News Aggregation And Summarisation

Cheah Weng Hoe
Asia Pacific University of Technology and Innovation
Kuala Lumpur, Malaysia
tp055533@mail.apu.edu.my

Mafas Raheem
Asia Pacific University of Technology and Innovation
Kuala Lumpur, Malaysia
raheem@apu.edu.my

Nirase Fathima Abubacker
Asia Pacific University of Technology and Innovation
Kuala Lumpur, Malaysia
nfathima.abubacker@apu.edu.my

***Abstract*—The proliferation of online content, spanning news articles, blogs, and social media updates, has precipitated a surge in information dissemination. This phenomenon presents significant challenges for individuals and organizations alike in maintaining currency with timely and accurate information. Consequently, news aggregation and summarization systems have emerged as indispensable tools to address this issue by gathering and consolidating articles from diverse sources into a centralized platform. The primary objective of this research is to comprehensively examine techniques in news aggregation and summarization. Additionally, it aims to evaluate the strengths and weaknesses of existing news aggregators such as Google News and Apple News+. Moreover, it explores challenges such as media bias and content quality, considering methodologies proposed by previous researchers. Furthermore, this study identifies emerging trends such as multimodal summarization, ethical considerations, and cross-platform integration, which promise to enhance the efficacy of news aggregators significantly.**

Keywords—*News Aggregation, News Summarisation, Google News, Apple News, Media Bias, Content Quality*

I. INTRODUCTION

In today's interconnected world characterized by globalization and rapid technological advancements, the sheer volume of online content, ranging from news articles to blogs and social media updates, has grown exponentially (Ali & Dahake, 2023). This surge in information poses a significant challenge: individuals, businesses, and organizations often struggle to access timely and accurate information essential for staying informed and making informed decisions. The consequence of inadequate access to up-to-date information is a heightened risk of making outdated decisions, which can have serious implications. Manually sifting through the vast array of news articles and updates is not only impractical but also time-consuming. This inefficiency underscores the critical need for news aggregation and summarization systems (Coddington, 2020). These systems play a pivotal role in addressing the challenge by automatically gathering, condensing, and presenting relevant information from diverse sources. By leveraging algorithms and artificial intelligence, these systems help users sift through the noise, providing concise summaries and insights that enable quicker comprehension of current trends and events. Thus, in the era of globalization and modernization, where information overload is a constant concern, effective news aggregation and summarization systems serve as indispensable tools. They empower users to navigate the vast digital landscape efficiently, ensuring they are equipped with the knowledge needed to make informed decisions in a timely manner.

Based on the definition provided by Bhujbal et al. (2023), a news aggregator is described as a platform or application that collects news articles from various sources and presents them in a centralized location. Typically, such systems include a built-in news summarization feature that generates concise versions of news articles while preserving their meaning and key elements (Mohamed et al., 2020). Therefore, it can be argued that this technology offers significant advantages over traditional methods of consuming news. This assertion is supported by research conducted by Mohite et al. (2023), which suggests that news aggregators with personalized features have the potential to greatly enhance user engagement and overall satisfaction with news content.

The primary aim of this study is to conduct a thorough investigation into news aggregation and summarisation, focusing on several key research inquiries. Firstly, the study aims to delineate the techniques utilized by previous researchers in this domain, providing a comprehensive overview of existing methodologies. Secondly, it seeks to identify and analyse the challenges and limitations inherent in current approaches to news aggregation and summarisation. Thirdly, the study endeavours to propose and evaluate potential techniques that could effectively mitigate these challenges and enhance current methodologies. Lastly, it aims to pinpoint the research gaps that persist from past studies, identifying areas where further investigation and innovation are needed to advance the field of news aggregation and summarisation. Through addressing these inquiries, this study aims to contribute new insights and advancements to the field, ultimately improving the efficacy and applicability of automated news processing technologies.

II. BACKGROUND & RELATED WORK

In this section, the researcher will explore the literature from previous studies. This will encompass a review of diverse news aggregation methods, summarization techniques, and algorithms.

A. BACKGROUND

According to Sanderson & Croft (2012), the emergence of news aggregators can be traced back to the evolution of information retrieval systems developed in the mid-20th century. During this period, pioneering researchers

introduced several innovative initiatives, including the establishment of databases and indexing systems. These systems enable users to store and access textual data from various sources. Consequently, they laid the foundation for subsequent advancements, particularly in automated information processing and retrieval.

