
REAL-TIME NEWS AGGREGATOR

Bhanuprakash BR*¹

*¹Student, Dept. Of Master Of Computer Application, PES Institute Of Technology And Management, Shivamogga, Karnataka, India.

ABSTRACT

"Real-Time News Aggregator" is a user-friendly Streamlit application designed to bring you the latest news in a concise and accessible format. Using Google News RSS feeds, it aggregates and summarizes news articles so you can stay informed effortlessly. This app leverages powerful Python libraries like Beautiful Soup for web scraping, newspaper3k for article processing, and NLTK for natural language processing, ensuring a smooth and efficient news delivery experience. With "News Quest," you can easily fetch the latest top news, explore articles across various categories such as world news, business, technology, and more, or search for news based on your interests. Each news article includes a title, summary, image, publication date, and a link to the full article, giving you a comprehensive snapshot of the news landscape. The interface is designed to be interactive and intuitive. You can select your preferred news category, specify the number of articles to display using a slider, and search for specific topics to find relevant news articles. The application maintains visual consistency by showing a placeholder image if an article lacks one. The app's execution flow involves setting up Streamlit configurations, downloading necessary NLTK data, and fetching and processing data based on your inputs. It then displays the articles efficiently, making "News Quest" a powerful tool for accessing summarized news content quickly and effectively, catering to a wide range of informational needs.

Keywords: Streamlit, News Aggregation, Google News RSS Feeds, Natural Language Toolkit (NLTK), Newspaper3k.

I. INTRODUCTION

In today's fast-paced digital world, staying updated on current events is crucial, yet the sheer volume of available news can be overwhelming. "News Quest: Personalized News Portal" addresses this challenge by providing a streamlined and personalized news consumption experience. Built on the robust Streamlit framework, this web application transforms how individuals access and interact with news content. By leveraging powerful Python libraries such as BeautifulSoup, `newspaper3k`, and NLTK, "News Quest" offers concise and relevant news summaries sourced from Google News RSS feeds. With a strong focus on user-centric design and functionality, this project simplifies navigating the vast sea of news articles, empowering users to stay informed with ease and efficiency.

A Streamlit application called "Real-Time News Aggregator" gathers and condenses news stories from Google News RSS feeds into an intuitive user interface. It offers a simplified method of staying updated by utilizing Python modules such as NLTK for natural language processing, newspaper3k for article processing, and BeautifulSoup for web scraping. Users can fetch the latest top news, explore articles by various categories (such as world news, business, technology, etc.), and search for news based on custom topics. Each displayed article includes a title, summary, image, publication date, and a link to the full article, offering a comprehensive news snapshot. The interactive and intuitive interface lets users select news categories, specify the number of articles to display, and search for specific topics, ensuring a consistent visual display with placeholder images if needed. The execution flow involves setting up Streamlit configurations, downloading necessary NLTK data, and processing and displaying articles efficiently. Overall, "News Quest" is a powerful tool for accessing summarized news content quickly and efficiently, catering to diverse informational needs.

II. RELATED WORK

The paper [1] presents a general framework for first story detection (FSD) that utilizes entities and their relations. By leveraging entity-based features and relations extracted from news articles, the proposed framework aims to identify emerging stories effectively. Experimental results demonstrate effectiveness of the framework in detecting first stories with high accuracy. The paper [2] discusses the challenges and trends in detecting events in online social networks (OSNs). It provides definitions and insights into event detection and

