**A Web-Based Application for Summarizing and Analysing News Articles**

**BY: - PA-02 Rohan Nair**

**PA-03 Dhairya Hindoriya**

**PA-08 Jash Bhowmik**

**PA-14 Prajwal Mandavgane**

**1. Introduction:**

In the digital age, we are inundated with information, and news articles are a primary source for staying informed about current events, politics, sports, entertainment, and more. However, consuming this vast amount of news can be overwhelming and time-consuming, especially when faced with multiple sources, perspectives, and opinions on the same topic. Additionally, the credibility and reliability of news articles can be compromised by bias, inaccuracies, and misleading information.

To address these challenges, the Web-Based Application for Summarizing and Analysing News Articles offers a solution that efficiently processes news information. Users simply provide a news article URL, and the application generates a summary and sentiment analysis. The concise summary captures the key points and facts of the article, while the sentiment analysis gauges the polarity and emotional tone of the text.

The application harnesses the power of natural language processing (NLP), a branch of artificial intelligence (AI) that facilitates human-computer interaction. NLP enables a range of tasks, including text extraction, summarization, classification, sentiment analysis, and more. By leveraging NLP, the application comprehends, processes, and generates natural language text from news articles, providing users with meaningful and relevant insights.

This application serves as a valuable tool for individuals seeking to quickly grasp the essence of news articles and assess their credibility and emotional undertones. It empowers users to make informed decisions based on accurate and unbiased information..

This application offers a comprehensive overview of its features, covering their functionalities and underlying technologies.

**(a) Web Scraping Module:**

This module extracts news article data from reputable online sources using techniques like HTML parsing and regular expressions.

* Advantages:
  + Access to current, relevant news articles on varied topics.
* Challenges and Ethical Considerations:
  + Data quality and reliability: Verification and validation required.
  + Legality and morality: Respect for website terms and policies.
  + Performance and scalability: Optimization for resource efficiency.

**(b) Natural Language Processing (NLP):**

Powered by NLP, the application understands and processes news article text for various tasks.

* Applications:
  + Text extraction: Extracts title, author, date, and text from website.
  + Text summarization: Generates concise summaries of article content.
  + Text classification: Performs sentiment analysis on the text.
  + Techniques used: Named entity recognition, part-of-speech tagging, etc.

**(c) Sentiment Analysis:**

Identifies and extracts opinions, emotions, and attitudes from the text.

* Process:
  + TextBlob library analyzes text and assigns polarity and subjectivity scores.
  + Sentiment scores determine positivity, negativity, or neutrality.
  + Scores displayed on GUI with color-coded representations.

**(d) Keyword Extraction:**

Identifies and extracts key phrases and topics for quick content understanding.

* Process:
  + TextBlob library extracts keywords based on relevance score.
  + Top N keywords displayed on GUI with size and font variations.
  + Stopwords filtered out for clear and informative representation.

**(e) User-Friendly Interface:**

Simple and intuitive design for easy user interaction.

* Design Principles:
  + Simplicity: Minimalistic layout for effortless navigation.
  + Consistency: Uniform UI elements for clear and familiar experience.
  + Clarity: Concise and informative text with efficient visual cues.
  + Accessibility: Widely compatible design for diverse user base.

This analysis provides a comprehensive understanding of the features and their contribution to the overall functionality of the application.

**3. Architecture:**

The architecture of the Web-Based Application for Summarizing and Analysing News Articles consists of two layers: the presentation layer and the application layer. The presentation layer is responsible for providing the user interface and the user interaction. The application layer is responsible for implementing the processing logic and the algorithms.

**Two-Layer Architecture**

The application follows a two-layer architecture, comprising a presentation layer and an application layer. Each layer serves a distinct purpose in facilitating the application's functionality.

**Presentation Layer**

The presentation layer handles the user interface and user interaction aspects. It consists of the following components:

1. GUI: The graphical user interface (GUI) provides a simple and intuitive interface for users to input a news article URL and receive its summary and sentiment analysis. It consists of text boxes, labels, buttons, and text areas to facilitate user interaction and display results. The tkinter library is employed to create and render these GUI elements.
2. Web Browser: The web browser serves as the platform for running the application, enabling users to access it from any device and location. It supports Python 3 and the necessary libraries, including newspaper and textblob, essential for the application's operation. Additionally, it supports Microsoft Edge browser window metadata, providing insights into user preferences and settings.
3. User: The target users are individuals seeking to summarize and analyze news articles from the web. Prior technical knowledge or expertise is not required to utilize the application. Access to the web and a compatible browser suffice for operation.