In the late 20th century, the emergence of the Internet ushered in a new era of digital news consumption. This transformation stemmed from the rapid growth of online news websites and social media platforms, allowing users unprecedented ease in accessing diverse global information (Gaskins & Jerit, 2012). This phenomenon significantly contributed to the democratization of information and the decentralization of news distribution channels, empowering users with real-time access to updated news.

In the early 21st century, the advent of real-time news streams signified a paradigmatic shift in how news is consumed. According to Huang et al. (2015), these streams of real-time data are characterized by continuous dissemination of news articles, updates, and notifications aimed at providing users with the latest information on unfolding global events. Similarly, Hasan et al. (2019) argued that social media platforms, such as Twitter, played a crucial role in popularizing real-time news streams by allowing users to engage with people, communities, and topics tailored to their interests. These connections serve to keep users informed about the latest updates on specific news stories.

In contrast, Coddington's (2020) research underscores the importance of aggregation in contemporary journalism when compared to traditional reporting. While some researchers have argued that aggregation is less effective than traditional reporting, Coddington (2020) contends that both methods play crucial roles in gathering and presenting information.

In this modern age, technological advancements, especially in artificial intelligence, have greatly bolstered the refinement of complex algorithms and tools (Mohite et al., 2023). These advances have enabled the automated extraction, classification, and summarization of news articles from diverse sources, enhancing the efficiency and effectiveness of information retrieval.

B. RELATED WORK

In this section, the researcher delves into the existing literature from previous studies. Various news aggregation and summarization techniques and algorithms are discussed.

1) News Aggregation

Thorne et al. (2018) lay the groundwork for similar systems in this field by proposing a fact extraction and verification system based on "Document Retrieval, Sentence Selection, and Recognising Textual Entailment (RTE)." Other researchers have compared the effectiveness of transformer-based models to conventional baseline models (Chen et al., 2019a; Chen et al., 2019b; Gretz et al., 2020). Carlebach et al. (2020) further demonstrate the capability of transformer-based models in this domain. Their implemented system includes topic modeling, semantic clustering, hypothesis extraction, and textual inference techniques to identify opinions expressed in news articles. The results show relative improvements ranging from 10-23% over baseline models. This study emphasizes the importance of model selection by showing a fivefold reduction in loss when using a compact T5 model for premise extraction compared to a sizable BART model, highlighting the efficiency benefits of careful model selection. Additionally, Carlebach et al. (2020) and Gretz et al. (2020) highlight the utility of pre-trained BERT-based models for textual entailment in identifying diverse viewpoints. This indicates the potential for these models to enhance the understanding and classification of entailments as positive, neutral, or negative.

Ali & Dahake (2023) explored the utilization of unsupervised learning techniques in aggregating relevant news articles. Various embedding methodologies, such as Bag-Of-Words, TFIDF, and BERT, were used to transform news articles into vector representations. Clustering algorithms, including hashing algorithms, were employed to categorize interconnected news articles. After comparing different embedding techniques, BERT demonstrated the best result with an accuracy of 82.39%. Although BERT shows superior performance, it is known as the most complex and computationally intensive embedding technique. This complexity may limit its scalability in real-world applications. Therefore, future research should focus on optimizing the efficiency of these embedding techniques or exploring alternative approaches that balance performance and computational cost. Additionally, the researchers claim that clustering algorithms could provide better results in terms of personalization and trend analysis, given that information retrieval and data exploration are conducted effectively.

Clustering algorithms have been utilized to group relevant news articles and enable personalized content delivery. However, further investigation is needed to evaluate their performance and effectiveness in real-world scenarios. Research should examine how various clustering algorithms impact the quality of personalized recommendations, as well as their scalability and adaptability to evolving user preferences and trends. Recently, Dahake & Ali (2023) conducted a follow-up study, gathering relevant business news articles using the approach presented in previous research. However, due to limited access to the journal article, the results of this research remain uncertain.

Conversely, Hamborg et al. (2020) introduced the Matrix-Based model in the field of news aggregation. Generally, the analysis workflow of Matrix-based News Aggregators involves the following steps:

1. **Data Collection:** Crawl news articles from online platforms

2. **Article Extraction:** Extracting important contents within the websites
3. **Grouping:** Find and categorise related articles on the same topic or event
4. **Summarisation:** Provide a summary for the respective news articles
5. **Visualisation:** Presenting topics according to the preferences of the users.

