Flask Web Framework based News Summarizer: Web Application

Apeksha Kamble[1], Adnan Gagan[2], Pooja Wagh[3], Umar Aziz[4], Dr. M. R. Rajput[5]

Students of Btech (CSE) [1,2,3,4], Assistant Professor [5], Department of Computer Science and Engineering
P.E.S. College of Engineering, Dr. Babasaheb Ambedkar Tehnological University, Lonere, Raigad (MH), India

Abstract: A summary condenses a lengthy text by emphasising key points. The reader can save time and choose whether to read the entire document by understanding everything upon reading the summary. It makes sure to just choose information that is relevant to the article since summaries should be shorter than the original. The main objective of a newspaper article summary web application is for readers to understand the essential points of the article without having to read the whole news. The proposed system in this study automatically assesses the information from multiple local online newspapers and summarizes it using a summarization API.

The system employs a web crawler, which crawls websites in search of pertinent content, to create summaries of the news articles. Abstractive and extractive summarization approaches are mostly used in computational linguistics techniques. Here, we make use of the advantages of both methods while also minimising its drawbacks. For the web application's front end, we used Streamlit to connect the back end to the front end. We used the Flask web framework. In accordance with the project's requirements, we additionally employed other python dependencies like pillow and urllib.

Keywords: AWS, API, Streamlit, anaconda, newspaper 3k, flask, pillow, Aylien, Text Analysis, meaning cloud

1. INTRODUCTION

The amount of information on the internet is increasing exponentially as a result of a booming global economy and advancement of technologies. Under which we miss out on a significant amount of information that would help in improving our knowledge. A hectic schedule makes it challenging to manage our daily operations and consume the daily newspaper in its entirety. Hence we came up with the concept of News Summarizer. It essentially displays the relevant details from an article while allowing us to browse through the articles of our choice. The architecture of News Summarizer ensures that we exclusively read the important parts of the article without wasting our time reading less important and in depth details of the article at first. The two main types of news summarising approaches are extractive and abstractive. Extractive summarization involves selecting key passages from the text and reproducing them word-for-word to create a subset of phrases from the original text. In contrast, abstractive summarization uses cutting-edge natural language processing to create new, shorter sentences that effectively communicate the most relevant information from the original after interpreting and analysing the text. With our web application and the assistance of natural language processing techniques based on a hybrid implementation of abstractive and extractive techniques our app is able to summarize the entire article within a few lines. Other tools or websites show us exactly what a normal Google news feed would show us.

Hence we came up with a web-based application which has been constructed using Python's Flask framework along with a few other dependencies. There are three main languages that are frequently used and make up the World Wide Web (WWW) standards, namely Hypertext Mark-up Language 5 (HTML5), Cascading Style Sheet (CSS), and JavaScript, in order to make websites look presentable and simple. These three fundamental languages serve as the foundation for most web page developments and are widely used. The maintainability of the code is impacted by the wide variety of languages used in a webpage. In order to handle the frontend, we used a python-based package called streamlit.

2. LITERATURE REVIEW

A considerable amount of literature has been published on various text summary techniques and news summarizers. All these techniques have shown that lots of time is saved by helping us to read the important articles only. It has been conclusively shown that every application's summarising strategy largely adheres to the techniques of Extractive, abstractive and Hybrid method. Different steps involved text summarization using NLP based techniques are explained. Steps involved are converting paragraphs to sentences, text pre-processing, tokenizing the sentences, and replacing words by weighted frequency in original sentences [1]. Django is a best framework for the deployment of the machine learning model in web applications. It provides lots of features to the developer for building the web application. For smaller web based projects flask framework and streamlit framework are suitable [2]. Table 1 gives comparative details of existing news/ article summarisers with respect to different techniques and tools of implementation along with pros and cons of each method.