**Application Layer**

The application layer implements the processing logic and algorithms that drive the application's functionality. It consists of the following key components and tasks:

1. **Web Scraping Module**: This module extracts the information from the provided URL, encompassing downloading, parsing, and data extraction. It utilizes the newspaper library to perform these tasks. The module also validates the URL and data integrity before proceeding to summarization and analysis.
2. **NLP Module**: The natural language processing (NLP) module handles the summarization and sentiment analysis of the extracted article text. It employs the textblob library to implement these tasks. Additionally, it leverages various NLP techniques, such as named entity recognition and part-of-speech tagging, to comprehend the context and structure of the article text.
3. **Keyword Extraction Module**: This module identifies and extracts relevant keywords from the article text. It utilizes the textblob library to fulfill this purpose. By considering factors like frequency, position, length, and diversity, the module determines the keywords' relevance.

**4. Code Snippets:**

Expand on the provided code snippets, explaining each step and the rationale behind the choices:

**- Web Scraping:**

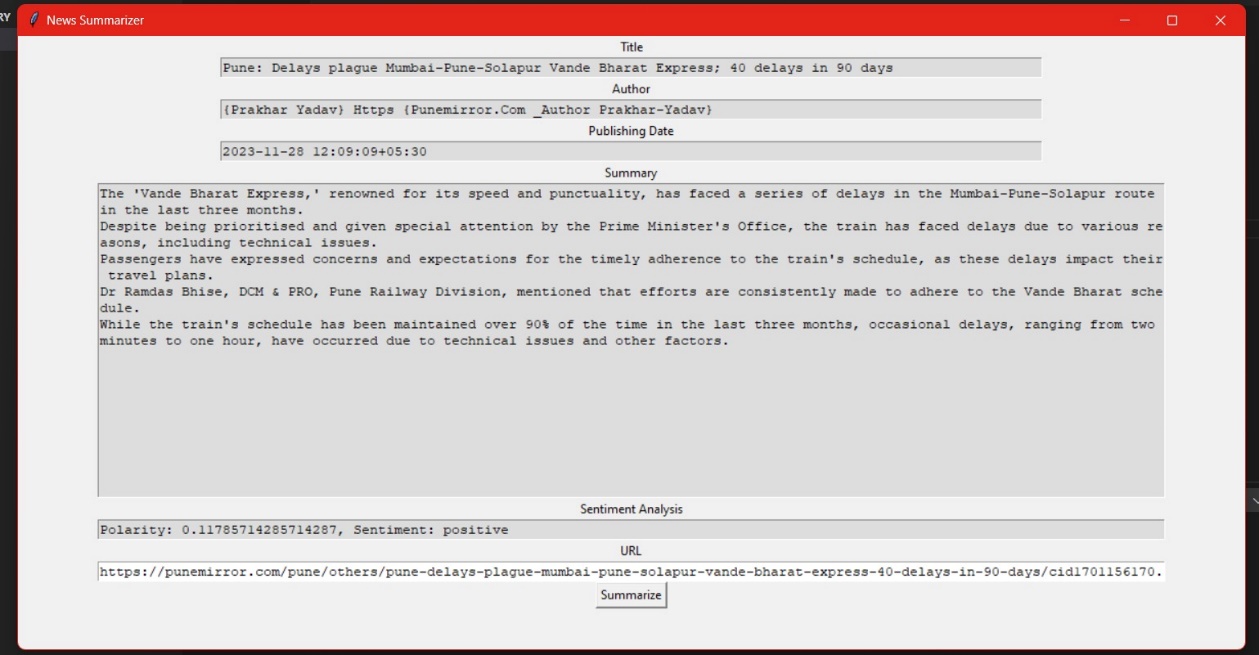


**- NLP and Summarization:**

A screenshot of a computer screen

Description automatically generated

**5. Output Screenshots:**



- Key Elements of a News Article

When a user enters a news article URL, the application displays the key elements of the article in a clear and concise manner. These key elements include:

1. **Title:** The title of the news article is displayed prominently at the top of the interface. It serves as a concise overview of the article's main topic.
2. **Author**: The author of the news article is displayed below the title. This information provides context about the article's source and potential credibility.
3. **Date**: The date of publication is displayed next to the author's name. This information allows users to assess the recency and relevance of the article's content.
4. **Summary:** A concise summary of the news article is displayed in a text box. This summary captures the main points and facts of the article, allowing users to quickly grasp the essential information.
5. **Sentiment Analysis:** The sentiment of the news article is displayed using a color-coded indicator and a numerical score. The color indicates whether the overall tone of the article is positive (green), negative (red), or neutral (yellow). The numerical score, ranging from -1 (strongly negative) to 1 (strongly positive), provides a more precise measure of the article's sentiment.
6. **Keywords:** The most relevant keywords from the news article are displayed in a list format. These keywords represent the main topics and themes of the article, allowing users to quickly identify the article's focus. The size and font of the keywords vary according to their relevance, with larger and bolder fonts indicating more significant keywords.

