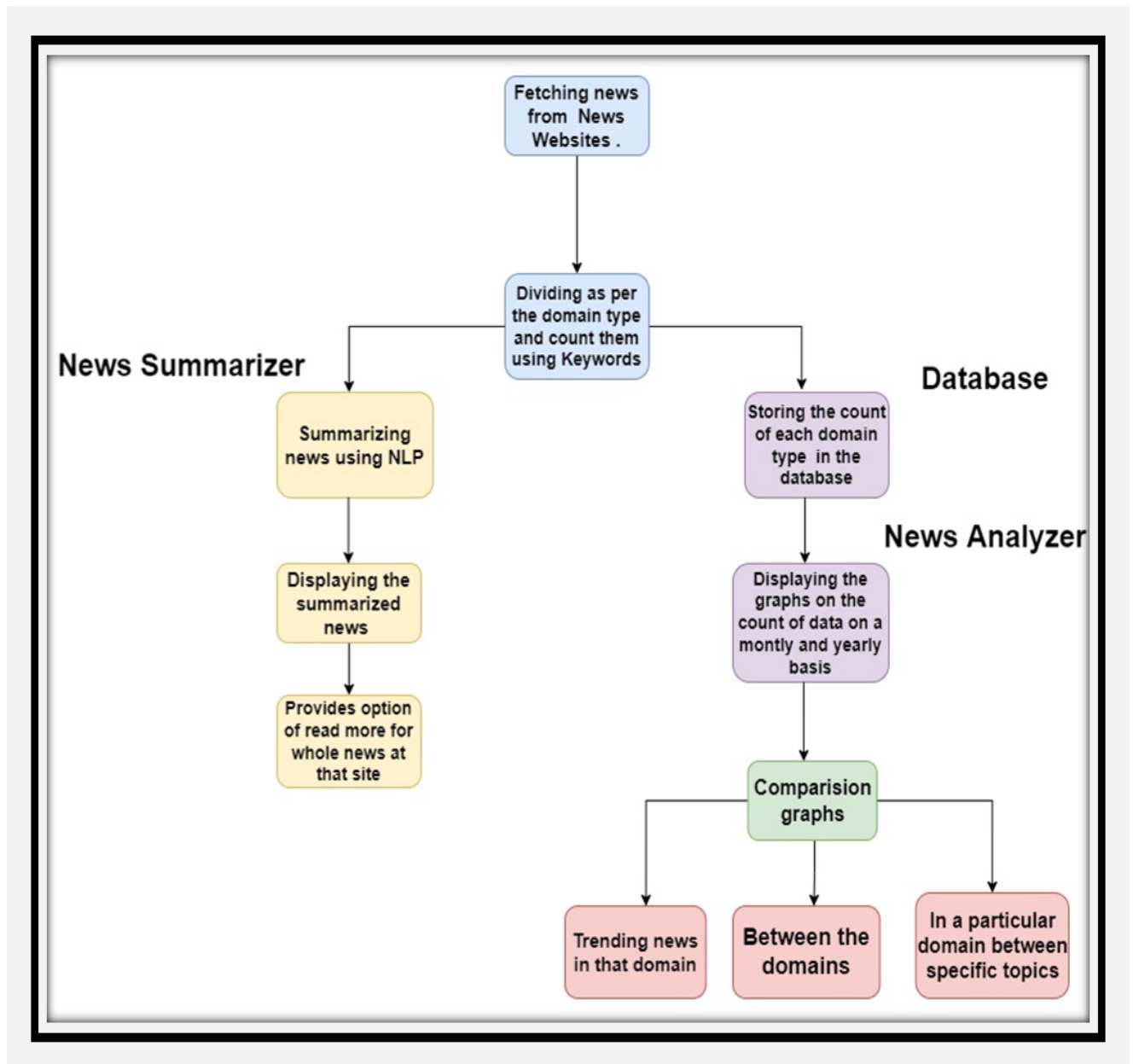
## **Objectives and Scope**

1. To extract data from a News Website.
2. To categorize the news as per the domains.
3. To provide summarize content of the news
4. To provide the analyzed news in the form of graphs, charts along with categories.
5. To extract data through news Website
6. Segregating that news as per the Domain
7. Extracting the domain data count for summarization and analyzation
8. To store that count of the domain data in the database
9. Summarizing the news and displaying them
10. After summarization is done then Extracting the domain data count from the database for analyzing.
11. Then as per category displaying that data in graphical representations

## **Implementation Technology**

1. Html 5/CSS 3/ - frontend
2. Python 3.10 - backend
3. NLP
4. Plotly
5. Application connected via: News Website
6. Database: MySQL

## Flowchart



## **Project Implementation**

### **1) Algorithm:**

Step 1: Import all the Modules.