Table 1. Comparative Study of Different News/ Text Summarisers

Methodology	Techniques	Tools	Pros	Cons
Ref. 3	computational linguistic techniques, language processing techniques, Abstractive and Extractive to summarize the text.	RDF triplets NER DARPA program	The large amount of data available about a topic is concisely presented and thus makes it easier to study and remain informed	The need to eliminate redundancy of data
Ref. 4	Automatic lossless-summarization of News Articles with Abstract meaning representation	AMR Tools and graphs AMR parsing , CNN	summarization where the amount of information loss will be based on the parameters specified by the user in accordance with ones suitability	cost is high
Ref. 5	Data mining	Weak naïve Bayes Classifier	a summary of something is a short account of it, which gives the main points but not the details	difficult to remove irrelevant contents.
Ref. 6	Neural Network model LSTM and GRU cells, RNN. BRNN	Glove 100d	generating one-sentence summarization that mimics the style of news titles given some paragraphs	more deep-learning method exploration, larger training time.
Ref. 7	NLP based text summarization techniques	TF metrics, Latent semantic analysis (LSA)	Sentimental Analysis provides perspective instead of bare facts	No logical relation for sentence arrangement
Ref. 8	end-to-end 'Segmentation-Based News Summarization, CNN	TOPICRANK, SEQ2SEQ	It aims to segment a news article into multiple sections and generate the corresponding summary to each section	No logical relation for sentence arrangement
Ref. 9	unsupervised extractive summarization, Text rank Algorithm , Python Genism library	ATS, Genism library.	It gives the output summary with specified number of words.	Specially used for Marathi text
Ref. 10	RBM, RNN , Text Rank	BM25RBM, RNN, Text Rank	we confirmed that a list of stop words helps quite significantly in all the tested approaches	final result of our testing was that the feature- based approach could not outperform
Ref. 11	Global vector algorithm, News Aggregator and Efficient Summarization system	Lemmatization and Stemming word frequency algorithm	Text Rank algorithm was the chosen approach to be applied in the summarization system over Word Frequency algorithm	The execution depends on the permanent existence of an expert one who knows the actual rules of summarization.
Ref. 12	Weight bigram, ROUGE evaluation metrics, use of facebook public posts to enhance trending news summarization	R-SU4, skip-.bigram	Experiments show that post information is useful for improving the performance.	It just generate the summary from the posts only

Table 2 compares various news aggregator text analytics andweb scraping methods with respect to parameters like cost pros and cons and approaches used for summarizing the news.

Table 2. Comparison of Various News Aggregator, Text analytics and Web Scraping Methods

Methodology	Cost	Pros	Cons	Extractive	Abstractive	Hybrid
Meaning Cloud	40000 in free requests plus $0.0003 every request over that limit each month	good number of free requests for a website with low to medium visitors	Because the summary is extractive, it contains grammar errors and statements that are incoherent.	✓		
Newspaper 3k	Monthly free calls are 100. an extra $0.01 per call	The context and summary are perfectly integrated. they are typically coherent and succinctly lay out all crucial details.	Summarization requires a lot of computer power. Due to the computing time needed to provide summaries, the website occasionally feels choppy.			✓
Aylien Text Analysis	varying price ranges from $200 to $1400 depending on the needs	This API has endpoints for ot her NLP providers	It employs extractive summarization, which accounts for the extensive summaries. The quality is very poor considering the price.	✓		

3. PROPOSED DESIGN

The flow of the proposed system is explained infigure 1 which gives a comprehensive overview.
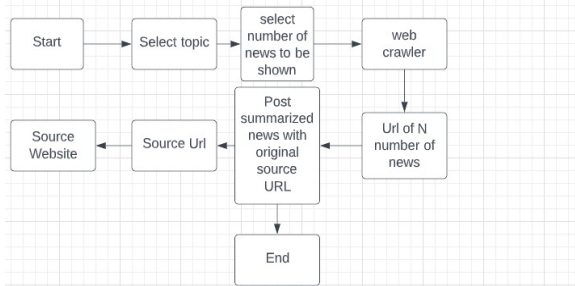

