



Department of Computer Science & Engineering
B. P. Poddar Institute of Management & Technology

PCA-2 PROJECT
IT WORKSHOP (PYTHON) – PCC-CS-393

Name	Roll No.
DEBANJAN GANGULY	11500122001
AMAN SANTRA	11500122009
ARGHA MALLICK	11500122014
AMRITASH BANERJEE	2422127

Project Title – News Summarization System

using Google RSS News Feed API & Natural Language Processing



Debanjan Ganguly

Aman Santra

Argha Mallick

Amritash Banerjee

Project Overview

The News Summarization System is an innovative machine learning-driven project designed to streamline and condense news articles fetched from the Google RSS News Feed API. The system is engineered to collect a vast array of current news articles from various sources, process their content, and present concise and easily digestible summaries to users. Leveraging machine learning techniques, natural language processing (NLP), and text summarization algorithms, the system's core functionality involves converting lengthy news articles into shorter, more coherent summaries.

Key Components and Workflow

Google RSS News Feed API Integration: The system interfaces with the Google RSS News Feed API, allowing it to retrieve a diverse range of news articles across multiple categories and sources. This integration ensures a constant inflow of up-to-date information.

Data Preprocessing: Upon retrieval of news articles, the data undergoes preprocessing, including text cleaning, tokenization, and filtering out redundant or irrelevant information. This step prepares the content for subsequent analysis.

Natural Language Processing (NLP): NLP techniques are applied to comprehend the context, semantics, and structure of the news articles. This includes tasks such as part-of-speech tagging, named entity recognition, and sentiment analysis to grasp the nuances of the text.

Text Summarization Algorithms: Employing machine learning-driven summarization algorithms such as extractive or abstractive methods, the system condenses the articles into brief, coherent summaries. Extractive methods may involve ranking and selecting key sentences, while abstractive methods might involve generating new sentences based on the content.

User Interface and Different Routes: The system provides a user-friendly interface through different routes or paths, allowing users to access summarized news in various categories or topics. This may include options to explore summarized news by subject, date, source, or user preferences. Also, user can search for any particular topics in the search bar to view the related news articles.

Output and User Interaction:

The summarized news articles are then displayed to the user through the chosen routes, enabling quick and efficient consumption of essential information. Users can access a wide array of summarized news articles, saving time while staying informed about current events.

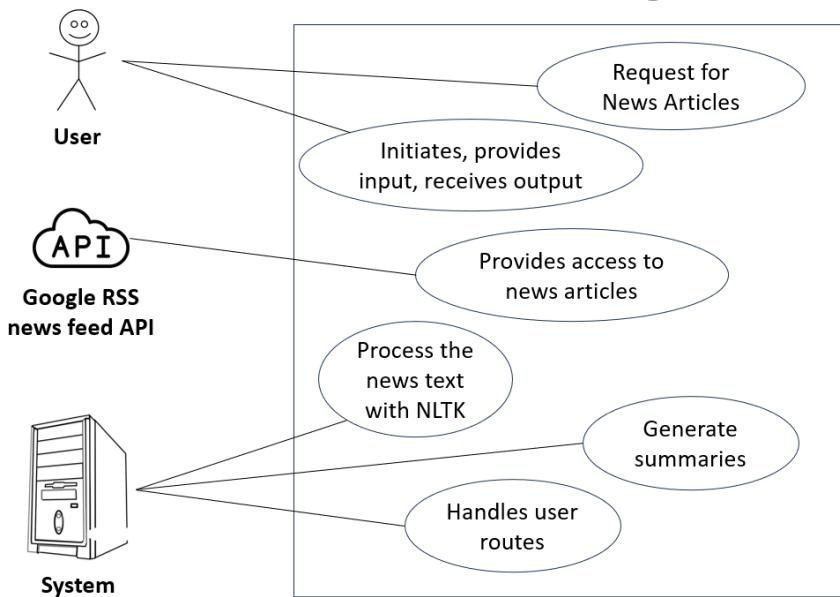
Benefits and Future Prospects

This project aims to offer users a convenient way to stay updated on current events without investing extensive time in reading lengthy articles. It facilitates quick access to crucial information, making it beneficial for individuals with time constraints or a need for rapid news consumption.