This approach is supported by research conducted by Bhujbal et al. (2023), which emphasizes the importance of web scraping in developing influential online news aggregators. This technique generally involves analyzing the HTML structure of a website to extract desired data, such as article titles and contents. The implemented system allows users to customize their news consumption by selecting preferred websites and receiving curated articles. This customization saves users time and effort and streamlines information gathering in daily routines. However, Chitra et al. (2023) argue that using keyword search algorithms to enable personalization in the news retrieval process would be more efficient. They contend that leveraging keyword search algorithms allows users to access news tailored to their preferences more easily than manually selecting websites. Despite this, limited research has been conducted in this area. Therefore, additional research is required to explore the feasibility and effectiveness of integrating keyword search algorithms into news aggregation systems and to evaluate their impact on user experience and engagement.

The main drawback of past research is the lack of emphasis on quality assurance measures to ensure both the precision and dependability of aggregated news articles. Although much of the research demonstrates promising results, it often lacks empirical data or experimental evidence to prove that the developed models can be effectively implemented in real-world scenarios. While recent research has shown promising advancements in news aggregation techniques, there is still room for improvement in future studies.

TABLE I. COMPARISION TABLE BASED ON NEWS AGGREGATION TECHNIQUES

Techniques	**News Aggregation**			
	Authors	***Main Focus***	***Methodology***	***Key Findings***
Graph-Based Models	(Thorne et al., 2018)	Fact extraction and verification	Document Retrieval, Sentence Selection, Recognising Textual Entailment (RTE)	Proposed system sets foundation for similar system. Emphasis on fact extraction and verification.
Transformer-Based Models	(Carleb ach et al., 2020)	Aggregating news articles	News aggregation, topic modelling, semantic clustering	Transformer-based embeddings shows efficiency gains and improved clustering quality. Relative improvements of 10-23% over baseline models.
Unsupervised Learning Technique	(Ali & Dahake, 2023)	Aggregating relevant news articles	Embedding techniques (Bag-Of-Words, TFIDF, BERT) combined with clustering algorithms	BERT demonstrates superior performance with an accuracy of 82.39% scalability concerns due to complexity.
Matrix-Based Model	Hamborg et al. (2020), (Bhujbal et al., 2023), (Chitra et al., 2023)	Web Scrapping and Personalisation	Data gathering, article extraction, grouping, summarisation, visualisation	Web scrapping enables customisation of news consumption. Personalisation customisations enable users to find news articles based on their preferences.

2) News Summarisation

The realm of automated summarization has garnered attention since the late 1950s. Summarizing a news article can be challenging for an individual with no knowledge of the field. However, this task can be even more difficult for machines (Gulati et al., 2023). According to Gulati et al. (2023), graph-based approaches hold great potential in the process of text summarization. Recently, various graph-based approaches have shown promising results. For instance, research conducted by Mohamed et al. (2020) utilizes the TextRank algorithm to summarize aggregated news articles.

Gulati et al. (2023) stated that the "conventional TextRank model is based on the PageRank algorithm, which leverages textual data as input rather than web pages." The utilization of the TextRank algorithm for summarization has demonstrated promising results in effectively condensing information and providing readers with understandable summaries without needing to refer to the original articles. Moreover, the researchers also compared the results between the TextRank algorithm and the Word Frequency algorithm. The results highlight the superiority of TextRank in providing more efficient and coherent summaries, emphasizing the importance of choosing the appropriate algorithm for text summarization tasks.

Research conducted by Kumar et al. (2020) proposed a hybrid model combining fuzzy logic with graph-based techniques (TextRank and LexRank) and one semantic-based technique (Latent Semantic Analysis). According to Gulati et al. (2023), fuzzy-logic techniques exhibit strong judgment thinking and cope well with ambiguity in an unsupervised manner. The researchers evaluated the proposed system using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE). The resulting ROUGE-1 score of 0.87653 demonstrates that the implemented hybrid model significantly outperforms any single implementation technique.

On the other hand, research by Gulati et al. (2023) introduced an improved version of the TextRank algorithm along with the BM25+ algorithm. They evaluated the model's performance with ROUGE-1, ROUGE-2, and ROUGE-L metrics, showing that the improved TextRank algorithm significantly enhanced the accuracy of the summarization results. Gulati et al. (2023) stated that "among the numerous variations of the similarity matrix implemented, the proposed algorithm performs better than the traditional TextRank and the BM25+ similarity model by 1.654% and 0.413% on the mean of F-score."

Both graph-based and fuzzy logic approaches demonstrate strong performance in text summarization. The main research gap emerging from past research is the exploration of hybrid approaches combining graph-based methods with other techniques, such as deep learning and semantic analysis, to further enhance text summarization performance.