Figure 1. Block diagram of news summarizer

The user enters the topic of interest after visiting the website, then choose the number of news items to be displayed using the number slider. We obtain the URL of the most recent news of the chosen topics by utilising the Google RSS webcrawler. then the website's url is supplied to newspaper3k api, which downloads the website's content as XML files and sends it to .nlp() function to summarise. The user can read the news that has been summarized and article's source URL.Figure 2 shows the working of Google RSS API.
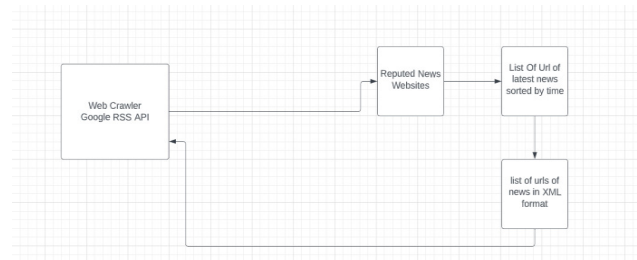

Figure 2. Block Diagram of Google RSS API

Google crawls RSS or Atom feeds for Google News using Feedfetcher. Users of an app or service can request feeds, which Feedfetcher keeps and periodically updates. Following the retrieval of the news article website urls, the list is sent to the summary phase.
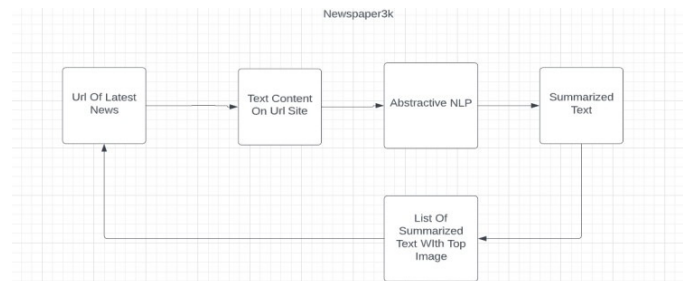

Figure 3. Block Diagram of Summarization using newspaper3k Library

Figure 3 shows how the news is summarized using newspaper3k summarizer. Newspaper3k's.download() function is used to download the content from the url that was received from the feedfetcher. The news is then analysed. The nlp() function is used, which applies a hybrid kind of data summarization. The article's author, image, time, and summarised data are subsequently shown on the user interface (UI).

4. RESULT AND ANALYSIS

The user interface is interactive and provides all of the resources suggested in the problem statement. Because the analyses are accurate, the work is completed quickly and amicably summarization from. The UI also shows the image from the source website of the article .Here is the view of the website after searching for a topic and choosing the number of news to be shown.
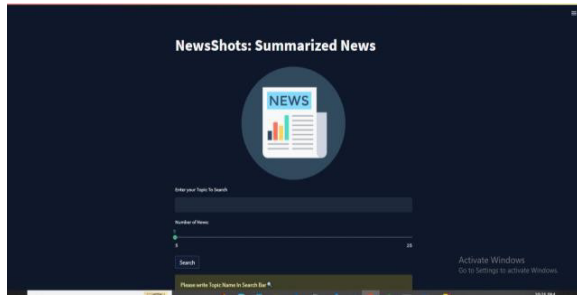

Figure 4. User Interface

As shown in the Figure. 4 the First page of the website is visible and the user interface is exactly same as it should be designed.
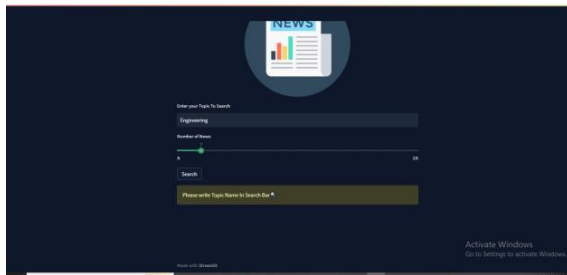

Figure 5. Interface showing the selection of the news and range

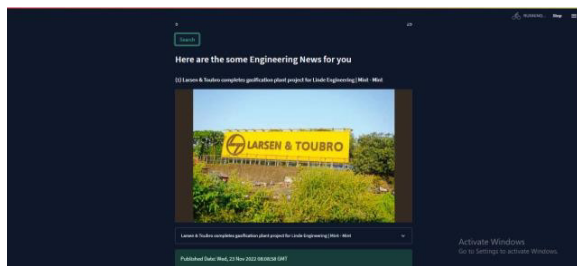Figure 5 shows the interface showing the news selection and range of number of news.