Future iterations of the project might incorporate improvements in summarization techniques, personalization features, user feedback integration, and enhanced user interfaces for an even more tailored and engaging experience.

This project holds significant potential in providing users with a more efficient and accessible method for staying informed about the latest news, and it showcases the powerful applications of machine learning in content summarization.

Use Case Diagram



SOURCE CODE

app.py

```
from flask import Flask, render_template, request
from bs4 import BeautifulSoup as soup
from urllib.request import urlopen
from newspaper import Article
import nltk
# nltk.download('punkt')

def fetch_news_search_topic(topic):
    topic = topic.replace(" ", "%20")
    site = 'https://news.google.com/rss/search?q={}'.format(topic)
    op = urlopen(site) # Open that site
    rd = op.read() # read data from site
    op.close() # close the object
    sp_page = soup(rd, 'xml') # scrapping data from site
    news_list = sp_page.find_all('item') # finding news
    return news_list

def fetch_top_news():
    site = 'https://news.google.com/news/rss'
    op = urlopen(site) # Open that site
    rd = op.read() # read data from site
    op.close() # close the object
    sp_page = soup(rd, 'xml') # scrapping data from site
    news_list = sp_page.find_all('item') # finding news
    return news_list

def fetch_category_news(topic):
    site = 'https://news.google.com/news/rss/headlines/section/topic/{}'.format(topic.upper())
    op = urlopen(site) # Open that site
    rd = op.read() # read data from site
    op.close() # close the object
    sp_page = soup(rd, 'xml') # scrapping data from site
    news_list = sp_page.find_all('item') # finding news
    return news_list
```

```

def display_news(list_of_news, news_quantity):
    c = 1
    response = []
    for news in list_of_news:
        news_data = Article(news.link.text)
        try:
            news_data.download()
            news_data.parse()
            news_data.nlp()
        except Exception as e:
            print("Error:", e)

        news_id = str(c) + 'NewsBits'
        news_title = news.title.text
        news_summary = news_data.summary
        news_source = news.source.text
        news_link = news.link.text
        news_date = news.pubDate.text
        news_image = news_data.top_image

        data_dict = {}
        data_dict["news_id"] = news_id
        data_dict["news_title"] = news_title
        data_dict["news_image"] = news_image
        data_dict["news_summary"] = news_summary
        data_dict["news_source"] = news_source
        data_dict["news_link"] = news_link
        data_dict["news_date"] = news_date

        response.append(data_dict)

        print(f"Fetched News {c}")

        if c >= news_quantity:
            break
        c += 1

    return response

app = Flask(__name__)

@app.errorhandler(404)
def not_found(e):
    print(e)
    return render_template('404.html')

default_news_quantity = 2

@app.route('/', methods=["GET", "POST"])
def home_page():
    if request.method == "GET":
        print("/")
        news_list = fetch_top_news()
        response = display_news(list_of_news=news_list,
                               news_quantity=default_news_quantity)
        return render_template('home.html', response=response, topic="Top Trending",
                           news_quantity=default_news_quantity, request_url="/")

    if request.method == "POST":
        load_more_range = int(request.form.get('load_more_range'))
        print(f"Loading more on '/' with news quantity: {load_more_range}")
        news_list = fetch_top_news()
        response = display_news(list_of_news=news_list,
                               news_quantity=load_more_range)
        return render_template('home.html', response=response, topic="Top Trending",
                           news_quantity=load_more_range, request_url="/")