Conversely, some researchers argue that topic modeling approaches, such as generative models, could perform better in this field (Reddy et al., 2022). To support this statement, Reddy et al. (2022) conducted research using topic modeling techniques, particularly Non-negative Matrix Factorisation (NMF) and Latent Semantic Analysis (LSA). The research evaluated the effectiveness of NMF and LSA models in their ability to identify hidden data within news articles. Experimental results demonstrated that LSA outperformed NMF in terms of mean coherence values and execution time efficiency. Langote & Gawande (2023) agree that the LSA algorithm could further enhance performance by integrating other Natural Language Processing (NLP) techniques. This is because NLP plays an important role in removing stop words, punctuation, and other irrelevant features.

Similar research was conducted by Mishra et al. (2023). In this study, the researchers leveraged NLP-based text summarization to address the issue of information overload. The main technique introduced was the utilization of text sentiment analysis on the news articles. The proposed system showed promising results in analysing the sentiment and emotions expressed in news articles. This analysis could provide valuable insights into public opinion, trends, and reactions to news events, contributing to a better understanding of the overall sentiment in the news articles.

In summary, further research is needed to compare the performance of graph-based models and topic modelling approaches to determine the most effective method for handling various types of content. Additionally, previous studies have suggested integrating text summarization with sentiment analysis to gain valuable insights into public opinion and reactions to news events. However, there has been limited research on how sentiment analysis affects the quality and relevance of summarized news. While existing research has made significant contributions to this field, there is still ample opportunity for improvement. Future studies should explore new avenues to enhance the accuracy and efficiency of the summarization process, especially in real-world applications.

TABLE II. COMPARISION TABLE BASED ON NEWS SUMMARISATION TECHNIQUES

Techniques	**News Aggregation**			
	Authors	***Main Focus***	***Methodology***	***Key Findings***
Graph-Based Model	Mohamed et al. (2020), (Gulati et al., 2023)	Text summarisation using TextRank Algorithm	Utilisation of TextRank algorithm for summarisation.	TextRank algorithm provides effective and coherent summaries. Modified TextRank algorithm improves summarisation accuracy
Hybrid Model	(Kumar et al., 2020), (Gulati et al., 2023)	Text summarisation using TextRank Algorithm Combining fuzzy logic with graph-based techniques	Hybrid model combining TextRank, LexRank, LSA and fuzzy logic	Hybrid model outperforms individual techniques in terms of summarisation accuracy
Topic Modelling	Topic Modelling (Reddy et al., 2022), (Langote & Gawande, 2023), (Mishra et al., 2023)	Non-negative Matrix Factorisation (NMF), Latent Semantic Analysis (LSA), Sentiment Analysis	Evaluation of NMF and LSA models for identifying hidden data Integration of LSA with NLP techniques Utilisation of Sentiment Analysis	LSA outperforms NML in terms of coherence values and execution time efficiency. Sentiment analysis provides insights into public opinion and reactions towards news events.

III. APPLICATION & USE CASES

In this section, the researchers delved into various existing news aggregators to discern the strengths and weaknesses inherent within these applications.

A. Google News

Google News serves as a sophisticated news aggregation platform accessible through Google. It employs a multi-step process to collect, organize, and present news content from a wide array of sources. Initially, Google News meticulously scans numerous news websites, blogs, and other online platforms to gather and categorize articles. However, as Engelmann et al. (2021) noted, the nature of this information is shaped not only by journalists but also by the algorithms employed by the news aggregators.

Once the news articles have been collected, they are meticulously categorized into various topics using advanced algorithms. This system enables users to effortlessly explore different areas of interest, such as world news, technology, business, and more. Additionally, personalization is crucial

in enhancing the user experience. Le et al. (2019) support this, noting that Google News can offer tailored news recommendations by analyzing individual preferences, search history, and interactions. However, Le et al. (2019) also emphasize that further empirical evidence is needed to fully understand the causes of filter bubbles and echo chambers.

Conversely, research by Basch et al. (2022) suggests that Google News should enhance its communication efforts to boost screening uptake and highlights the need for better information dissemination. This recommendation stems from the observation that many articles recommended by Google News did not adequately address crucial aspects of Colorectal Cancer (CRC), particularly during the National Colorectal Cancer Awareness Month in March 2022.