explores various techniques for identifying events from OSN data. The paper offers valuable perspectives on the evolving landscape of event detection in the context of social networks. The paper [3] aims on the detection and delineation of events and sub-events in social networks. It proposes a framework that leverages network analysis techniques to identify events and their associated sub-events from social media data. The framework demonstrates promising results in accurately detecting and delineating events in real-world scenarios. The paper[4] presents an automatic text-summarization approach of news articles. The proposed method utilizes techniques such as sentence extraction and ranking to generate concise summaries of news content. Experimental results indicate that the approach effectively summarizes news articles while preserving essential information. The paper [5] introduces a method for aspect-based sentiment analysis using BERT, a pre-trained language model. By constructing auxiliary sentences, the proposed approach effectively captures aspect-level sentiment from text data. Experimental results demonstrates the effectiveness of the method in accurately identifying sentiment across different aspects. The paper[6] presents Twittermonitor, a system for trend detection over the Twitter stream. By analyzing tweet data in real-time, Twittermonitor identifies trending topics and events from the Twitter stream. When it comes to tracking conversations on social media sites and identifying new trends, the system performs admirably. A poorly supervised learning strategy for Twitter fake news identification is presented in the publication [7]. The suggested strategy accurately detects fake news tweets by utilizing metadata and user interactions, which are weak supervision signals. The strategy is successful in reducing the dissemination of false information on social media, as demonstrated by the experimental results. Using a unified text-to-text Transformer model, the research [8] investigates the boundaries of transfer learning. The study explores the effectiveness of transfer learning across domains by optimizing pre-trained models on a range of natural language processing. The results shed light on the capabilities and limitations of transfer learning in NLP applications. The paper [10] introduces News Monitor, a framework for querying news in real-time. By leveraging advanced indexing and retrieval techniques, News Monitor enables users to query and access news articles efficiently. The framework offers valuable insights into real-time news monitoring and retrieval systems, facilitating timely access to relevant news content.

III. SYSTEM DESIGN

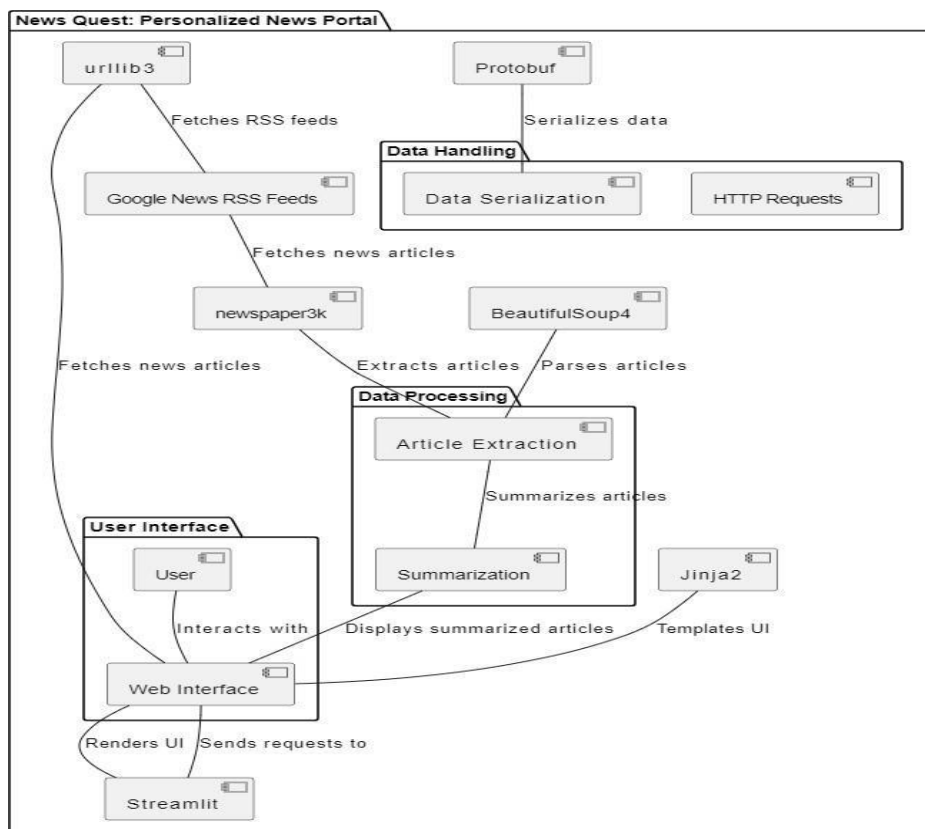


Fig 1: System Architecture

The system architecture of the "Real-Time News Aggregator" is built to be robust and scalable, ensuring efficient news aggregation and summarization. At its heart is Streamlit, a Python framework that provides an interactive and user-friendly web interface. The system integrates with Google News RSS feeds to dynamically fetch news articles, which are then processed using the `newspaper3k` library for content extraction and summarization. Key components like BeautifulSoup4 for web scraping and Protobuf for data serialization ensure seamless data processing and transmission.

The architecture is designed with modularity in mind, making it easy to add new features and updates. The use of Jinja2 for template rendering and urllib3 for handling HTTP requests enhances the system's performance and reliability. This meticulous design facilitates streamlined news consumption while accommodating future enhancements and scalability, ensuring that users have an efficient and up-to-date news portal.