```

```

@app.route('/topics/<string:topic>', methods=["GET", "POST"])
def search_by_topics(topic):
    news_list = fetch_category_news(topic)
    if request.method == "GET":
        print(f"GET '/topics/{topic}'")
        response = display_news(list_of_news=news_list,
                                news_quantity=default_news_quantity)
        return render_template('home.html', response=response, topic=topic.upper(),
news_quantity=default_news_quantity, request_url="/topics/"+topic)

    if request.method == "POST":
        load_more_range = int(request.form.get('load_more_range'))
        print(f"Loading more on '/topics/{topic}' with news quantity: {load_more_range}")
        response = display_news(list_of_news=news_list,
                                news_quantity=load_more_range)
        return render_template('home.html', response=response, topic=topic.upper(),
news_quantity=load_more_range, request_url="/topics/"+topic)

@app.route('/search', methods=["GET", "POST"])
def search():
    if request.method == "GET":
        print("GET '/search'")
        return render_template('home.html', request_url="/search")

    if request.method == "POST":
        try:
            load_more_range = int(request.form.get('load_more_range'))
        except:
            load_more_range = default_news_quantity
        topic = request.form.get('topic')
        print(f"Searching for {topic} with news quantity: {load_more_range}")
        news_list = fetch_news_search_topic(topic)
        response = display_news(list_of_news=news_list,
                                news_quantity=load_more_range)
        return render_template('home.html', response=response, topic=topic, news_quantity=load_more_range,
request_url="/search")

if __name__ == "__main__":
    app.run(host='0.0.0.0', port=3000, debug=True)

```

templates/layout.html

```
<!doctype html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <link href="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/css/bootstrap.min.css" rel="stylesheet"
        integrity="sha384-EVSTQN3/azprG1Anm3QDgpJLIm9Nao0Yz1ztcQtTwFspd3yD65VohppuC0mLASjC"
        crossorigin="anonymous">
    <title>News Summerization System</title>
    <link type="image/png" sizes="32x32" rel="icon" href="{{ url_for('static',
        filename='/images/favicon.png') }}"/>
    <link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-
        awesome.min.css">
    <link rel="stylesheet" href="{{ url_for('static', filename='/css/style.css') }}"
    <link rel="stylesheet" href="{{ url_for('static', filename='/css/scrollbar.css') }}>
</head>

<body>
    <nav class="navbar navbar-dark bg-dark navbar-expand-lg" style="background-color: #735d7e !important;">
        <div class="container-fluid">
            <a class="navbar-brand" href="/">NewsBits</a>
            <button class="navbar-toggler" type="button" data-bs-toggle="collapse"
                data-bs-target="#navbarSupportedContent" aria-controls="navbarSupportedContent" aria-
                expanded="false"
                aria-label="Toggle navigation">
                <span class="navbar-toggler-icon"></span>
            </button>
            <div class="collapse navbar-collapse" id="navbarSupportedContent">
                <ul class="navbar-nav ms-auto mb-2 mb-lg-0" style="padding-right: 10px;">
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/">HOME</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/world">WORLD</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/nation">NATION</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/business">BUSINESS</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/technology">TECH</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/entertainment">ENTERTAINMENT</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/sports">SPORTS</a>
                    </li>
                    <li class="nav-item">
                        <a class="nav-link active" aria-current="page" href="/topics/science">SCIENCE</a>
                    </li>
                </ul>
            </div>
        </div>
    </nav>

    <main class="bg-dark min-h-100 pb-5">
        <div class="container">
            
            <h1 class="text-light text-center">Welcome to NewsBits  Read the Summarized News Articles</h1>
            <form action="/search" method="POST" class="p-3">
                <div class="form-group mb-3 form-content">
```

```
        {% if topic %}
        <input value="{{topic}}" name="topic" type="text" class="form-control" id="topic"
               placeholder="Search Topics" required>
        {% else %}
        <input name="topic" type="text" class="form-control" id="topic" placeholder="Search
Topics"
               required>
        {% endif %}
        <button type="submit" class="btn btn-primary"><i class="fa fa-search"></i></button>
    </div>
</form>
</div>

{% block body %}

{% endblock body %}

</main>

<footer>
    <p>News Summerization System &copy; Made with Flask by Machine Learning Integrity</p>
</footer>

<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.0.2/dist/js/bootstrap.bundle.min.js"
       integrity="sha384-MrcW6ZMFYlzcLA8Nl+A7MsXsP1UyJoMp4YLEuNSfAP+JcXn/tWtIxVXM"
       crossorigin="anonymous"></script>
</body>

</html>
```