B. Apple News+

Apple News+ is a subscription service offered by Apple that provides access to a wide range of global publications, including magazines, newspapers, and digital publishers (Apple News+, 2024). It combines technological algorithms with human editorial input to curate news content for users, reflecting a hybrid approach to news aggregation.

Research conducted by Bandy and Diakopoulos (2019) delved into the mechanisms behind Apple News and the content it presents, emphasizing the social, political, and economic consequences of its curation methods. The researchers employed multiple audit methods to analyze the update frequency of the application, content adaptation, source concentration, and more. The results demonstrated that the "human-curated Top Stories" section offers fewer news items each day but has higher source diversity and evenness compared to the "algorithmically-curated Trending Stories" section.

Conversely, Bandy and Diakopoulos (2019) argue that Apple News tends to show the same "Top Stories" and "Trending Stories" to all users to minimize individual filter bubble effects. However, this approach has a major drawback: minimal content personalization in the Trending Stories section. While personalization could enhance the user experience, the lack of extensive personalization ensures broader exposure to a variety of news topics and sources.

TABLE III. COMPARISION TABLE ON GOOGLE NEWS & APPLE NEWS+

Aspect	Comparison Table	
	Google News	***Apple News+***
Collection Process	Systematically scans news websites, blogs and online platforms.	Combines technological algorithms with human editorial input.
Categorisation	Utilises sophisticated algorithms to categorise news articles into various topics.	Features both algorithmically curated and huma-curated sections.
Personalisation	Tailored news recommendations based on user preferences, search history and interactions.	Minimises individual filter bubble effect by showing the same top and trending stories to every user.
Content Diversity	Offers a wide range of news content from diverse sources	Human-curated sections exhibit greater source diversity and evenness.
Communication Efforts	Needs to improve communication efforts to disseminate important information.	Content adaption and minimal personalisation ensure broader exposure to various news topics and sources.
Weakness	Relies heavily on algorithms for content curation, potentially leading to filter bubbles or echo chambers.	Limited content personalisation may result in less tailored user experience.
Strength	Provides comprehensive coverage of news topics with personalised recommendations.	Offers a hybrid approach that combines algorithmic curation with human editorial input for diverse and balanced content.

IV. CHALLENGES & LIMITATION

In this section, the researcher aims to identify the challenges and limitations that require attention, drawing from previous research efforts. Various prospective methodologies proposed by earlier researchers have also been highlighted and discussed.

A. Media Bias

Media bias in news aggregation and summarization can significantly alter an individual's awareness and perception of a particular topic. However, Hamborg et al. (2020) stated that most current news aggregators focus solely on gathering and delivering relevant information in news articles without presenting opposing viewpoints. This lack of balanced perspectives can become a critical issue, particularly affecting individuals' decisions on high-impact social events, such as elections or attitudes towards war.

Hamborg et al. (2020) introduced an innovative method for news exploration known as matrix-based news aggregation (MNA). The researchers claim that this approach can present users with different opinions on the same news topic, providing a comprehensive and diversified understanding of the news. Meanwhile, Ziashahabi et al. (2020) introduced an incentive layer using blockchain to address this issue. This incentive layer considers that users may have differing perspectives on news, enabling them to contribute their wide range of perspectives to the network. This technique could be beneficial for news aggregators in eliminating media bias.

Zhou and Tan (2023) reveal that political bias in this field is influenced by entity prominence. To address this, they developed an entity-based computational framework to evaluate the sensitivity of various extractive and abstractive summaries, focusing particularly on politicians Donald Trump and Joe Biden. The results demonstrate significant potential in identifying biases by de-emphasizing the presence of Trump and Biden compared to other presidents within the same article context, which may skew reader perception of events. Although this technique shows promise, it may not capture all nuances of political bias. Therefore,

future studies could utilize the base model of the entity replacement method to address this limitation.

B. Content Quality & Credibility

The challenges of content quality and credibility in news aggregation and summarization stem from "extrinsic hallucination," where summaries include facts do not present in the source document due to reliance on world knowledge, leading to erroneous information (Mishra et al., 2023; Xu et al., 2021). This issue arises because human-written summaries contain approximately 36% of external facts, which may not reflect the truth behind certain news stories (Maynez et al., 2020). Furthermore, the high volume of data generated globally each day exacerbates the problem, making it difficult for analysts to discern essential information from the noise. However, not much recent research has addressed this issue.

Research conducted by Diel and Roberts (2021) found that news organizations could enhance the perceived credibility of aggregated news stories by providing clear attributions to the original sources. Participants rated stories with clear attributions as more credible compared to those without attributions. Additionally, the results indicated that the perceived credibility of news stories was influenced by the reputation of the original source. Stories attributed to a reputable news organization were perceived as more credible than those attributed to lesser-known sources.