IV. DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) is a visual tool used to illustrate the flow of data within a system, showing how information moves from input to output through various processes. It helps in understanding a system's functionality by mapping out data sources, destinations, and storage points. DFDs are usually divided into levels to offer different degrees of detail, starting with Level 0 for a high-level overview and progressing to more detailed levels like Level 1 and Level 2, which break down the system into specific components and processes. By depicting data interactions, DFDs help identify potential inefficiencies and ensure all data-related requirements are met, making them essential for systems analysis and design.

News Quest - Level 1 DFD

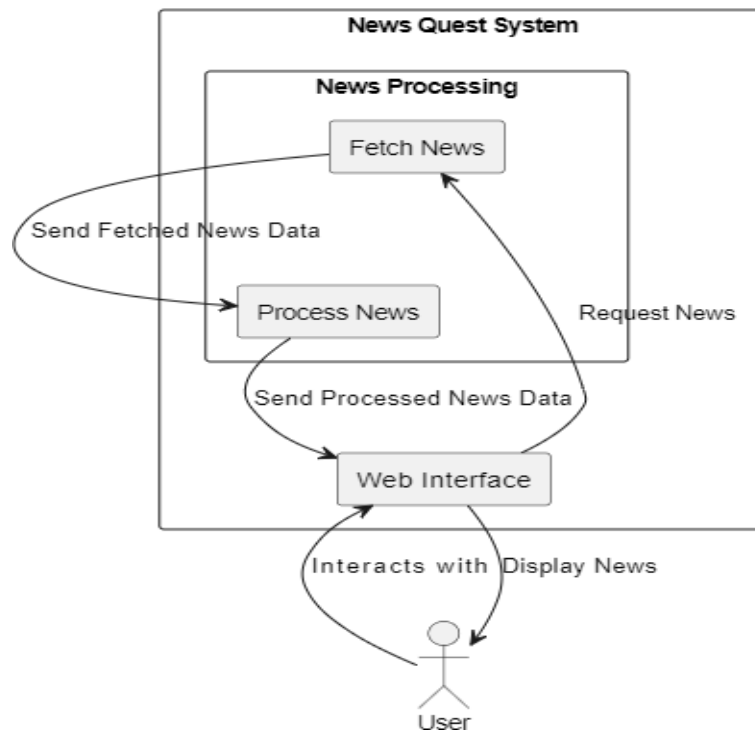


Fig 2: Level 1 DFD

The Level 0 DFD, also known as the Context Diagram, provides a high-level overview of the News Quest system, showing how it interacts with external entities. In this case, the external entity is the User, who interacts with the News Quest system to receive summarized news articles. The system fetches news from Google News RSS Feeds and processes this information to deliver concise articles to the user. This diagram defines the system's boundaries and highlights the primary interactions between the system and its external environment, setting the stage for a deeper understanding of the system's functionality.