templates/home.html

```
{% extends 'layout.html' %}

{% block body %}

{% if response %}

<h2 class="text-center text-light pb-3">Showing News for <strong>{{topic}}</strong></h2>
<div class="container home text-light">
    {% for news in response %}
        <div class="news-card">
            
            <div class="accordion" id="accordionExample{{news['news_id']}}">
                <div class="accordion-item">
                    <h2 class="accordion-header">
                        <button class="accordion-button text-light" type="button" data-bs-toggle="collapse"
                               data-bs-target="#collapseOne{{news['news_id']}}" aria-expanded="true"
                               aria-controls="collapseOne{{news['news_id']}}"
                               style="background-color: #735d7e;">{{news['news_title']}}
                    </button>
                </h2>
                <div id="collapseOne{{news['news_id']}}" class="accordion-collapse collapse show"
                     data-bs-parent="#accordionExample{{news['news_id']}}">
                    <div class="accordion-body bg-dark">
                        {{news['news_summary']}}
                        <br>
                        <p class="font-weight-bold"><a class="text-warning font-weight-bold" target="_blank"
                           href="{{news['news_link']}}>Read
                           full
                           article on
                           {{news['news_source']}}</a></p>
                    </div>
                </div>
            </div>
            <p class="bg-success text-center">Published Date {{news['news_date']}}</p>
        </div>
    {% endfor %}

</div>
<div class="container mt-4">
    <form action="{{request_url}}" method="POST">
        <div class="load_more">
            <input value="{{topic}}" name="topic" type="text" class="form-control" id="topic"
                   placeholder="Search Topics" hidden>
            <label class="btn btn-light">Number of Articles {{news_quantity}}</label>
            <button type="submit" class="btn btn-primary">Go</button>
        </div>
        <input class="w-100" type="range" name="load_more_range" id="load_more_range" step="1" min="1"
               max="25"
               value="{{news_quantity}}>
        <div class="load_more text-light">
            <span>01</span><span>05</span><span>10</span><span>15</span><span>20</span><span>25</span>
        </div>
    </form>
</div>

{% else %}

<div class="container">
    <a href="/"></a>
</div>

{% endif %}

{% endblock %}
```

OUTPUT SCREEN

End point: Homepage('/')

NewsBits

HOME WORLD NATION BUSINESS TECH ENTERTAINMENT SPORTS SCIENCE

hello readers

Welcome to NewsBits  Read the Summarized News Articles

Top Trending

Showing News for Top Trending





Apple Warns Top Indian Opposition Leaders, Journalists About 'State-Sponsored' Attack on Phone - The Wire 

"The reports of threat notifications from Apple need to be taken very seriously and require investigation to determine the source and the extent of the malware attack. "Firstly, reports indicate that India has been a ground for deploying Pegasus spyware by NSO Group, an Israeli firm. In October, 2019, state attackers targeted activists, and in July, 2021 they extended their reach to public officials and journalists. The Indian government is scouring the globe for spyware it could use which has a "lower profile" than Pegasus. If you have received such an email from Apple, please get in touch with us at ."

[Read full article on The Wire](#)

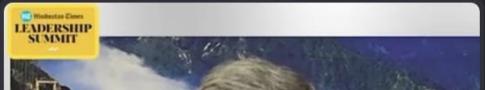
Published Date Tue, 31 Oct 2023 11:02:04 GMT

"When I Was In Trouble...": Out On Bail, Chandrababu Naidu's Message - NDTV 

READ | The ₹ 371-Crore Scam That Led To Chandrababu Naidu's Arrest: Explained Massive celebrations broke out in Rajahmundry and other parts of Andhra Pradesh once Mr Naidu walked out of jail. I will never forget the affection shown to me... not only in Andhra Pradesh but also Telangana and across the world." #WATCH | Supporters of former Andhra Pradesh CM and TDP Chief N Chandrababu Naidu surround him as he walks out of Rajahmundry jail. "Yesterday, Chandrababu Naidu was named Accused No 3 in a case of allowing illegal licences to liquor distillery. Mr Naidu's arrest last month triggered political turmoil and led to violent clashes.