Xu et al. (2021) introduced the MiraNews dataset, which includes multiple supplementary resource documents to assist in generating summaries, addressing the issue of “extrinsic hallucination.” Extrinsic hallucination refers to inaccuracies derived from world knowledge rather than the source document. The study found that assisted summarization significantly reduces hallucinations by 55% compared to models trained solely on the main article, highlighting the effectiveness of leveraging multiple resources to enhance data quality and summary accuracy. However, the generalization of these models to diverse news domains or languages beyond the scope of the dataset is not extensively discussed, which could be a limitation in real-world applications where models need to adapt to various contexts.

V. TRENDS & FUTURE DIRECTIONS

In this section, the researchers explored an analysis of trends and future directions that have the potential to enhance the effectiveness of news aggregation and summarization.

A. Multimodal Summarisation

Multimodal summarization offers promising advancements in the performance of news aggregators by integrating essential information from diverse sources, thus improving the quality and reliability of summaries. Kumar et al. (2020) and Qiu et al. (2022) support this view, highlighting that multimodal summarization can produce more concise and informative summaries through cross-modal information integration. However, He et al. (2023) points out a limitation in current methods: they often neglect the temporal connections between modalities and the inherent correlations between samples. This oversight can diminish the effectiveness of unimodal summarization. To address these challenges, He et al. (2023) introduced the Align and Attend Multimodal Summarization (A2Summ) model. This innovation underscores the importance of advanced multimodal summarization techniques in effectively summarizing both news articles and videos, thus helping to tackle the information overload prevalent in today’s media landscape.

B. Ethical Consideration & Transparency

Ethical considerations and transparency play an essential role in the future development of news aggregators (Bandy & Diakopoulos, 2019). The proliferation of online news sources has resulted in a flood of information, making it challenging to distinguish between authentic and false reports (Dwivedi & K, 2022). Therefore, transparency in collecting user data is essential in addressing this ethical issue. However, achieving genuine transparency is complex and may not always benefit users as intended. Andrunyk et al. (2020) emphasize that understanding user needs, implementing transparent processes, and ensuring ethical data handling are key attributes to consider when developing intelligent information systems for news aggregation. Although news aggregators can aid in summarizing news effectively, ethical considerations are crucial to maintaining user trust and ensuring the accuracy of the information presented.

C. Cross-Platform Integration & Collaboration

Cross-platform integration and collaboration could be crucial in the further development of news aggregators. This approach enhances the efficiency and comprehensiveness of the news summarization process. Mohamed et al. (2020) support this, stating that integrating content from various sources, including news media streams and social media platforms, provides a more holistic view. Coddington (2020) also asserts that collaboration among different news agencies enables the sharing of resources and expertise. Additionally, Xu et al. (2021) emphasize that cross-platform integration facilitates the gathering of large-scale multilingual datasets for an inclusive and effective summarization system. This collaborative approach not only saves time for end-users by consolidating information in one place but also expands the news ecosystem, benefiting both news sites and end-users.

VI. CONCLUSION

In conclusion, this literature review offers insights into the evolving landscape of news aggregation and summarization by examining the background and related work of previous researchers. Additionally, it explores various existing news aggregation and summarization applications, such as Google News and Apple News+, to identify their inherent strengths and weaknesses.

Moreover, this literature review examines the challenges and limitations in this domain. Although most of the existing techniques, such as graph-based algorithms and fuzzy logic, demonstrate high potential in this field, there are persistent challenges and limitations, such as media bias and content

quality. Hence, it is necessary to conduct further research and efforts to address these limitations.

Finally, the researchers identified trends that could be implemented in future studies. These trends include multimodal summarization, ethical considerations, and cross-platform integration. Previous researchers have proven that these trends hold significant potential in enhancing the overall performance of news aggregators. Stakeholders can leverage these innovative approaches to address the mentioned limitations, thereby developing a robust system that provides timely, accurate, and diverse news content to users worldwide.

REFERENCES

Ali, S. M., & Dahake, V. G. (2023). Unsupervised Learning-based News Aggregation: A Comparative Study of Different Embedding and Clustering Techniques. *Indian Society for Technical Education*, *46*(3), 18–27.