Figure 6. Hypertext Source

Figure 6 shows the snapshot of the summarized text along with the hypertext which leads to the source website.
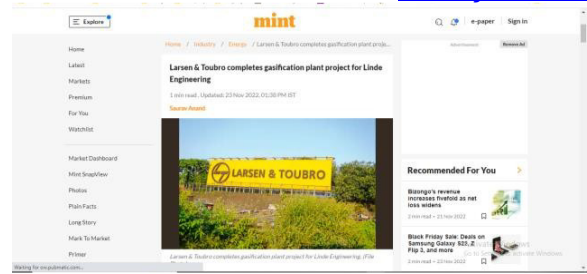

Figure 7. Brief news about the selected article with the main refrence page

Figure 7 shows brief news about the selected article with the main reference page.

6. CONCLUSION AND FUTURE SCOPE

An app for news summarization can let consumers read more news items in less time. Newspaper3k Library's pricing and summary quality makes it the finest option. Since homepage just has one page and numerous ML models have been incorporated into it, flask is the best choice. Reduced costs and improved maintenance were greatly aided by streamlet. We also want to leverage the AWS cloud to deploy this model and put it online for usage by the general public. Such an app would minimise the time and effort needed to increase people's knowledge of their surroundings. People with little to no time would find it to be quite appealing. They can read this while going about their daily business, such as walking or waiting for someone.

The proposed app can be deployed on any cloud platform or can be launched as an application.

References

[1] Usman Malik, Artcle on "Text summarization using NLTK in Python", Stack Abuse, July 2022

[2] Himangi Dani, Pooja Bhople, Hariom Waghmare , Kartik Munginwar and Ankush Patil, "Review on Frameworks Used for Deployment of Machine Learning Model", International Journal for Research in Applied Science & Engineering Technology (IJRASET) Vol(10), Issue(II), Feb. 2022.

[3] Laxmi B. Rananvare P.Venkata Subha Reddy, "Automatic news article summarization" International Journal of Computer Science and Engineering ,Vol(6), Issue(2), pp.230-237,Feb 2018.

[4] Ritwik Mishraa, Tirthankara Gayen, "Automatic losslesssummarization of news article with abstract meaning", Procedia Computer Science,Vol(135), pp.178-185, 2018

[5] Ji Eun Lee, Hyun Soo Park, kyungjoongkim, Jae Chun No, "Learning to predict the need of the summarization of news article", Procedia Computer Science,Vol(24), pp.274-279, 2013

[6] Hujia Yu, Chang Yue and Chao Wang, "News Article Summarization with Attention-based Deep Recurrent Neural Networks",

[7] Sara Tarannum, Piyush Sonar, Aashi Agrawal and Krishnai Khairnar "NLP based Text Summarization Techniques for News Articles: Approaches and Challenges, International

Research Journal of Engineering and Technology, Vol(8), Issue(12), Dec.2021

[8] Yang Liu, Chenggang Zhu and Michael Zeng " End-to-End Segmentation-Based News Summarization Microsoft Cognitive Services Research" , Book-ACL, May 2022.
[9] Apurva D. Dhawale, Sonali B. Kulkarni, Vaishali M. Kumbhakarna, "Automatic Unsupervised Extractive Summarization of Marathi Text Using Natural Language Processing, IOSR Journal of Computer Engineering (IOSR-JCE), Volume 22, Issue 6, PP 21-25 (Nov. – Dec. 2020).
[10] Nikhil S. Shirwandkar, Samidha Kulkarni, "Extractive Text Summarization using Deep Learning", International Conference on Computing Communication Control and Automation (ICCUBEA) 16-18 August 2018.

[11] Alaa Mohamed, Marwan Ibrahim, Mayar Yasser, Mohamed Ayman, Menna Gami, Walaa Hassan, "News Aggregator and Efficient Summarization System", International Journal of Advanced Computer Science and Applications(IJACSA), Volume 11 Issue 6, 2020.
[12] Chen Li, Zhonghu Wei, Yang Liu, Yang Jin, Fei Huang, "Using Facebook Public Posts to Enhance Trending News Summarization", Meta Research, Dec.206

.

/