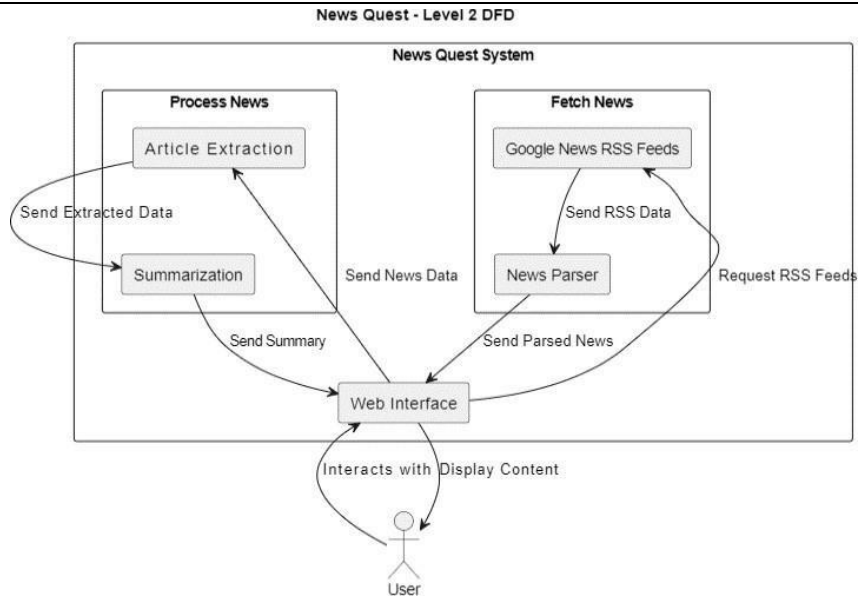


Fig 3: Level 2 DFD

The Level 1 Data Flow Diagram provides a detailed view of the internal processes within the News Quest system. It breaks down the system into key processes, such as Fetching News, Processing News, and the Web Interface through which users interact with the system. The Fetch News process initiates by retrieving RSS feeds from Google News, which are then parsed by the News Parser. This ensures that the system gathers the latest news updates efficiently. In the Processing News section, Article Extraction and Summarization processes work to condense the news articles into concise summaries. This step is crucial in delivering streamlined information to users. Lastly, the Web Interface manages user requests and displays the summarized news content, providing an intuitive platform for users to access and navigate through news articles.

This level of detail in the Level 1 DFD helps to comprehend the main functional components and the flow of data within the News Quest system, highlighting how each process contributes to delivering an effective news aggregation and summarization service.

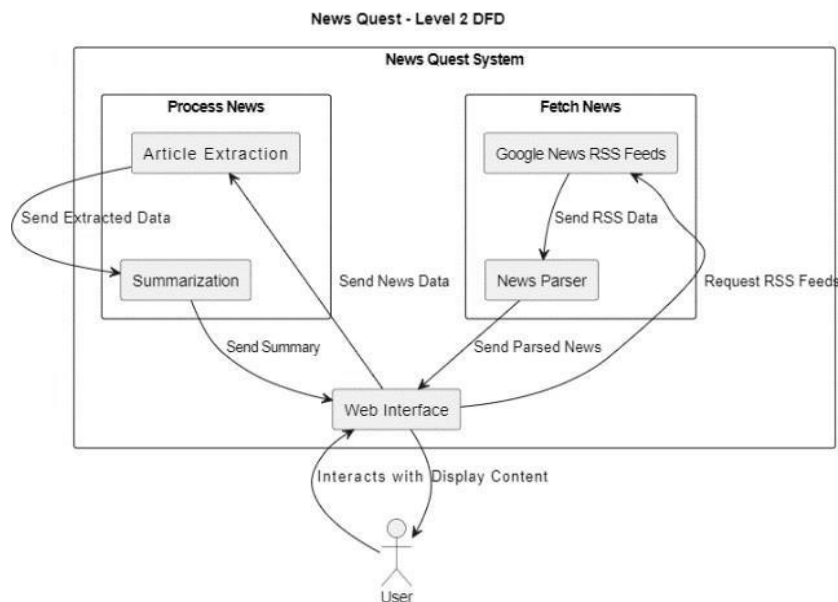


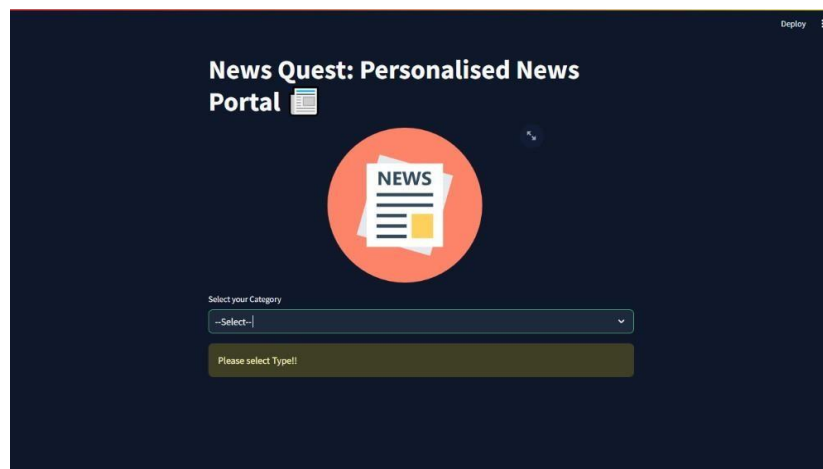
Fig 4: Level 2 DFD

The Level 2 Data Flow Diagram of the News Quest system intricately maps out the sequential processes involved in fetching and processing news articles. Users interact through Web Interface to request news content, initiating the Fetch News process. This procedure receives RSS data from Google News RSS Feeds, parses it using the News Parser, and then returns the results to Web Interface. The parsed data is then sent by Web Interface to the Article