[Read full article on NDTV](#)

Published Date Tue, 31 Oct 2023 12:03:01 GMT





HTLS Day 1: Ashwini Vaishnaw explains PM Modi's 20-year semiconductor plan - Hindustan Times

"Semiconductors are a foundational industry, and we will become a major player in the semiconductor industry going forward. Elaborating on the global semiconductor market, he said India has been trying to carve out its niche in an untapped area of the industry. It gives the ability to companies to test out their products in a large market. On telecom and 5G internet Ashwini Vaishnaw said Prime Minister Narendra Modi's government completely turned around the telecom sector that had been 'bleeding'. I read a recent report that hiring in the telecom sector is going on at 20 percent.

[Read full article on Hindustan Times](#)

Published Date Tue, 31 Oct 2023 12:54:40 GMT



Cash for queries scam: Mahua Moitra drops defamation lawsuit against media houses and social media platforms, informs Delhi High Court - OpIndia

The Trinamool Congress leader Mahua Moitra's Lawyer on Tuesday informed the Delhi High Court that they are not pressing for any relief in the suit against media houses and social media intermediaries. Earlier on October 20, 2023, the Delhi High Court adjourned TMC MP Mahua Moitra's defamation plea against Nishikant Dubey for October 31. While Senior Advocate Sankaranaryan appearing for Mahua Moitra, was putting forth his submissions, Adv Jai Anant Dehradrai, against whom also injunction was sought, objected to him appearing in the case. I immediately withdrew and said that I don't want to do this case. The Delhi High Court earlier had issued notices to the defendants on Moitra's suit. Dubey claimed that an Advocate, Jai Anant Dehradrai had provided him with proof of a bribe.

[Read full article on OpIndia](#)

Published Date Tue, 31 Oct 2023 07:39:27 GMT



Kerala blasts accused worked in Middle East as electrical foreman, carried 2 IEDs to convention centre: Police - The Indian Express

The key accused in the Kochi blasts was working in the Middle East as an electrical foreman until two months ago, and had gained skills to experiment with IEDs using locally available materials, initial probe has revealed. Police also found that Dominic Martin extensively searched the internet to learn about how to make explosives, and that he had made more than one IED at his house at Thammanam in Kochi, which he carried in a plastic bag to the convention centre in Kalamassery early Sunday morning. After placing the plastic bags containing the IEDs at the convention hall before participants entered, he waited nearby and triggered the blast using two remotes. He then rode his two-wheeler towards Thrissur and checked into a hotel, from where he uploaded a confessional video. After spending less than half an hour at that hotel, he turned up at Kodakara police station in Thrissur claiming responsibility for the blasts.

Published Date Mon, 30 Oct 2023 22:53:31 GMT



Israel expands Gaza offensive, sends troops into Hamas tunnels - Firstpost

Israel announced on Tuesday that it had launched an attack on Hamas terrorists within the Islamists' extensive network of tunnels beneath Gaza. According to witnesses, Israeli forces attacked Gaza City from two directions on Monday, focusing on the city's main north-south road. The armed wing of Hamas, known as the al-Qassam brigades, reported that terrorists and Israeli forces engaged in combat early on Tuesday. Calls for a ceasefire are calls for Israel to surrender to Hamas, to surrender to terrorism, to surrender to barbarism. Israeli forces are advancing slowly in their ground offensive, according to military experts, in part to leave open the possibility that Hamas terrorists will mediate the hostages' release.

[Read full article on Firstpost](#)

Published Date Tue, 31 Oct 2023 10:05:43 GMT

Number of Articles (6)

01

05

10

15

20

Go

25

End point: Topics ('/topics/sports')