Andrunyk, V., Vasevych, A., Chyrun, L., Chernovol, N., Antonyuk, N., & Gozhyj, A. (2020). *Development of Information System for Aggregation and Ranking of News Taking into Account the User Needs.* 1127–1171. https://typeset.io/papers/development-of-information-system-for-aggregation-and-39drigx8gv

Apple News+. (2024). Apple. https://www.apple.com/apple-news/

Bandy, J., & Diakopoulos, N. (2019). Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*, 36–47. https://doi.org/10.1609/icwsm.v14i1.7277

Basch, C. H., Hillyer, G. C., & Jacques, E. T. (2022). *News Coverage of Colorectal Cancer on Google News: Descriptive Study*. https://doi.org/10.2196/39180

Bhujbal, Mr. M., Bibawanekar, Ms. B., & Deshmukh, Dr. P. (2023). News Aggregation using Web Scraping News Portals. *International Journal of Advanced Research in Science, Communication and Technology*, *3*(2), 275–284. https://doi.org/10.48175/IJARSCT-12138

Carlebach, M., Cheruvu, R., Walker, B., Ilharco, C., & Jaume, S. (2020). News Aggregation with Diverse Viewpoint Identification Using Neural Embeddings and Semantic Understanding Models. *Harvard University*, 59–66.

Chen, S., Khashabi, D., Callison-Burch, C., & Roth, D. (2019). *PerspectroScope: A Window to the World of Diverse Perspectives* (arXiv:1906.04761). arXiv. https://doi.org/10.48550/arXiv.1906.04761

Chen, S., Khashabi, D., Yin, W., Callison-Burch, C., & Roth, D. (2019). *Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims* (arXiv:1906.03538). arXiv. https://doi.org/10.48550/arXiv.1906.03538

Chitra, N. T., Sreenivasulu, G., Subhashini, P., Sharma, G. K., & Sandhya, N. (2023). An Effective approach for preparation of News Summary using web scraping techniques. *2023 International Conference on Computer Communication and Informatics (ICCCI)*, 1–4. https://doi.org/10.1109/ICCCI56745.2023.10128366

Coddington, M. (2020a). Gathering evidence of evidence: News aggregation as an epistemological practice. *Journalism*, *21*(3), 365–380. https://doi.org/10.1177/1464884918817608

Coddington, M. (2020b). Gathering evidence of evidence: News aggregation as an epistemological practice. *Journalism*, *21*(3), 365–380. https://doi.org/10.1177/1464884918817608

Dahake, G. V., & Ali, M. Sadiq. (2024). Developing Unsupervised Learning Techniques for Business News Articles. *2023 6th International Conference on Advances in Science and Technology (ICAST)*, 547–551. https://doi.org/10.1109/ICAST59062.2023.10454905

Diel, S., & Roberts, C. (2021). News story aggregation and perceived credibility. *SAGE Publications Inc*, *42*(2), 162–181. https://doi.org/10.1177/07395329211013488

Dwivedi, R. K., & K, V. (2022). A Deep Investigation on News Aggregation and Recommendation System: NARS. *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 1–5. https://doi.org/10.1109/ICERECT56837.2022.10060379

Engelmann, I., Luebke, S. M., & Kessler, S. H. (2021). Effects of News Factors on Users' News Attention and Selective Exposure on a News Aggregator Website. *Journalism Studies*, *22*(6), 780–798. https://doi.org/10.1080/1461670X.2021.1889395

Gaskins, B., & Jerit, J. (2012). Internet News: Is It a Replacement for Traditional Media Outlets? *The International Journal of Press/Politics*, *17*(2), 190–213. https://doi.org/10.1177/1940161211434640

Gretz, S., Friedman, R., Cohen-Karlik, E., Toledo, A., Lahav, D., Aharonov, R., & Slonim, N. (2020). A Large-Scale Dataset for Argument Quality Ranking: Construction and Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(05),

Article 05. https://doi.org/10.1609/aaai.v34i05.6285

Gulati, V., Kumar, D., Popescu, D. E., & Hemanth, J. D. (2023a). Extractive Article Summarization Using Integrated TextRank and BM25+ Algorithm. *MDPI*, *12*(2), Article 2. https://doi.org/10.3390/electronics12020372

Gulati, V., Kumar, D., Popescu, D. E., & Hemanth, J. D. (2023b). Extractive Article Summarization Using Integrated TextRank and BM25+ Algorithm. *Electronics*, *12*(2), Article 2. https://doi.org/10.3390/electronics12020372