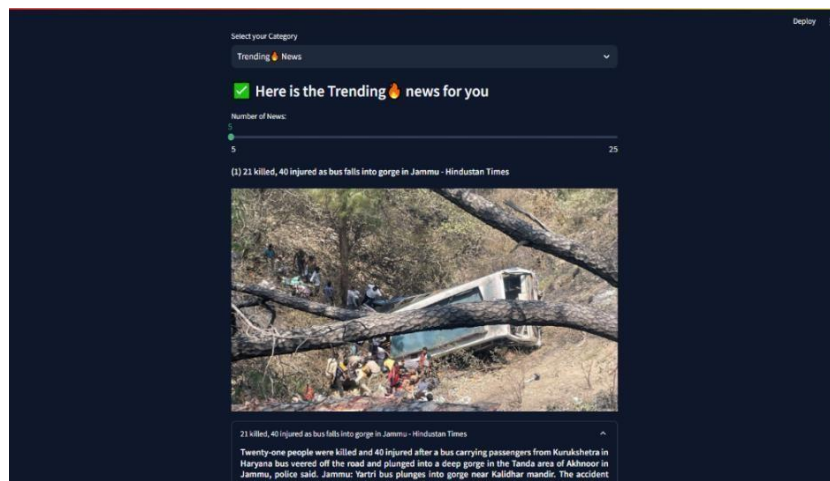
Extraction module, which uses it to extract pertinent data from the articles. The user is ultimately supplied with succinct summaries of the extracted data via Web Interface, following the summary module's summary process. A complete comprehension of the operational flow within the News Quest environment is ensured by this comprehensive graphic, which clearly illustrates the data flow through each stage of the system.

V. RESULTS AND DISCUSSIONS

This project focuses on creating a summarized news portal by extracting and processing data from news articles. It employs web scraping techniques to gather data, which is then processed to extract key information like titles, summaries, and publication dates using Python libraries such as BeautifulSoup and newspaper3k. The goal is to offer users a user-friendly interface for efficient access to news articles. The portal supports various news categories and enables users to search for specific topics, emphasizing convenience and keeping users informed about current events effectively.



(a)

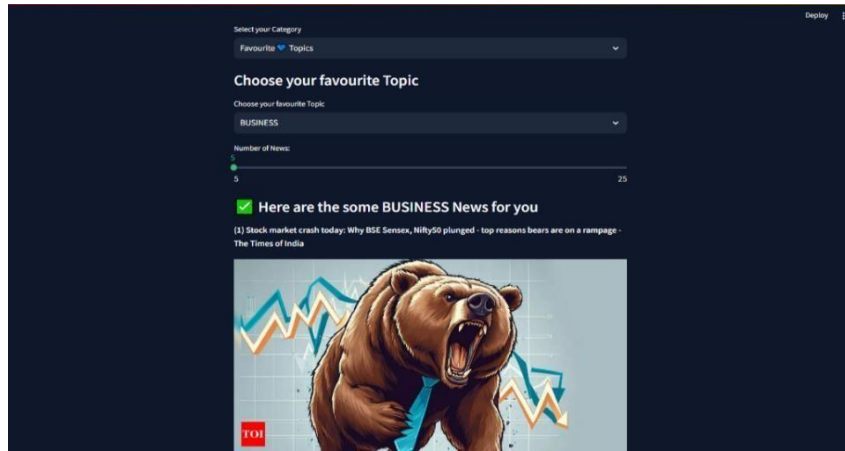


(b)