Step 2: Initialize a variable with 'Analyzer.xlsx'.

Step 3: Initialize function search topic.

Step 4: Open site.

Step 5: Read

Step 6: Initialize news\_list=sp\_page.findall('item') #finding news

Step 7: return news\_list

Step 8: Repeat from steps 3 for various functions of various categories.

Step 9: Initialize function news\_poster.

Step 10: Open poster link

Step 11: Read data

Step 12: Open Image

Step13: Close function

Step 14: Initialize run() function

Step 15: Select category=[' --select-- ', 'Trending News', 'Favorite Topics', 'Search Topic']

    If category[0] warning("Please select type !")

    elif category[1] show trending news

    elif category[2] select favorite topic

        show favorite topic

    Elif category[3] enter your topic

        show entered topic

Step 16: End

## 2) Methodology:

1. Extracting news from NDTV News website.
2. Fetching news using RSS Feed.
3. Segregating the news based on different domains i.e., sports, technology, etc.
4. Summarizing the news using NLP.
5. After summarization we perform analysis on the news.
6. For performing analysis we store the news count of all the domains.
7. The news count of domains is stored in the database with the help of keywords.
8. Once the count got stored in the database, we converted it into a csv file for performing analysis.
9. Analysis part was implemented by using Plotly.
10. By using plotly we generated pie charts for all the domains and also a line graph that will represent the weekly analysis of news.

## Output Screenshots





(1) Scared, Without Food, No Place To Go: Thousands Flee As Israel Pounds Gaza

Scared, Without Food, No Place To Go: Thousands Flee As Israel Pounds Gaza

Published Date: Sun, 08 Oct 2023 23:46:13 +0530

(2) Iran Funding Hamas Operations, Says Israel's Representative At UN

Iran Funding Hamas Operations, Says Israel's Representative At UN

Israel has launched a massive counterattack at Gaza Strip after the Hamas attack. Iran has been funding the Hamas group's operations, Israel's envoy to the UN said today amid the war with Palestine that started after Hamas launched its biggest attack on Israel from the Gaza Strip. Israel's representative to the UN Gilad Erdan said they are quite sure that Iran has been backing Hamas' operations and by extension the terrorist attack on Saturday morning that left over 600 Israelis dead and thousands injured. "Iran supports the legitimate defence of the Palestinian nation," Mr Raisi said, quoted by state television. He urged Muslim governments to "support the Palestinian nation" while praising "resistance" efforts by Hamas and other groups in countries including Syria, Lebanon and Iraq. Iran had talks with leaders of Hamas and Islamic Jihad in June."



## **Result**

In this project, we successfully summarized news articles and presented the data through a user-friendly interface. We focused on specific domains and added extra features to enhance information retrieval. Using various graphics, we displayed data in a visual format, allowing users to analyze the global situation.

We gathered the latest news from different sources using the News API, organized them, and presented them in one place. The news summaries provided users with a general understanding of the topics and their global impact. We noticed that starting phrases with repeated nouns consistently received positive feedback.

Through this project, we gained insights into daily news trends, such as popular news topics and keywords driving their popularity. We built the user interface using a Python module called Streamlit, adding dynamic elements for a better user experience. Unlike many text analytics projects using R, we chose Python to improve efficiency and accuracy in delivering information.

## **References**

[1]Rananavare, Laxmi & Reddy, P.. (2018). Automatic News Article Summarization.

International Journal of Computer Sciences and Engineering. 6. 230-237. doi:

10.26438/ijcse/v6i2.230237.

[2]P. Sethi, S. Sonawane, S. Khanwalker and R. B. Keskar, "Automatic text summarization of news articles," 2017 International Conference on Big Data, IoT and Data Science (BID), 2017, pp. 23-29, doi: 10.1109/BID.2017.8336568.

[3]Nahar, J., Kline, D, Layman, L., Modares Nezhad, M. (2019) Daily Text Analytics of News and Social Media with Power BI. Annals of the Master of Science in Computer Science and Information Systems at UNC Wilmington, 13(2) paper 2. <http://csbapp.uncw.edu/data/mscsis/full.aspx>

[4] Roth, Robert. (2017). User Interface and User Experience (UI/UX) Design. Geographic Information Science & Technology Body of Knowledge. 2017. doi:10.22224/gistbok/2017.2.5.

[5] Developer Documentation for Web Development