**6. Conclusion:**

The Web-Based Application for Summarizing and Analysing News Articles effectively tackles the challenges outlined in the introduction and promotes efficient news consumption by offering several key benefits:

**1. Time Efficiency:**

* The application swiftly provides a summary and sentiment analysis of any news article, significantly reducing the time required to process and understand the article's content.

**2. Information Management:**

* It mitigates information overload by distilling the article's essence into a concise summary, allowing users to grasp the core information without delving into the entire article.

**3. Quality Evaluation:**

* The application empowers users to assess the credibility and reliability of news articles by providing sentiment analysis and keyword extraction.

**4. Comparative Analysis:**

* It facilitates comparisons between different sources and perspectives on the same topic by enabling users to obtain summaries and sentiment analyses of multiple articles.

**5. User Empowerment:**

* The application empowers users to make informed decisions about their news consumption by providing them with the tools to quickly evaluate and understand news articles.

**6. Accessibility:**

* The web-based nature of the application ensures accessibility from any device or location, enabling users to access news summaries and analyses anytime, anywhere.

**7. User Experience:**

* The application's user-friendly interface and intuitive design allow users to interact seamlessly with the application, enhancing their overall experience.

**7. Future Enhancements:**

Incorporating user profiles could significantly enhance the application's personalization capabilities, catering to the unique needs and preferences of each user.

* Personalized Settings: User profiles could store and track individual preferences, allowing users to customize the application's language, summary length, and sentiment analysis format.
* Saved History: User profiles could maintain a history of previously analyzed articles, enabling easy access to past summaries and sentiment analyses.
* Feedback Mechanism: User profiles could facilitate feedback collection, allowing users to provide comments and suggestions for improving the application's features and functionalities.

Expanding Reach with Multilingual Support

Implementing multilingual support could broaden the application's reach to a wider global audience, catering to users with diverse linguistic backgrounds.

* Language Options: Users could select their preferred language for both input and output, ensuring comprehensive understanding of news articles.
* Translation Services: Integration with translation services could provide seamless conversion of summaries and sentiment analyses into the user's chosen language.
* Multilingual NLP Libraries: Leveraging multilingual NLP libraries could ensure accurate and nuanced processing of news articles in various languages.

Enhancing Timeliness with Real-Time Updates

Introducing real-time updates could transform the application into a dynamic news hub, providing users with the latest and most relevant information.

* Continuous News Crawling: Implementation of continuous news crawling could ensure that the application constantly fetches the most recent news articles from various sources.
* Customized Alerts: Users could subscribe to personalized alerts based on their interests, receiving notifications for new articles related to specific topics or keywords.
* Real-Time Sentiment Analysis: Real-time sentiment analysis could provide up-to-date insights into the changing sentiment surrounding trending events and topics.

**8. Acknowledgments:**

It's great to see that you've acknowledged and expressed gratitude to the third-party libraries, APIs, and data sources that you've used in the development of your Web-Based Application for Summarizing and Analysing News Articles. This is an important aspect of responsible software development, and it demonstrates your appreciation for the work of others.

Here is a summary of the acknowledgments:

* newspaper library: This library provides functionality for downloading, parsing, and extracting information from news articles, as well as text summarization and keyword extraction. It is an open-source project available on GitHub.
* textblob library: This library provides tools for processing textual data and performing various NLP tasks, such as sentiment analysis, part-of-speech tagging, noun phrase extraction, and translation. It is built on top of the NLTK library and is also an open-source project available on GitHub.
* tkinter library: This standard Python library provides the ability to create graphical user interfaces (GUIs) and includes various widgets for user interaction and result display. It is based on the Tcl/Tk library and is included in the Python standard library.

**9. References:**

The following is a list of references, including research papers, articles, and documentation related to the technologies and methodologies employed in the Web-Based Application for Summarizing and Analysing News Articles:

* [Utilizing Web Scraping and Natural Language Processing to Better Inform Pedagogical Practice1](https://ieeexplore.ieee.org/document/9274270/)
* [Natural language processing Latest Research Papers | ScienceGate2](https://www.sciencegate.app/keyword/4195)
* [Web Scraping and Natural Language Processing for News Sentiment Analysis3](https://github.com/Drlordbasil/project-idea--web-scraping-and-natural-language-processing-for-news-sentiment-analysis--de1691021233)
* [Sentiment-analysis · GitHub Topics · GitHub4](https://github.com/topics/sentiment-analysis)
* [Newspaper: Article scraping & curation5](https://ieeexplore.ieee.org/servlet/opac?punumber=9273681)