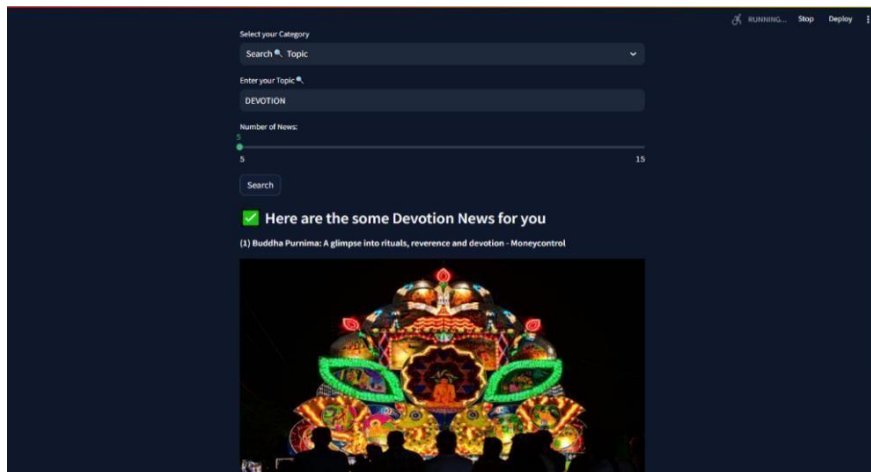
Fig 5: (a) News Quest- Home Page (b) News Quest- Trending News Results Page

In the above figure (a) it shows the streamlit applications results of News Quest- Home Page from where we can select the Trending news, Favourites or Search news options.

The above figure(b) shows the results of News Quest – Trending News Page section, which has the option for choosing the number of news to be displayed.



(c)



(d)

Fig 6: (c) News Quest- Favorite Topics Results Page (d) News Quest- Search Topic Results Page

The above figure(c) shows the result of Favourite Topics, business is selected as favourite topic rather than business we have 7 other favourite topics from which we can fetch news of user’s interest.

The above figure(d) shows the results of the Search Topic Page where the user can search the news according to their interest by entering the topic of their own. Here it is selected as Devotion.

VI. CONCLUSION

The development of "News Quest: Personalized News Portal" showcases the effective integration of modern technologies to enhance user engagement and streamline information accessibility. Through the strategic use of Python libraries like BeautifulSoup, newspaper3k, NLTK, and Streamlit, we have engineered a robust platform capable of aggregating and summarizing news articles with efficiency. This portal addresses the challenge of information overload by delivering concise and relevant news updates tailored to individual user preferences.

Rigorous testing procedures have been implemented to ensure the reliability and satisfaction of our users. Looking ahead, we are committed to further enhancing personalization and interactivity, promising continuous improvements in how news is consumed and accessed. "News Quest" stands as a scalable, user-friendly solution designed to keep users informed amidst the rapid pace of the digital age, reflecting our dedication to innovation and advancing news consumption experiences.

VII. REFERENCES

- [1] Panagiotou, N.E.; Akkaya, C.; Tsioutsoulou, K.; Kalogeraki, V.; Gunopulos, D. A General Framework for First Story Detection Utilizing Entities and their Relations. *IEEE Trans. Knowl. Data Eng.* 2020, 33, 3482–3493.

-
- [2] Panagiotou, N.; Katakis, I.; Gunopulos, D. Detecting events in online social networks: Definitions, trends and challenges. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 42–84.
- [3] Saravanou, A.; Katakis, I.; Valkanas, G.; Gunopulos, D. Detection and Delineation of Events and Sub-Events in Social Networks. In *Proceedings of the IEEE 34th International Conference on Data Engineering (ICDE)*, Paris, France, 16–19 April 2018; pp. 1348–1351.
- [4] Sethi, P.; Sonawane, S.; Khanwalker, S.; Keskar, R. Automatic text summarization of news articles. In *Proceedings of the 2017 International Conference on Big Data, IoT and Data Science (BIGDATA)*, Pune, India, 20–22 December 2017; pp. 23–29.
- [5] Sun, C.; Huang, L.; Qiu, X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the NAACL-HLT 2019*, Minneapolis, MN, USA, 2–7 June 2019; pp. 380–385.
- [6] Mathioudakis, M.; Koudas, N. Twittermonitor: Trend detection over the twitter stream. In *Proceedings of the 2010 International Conference on Management of data (SIGMOD '10)*, 6–10 June 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1155–1158.
- [7] Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 2020, 21, 1–67.
- [8] Saravanou, A.; Panagiotou, N.; Gunopulos, D. News Monitor: A Framework for Querying News in Real Time. In *Proceedings of the 43rd European Conference on IR Research, ECIR 2021*, Virtual Event, 28 March–1 April 2021; pp. 543–548.
- [9] Allan, J.; Lavrenko, V.; Malin, D.; Swan, R. Detections, Bounds, and Timelines: Umass and tdt-3. Available online: <http://ciir.cs.umass.edu/pubfiles/ir-201.pdf> (accessed on 20 December 2021).