Hamborg, F., Meuschke, N., & Gipp, B. (2020). Bias-aware news analysis using matrix-based news aggregation. *International Journal on Digital Libraries*, *21*(2), 129–147. https://doi.org/10.1007/s00799-018-0239-9

Hasan, M., Orgun, M. A., & Schwitter, R. (2019). Real-time event detection from the Twitter data stream using the TwitterNews+ Framework. *Information Processing & Management*, *56*(3), 1146–1165. https://doi.org/10.1016/j.ipm.2018.03.001

He, B., Wang, J., Qiu, J., Bui, T., Shrivastava, A., & Wang, Z. (2023). *Align and Attend: Multimodal Summarization with Dual Contrastive Losses* (arXiv:2303.07284). arXiv. https://doi.org/10.48550/arXiv.2303.07284

Huang, Y., Cui, B., Zhang, W., Jiang, J., & Xu, Y. (2015). TencentRec: Real-time Stream Recommendation in Practice. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 227–238. https://doi.org/10.1145/2723372.2742785

Kumar, A., Sharma, A., & Nayyar, A. (2020a). Fuzzy Logic based Hybrid Model for Automatic Extractive Text Summarization. *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, 7–15. https://doi.org/10.1145/3385209.3385235

Kumar, A., Sharma, A., & Nayyar, A. (2020b). Fuzzy Logic based Hybrid Model for Automatic Extractive Text Summarization. *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, 7–15. https://doi.org/10.1145/3385209.3385235

Langote, M. M., & Gawande, Dr. R. (2023). Effective News Text Summarization Techniques. *International Journal of Advanced Trends in Computer Science and Engineering*, *12*(3), 140–143. https://doi.org/10.30534/ijatcse/2023/071232023

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring Political Personalization of Google News Search. *The World Wide Web Conference*, 2957–2963. https://doi.org/10.1145/3308558.3313682

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

Mishra, A., Sahay, A., Pandey, M. anjusha, & Routaray, S. S. (2023). News text Analysis using Text Summarization and Sentiment Analysis based on NLP. *2023 3rd International Conference on Smart Data Intelligence (ICSMDI)*, 28–31. https://doi.org/10.1109/ICSMDI57622.2023.00014

Mohamed, A., Ibrahim, M., Yasser, M., Ayman, M., Gamil, M., & El-Ashmawi, W. (2020). News Aggregator and Efficient Summarization System. *International Journal of Advanced Computer Science and Applications*, *11*, 636–641. https://doi.org/10.14569/IJACSA.2020.0110677

Mohite, S., Shukla, A., Jadhav, S., Jain, D., Aggarwal, A., & Jaiswal, S. (2023). *Insight Now: A Cross-Platform News Application for Real-Time and Personalized News Aggregation*. 7. https://doi.org/10.1109/ICCST59048.2023.10474234

Qiu, J., Zhu, J., Xu, M., Dernoncourt, F., Bui, T., Wang, Z., Li, B., Zhao, D., & Jin, H. (2022). *Semantics-Consistent Cross-domain Summarization via Optimal Transport Alignment* (arXiv:2210.04722). arXiv. https://doi.org/10.48550/arXiv.2210.04722

Reddy, B. R., Reddy, D. S. T., Preetham, M. C. S., Rajasekhar, A. H. N., & Subramani, R. (2022). Comparative Study Analysis on News Articles Categorization using LSA and NMF Approaches. *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT54827.2022.9984466

Sanderson, M., & Croft, W. B. (2012). The History of Information Retrieval Research. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1444–1451. https://doi.org/10.1109/JPROC.2012.2189916

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). *FEVER: A large-scale dataset for Fact Extraction and VERification* (arXiv:1803.05355). arXiv. https://doi.org/10.48550/arXiv.1803.05355

Xu, X., Dušek, O., Narayan, S., Rieser, V., & Konstas, I. (2021). *MiRANews: Dataset and Benchmarks for Multi-Resource-Assisted News Summarization*

(arXiv:2109.10650). arXiv. https://doi.org/10.48550/arXiv.2109.10650

Zhou, K., & Tan, C. (2023). Entity-Based Evaluation of Political Bias in Automatic Summarization. *arXiv: Distributed, Parallel, and Cluster Computing*. https://doi.org/10.48550/arXiv.2305.02321 Focus to learn more

Ziashahabi, A., Maddah-Ali, M. A., & Heydarnoori, A. (2020). *Bias-Resistant Social News Aggregator Based on Blockchain* (arXiv:2010.10083). arXiv. https://doi.org/10.48550/arXiv.2010.10083