The screenshot shows the NewsBits website interface. At the top, there's a purple header bar with the 'NewsBits' logo on the left and navigation links for 'HOME', 'WORLD', 'NATION', 'BUSINESS', 'TECH', 'ENTERTAINMENT', 'SPORTS', and 'SCIENCE' on the right. Below the header is a large, stylized purple oval containing the text 'hello readers' and an icon of five books. Underneath the oval, a banner reads 'Welcome to NewsBits 📰 Read the Summarized News Articles'. A search bar with a magnifying glass icon is positioned below the banner. The main content area is titled 'Showing News for SPORTS'. It features two large thumbnail images: one of a cricket match between Bangladesh and Pakistan, and another of a Sri Lankan player lying on the ground. Below each thumbnail is a news summary card. The first card is for 'Pakistan vs Bangladesh highlights, World Cup 2023: Fakhar Zaman's 81 helps PAK beat BAN by 7 wickets - The Indian Express'. It includes a photo of Fakhar Zaman batting and a detailed description of his performance. The second card is for 'Boy Faints Under Pune Sun During Sri Lanka Anthem At Cricket World Cup Game vs Afghanistan. Watch - NDTV Sports', featuring a photo of a boy fainting and a brief description of the incident. Both cards have a green footer bar indicating they were published on Tuesday, October 31, 2023, at 16:00:00 GMT and 10:44:55 GMT respectively. At the bottom of the page, there's a search input field labeled 'Number of Articles (2)' with a 'Go' button, and a pagination bar with buttons for 01, 05, 10, 15, 20, and 25.

Other similar endpoints are:

[/topics/world](#)

[/topics/entertainment](#)

[/topics/nation](#)

[/topics/science](#)

[/topics/business](#)

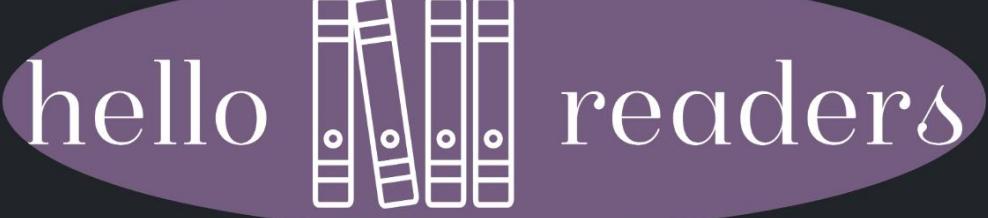
[/topics/tech](#)

Search Endpoint: /search

NewsBits

HOME WORLD NATION BUSINESS TECH ENTERTAINMENT SPORTS SCIENCE

hello readers



Welcome to NewsBits Read the Summarized News Articles

durga puja in kolkata 

Showing News for durga puja in kolkata



Restaurants make Rs 1,100 cr as Kolkata splurges on food during Durga puja - Business Standard 

During the just-concluded Durga puja festival, revellers of Kolkata not only spent hours and overwhelmed by the pandals, idols and overall decorations, besides thronging restaurants to satisfy their taste buds. This is the second Durga puja after all curbs related to the Covid-19 pandemic were withdrawn. "As per preliminary estimates, restaurants in the city did over Rs 1,100 crore business during the period," Poddar told PTI. He said the sales in some popular restaurants in the city were even better than the pre-Covid days of 2019. In comparison to last year, there was a 10-12 per cent growth in sales across all outlets during the Durga Puja period, Ather said.

[Read full article on Business Standard](#)

Published Date Sun, 29 Oct 2023 10:03:07 GMT

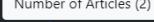


For foreign tourists, Durga Puja turns Kolkata into a city that never sleeps - Times of India 

Kolkata: This Durga Puja, the city of joy played host to a significant number of foreigners with tour operators and hotel associations estimating their footfall close to 20,000. The pace of life picks up as the days proceed and ebbs around late evening. But during the festival, Kolkata metamorphosed into a city that never sleeps. He has shot several videos of the festival and wants to do a docu-feature of the festival in Italian and screen it in the country. What amazed professor Vincenzo Gagliardi from the Naples Academy of Fine Arts was the huge effort and resources that went into crafting the pandals for one of the world's biggest public art events. "The way the city and its people participate in making this grand festival happen, it surely stems out of a divine connection," he remarked.

[Read full article on Times of India](#)

Published Date Sun, 29 Oct 2023 02:51:24 GMT

Number of Articles (2)  Go

01 05 10 15 20 25

[Link to the deployment \(demo\):](#)

<https://newsbits.onrender.com/>