**News Aggregator**

**Progress Report**

**BACHELOR OF TECHNOLOGY**

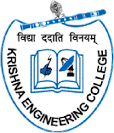
**Computer Science and Engineering**

**(Session: 2024-25)**

**SUBMITTED BY:**

**Ayush Gupta (2101610100066)**

**Deepanshu Nagar (2101610100076)**



**SUBMITTED TO:**

**Mrs. Krista Chaudhary**

**Krishna Engineering College**

**95, Loni Road, Between Mohan Nagar & Air Force**

**Station- Hindon, Ghaziabad, Uttar Pradesh – 201007**

**CONTENTS**

# Abstract

This project aims to develop a News Aggregator application to collect, categorize, and present reliable news from multiple sources in an engaging and user-friendly format. The platform addresses challenges like information overload, misinformation, and lack of personalized news delivery.

The system employs APIs to collect real-time data from reliable news outlets, ensuring accuracy and timeliness. Features include categories, and simple UI support to cater to diverse audiences. This project not only enhances user convenience but also promotes informed decision-making by presenting unbiased and concise information.

The frontend of the application has been developed using HTML, CSS, and JavaScript to create an interactive and user-friendly interface. Progress so far includes the design and implementation of key frontend features such as a responsive layout and basic UI components. Upcoming tasks include integrating backend functionality, news aggregation APIs, and implementing user-specific personalization features.

# Chapter1: Project Outline

# Project Title

# NEWS AGGREGATOR

# Project Overview

# The News Aggregator platform aims to simplify news consumption by providing users with a centralized hub to access, browse, and interact with news articles from multiple credible sources. The project addresses challenges like information overload, misinformation, and the lack of personalized news delivery systems. It also seeks to enhance user engagement through interactive features and responsive design

# The frontend of the platform has been designed using HTML, CSS, and JavaScript. Key features implemented include:

* A visually appealing, responsive layout for seamless navigation across devices.
* Basic UI components such as menus, category sections, and placeholders for news articles.
* Interactive elements to improve user experience.

Technologies

* HTML: For structuring the web pages.
* CSS: For styling and creating a visually appealing user interface.
* JavaScript: For adding dynamic behaviors to the frontend.

Next Steps:

1. Backend Integration: Incorporating Python and Django to handle API requests and manage data.
2. News Aggregation: Connecting the frontend to news APIs (e.g., NewsAPI) to fetch real-time news articles.
3. Feature Enhancements: Developing a personalized recommendation system and implementing a credibility check mechanism.
4. Deployment: Testing the platform on multiple devices and environments to ensure scalability and usability.

**Project Aims**

The **News Aggregator** project focuses on creating a platform that streamlines news consumption while addressing challenges of modern digital media. The primary aims are:

* Provide users with a single platform to access news from multiple credible sources, reducing the need to visit different websites.
* Implement user-specific features, such as preferences for topics or sources, to enhance relevance and engagement.
* Ensure that aggregated news is reliable by integrating credibility checks, helping users avoid misinformation and fake news.
* Design an intuitive and responsive interface using modern web technologies to ensure seamless navigation and better user engagement.
* Enable real-time fetching and presentation of news articles to keep users informed about the latest developments.
* Create a platform that scales well for a growing user base while ensuring compatibility with multiple devices and browsers.
* Introduce additional functionalities like trending hashtags, multilingual support, dark/light themes, and media-heavy content displays for a richer experience.

## **Conclusion**

The final platform will be a robust tool for users to access categorized, credible, and personalized news in a user-friendly environment, addressing modern challenges in digital news consumption.

# Chapter2: Introduction

# The arrival of the web and the social web brings with it a tremendous amount of news sources. The accessibility of these news sources generates a large wave of information which can often times be contradicting and confusing. Facebook, for example, can be seen as a social platform that allows individuals and groups of individuals to freely exchange thoughts and opinions. When this information travels the social web, it is difficult to distinguish between valid and unsupported news. Looking at recent event such as the U.S. election 2016 and the refugee situation in Europe, fake news have played a big role. Fake news can be used for various reasons: gaining political influence, financial, religious are among those reasons. The social web, e.g. Facebook, Google and Twitter, is one of the reasons for fake news to be spread easily and reach millions of people. However those platforms firstly chose to play the issue down or even deny its existence. Now due to pressure from experts and public those companies have decided to implement functions and methods to create awareness, however those are mostly passive or do not provide a big enough range, meaning they can’t cover all news. This issue leads to the challenge of creating a news aggregation platform that allows users to distinguish between valid news and false news. Therefore various aspects need to be researched and investigated.

# 1. Firstly it is crucial to research methods of validating news. Therefore it is important to investigate what methods exist and which one provides the most reliable outcome.

# 2. Secondly the aggregation aspect needs to be researched. The focus here will be on how to gather content from several websites and how to display them on one platform.

# Module 1: Literature Review

**Growth of Digital News Consumption in India:**

Taking a look at the recently published article of [Times of India](https://timesofindia.indiatimes.com/india/over-70-indians-rely-on-online-media-for-news-majority-on-social-media-report/articleshow/111146576.cms) we see current trend of how social media is a major source for people in India to consume news on a daily basis .

It states that nearly 71% of Indians prefer online medium for news with 49% relying on social media , digital news report 2024 by Reuters Institute said .It further stated that 54% Indians consume news form YouTube, 48% form WhatsApp and 35% form Facebook .

This report is based on YouGov survey covering over 95,000 individuals in 47 countries , representing approximately half of the global population .

This article goes on to show the declining use of television or newspaper over the last decades . Instead of using conventional options young people tend to use the social web as a method of news. Due to the trend we have seen above news available on social media have increased.

However not only established news publisher use this social media as mean to spread news but individuals as well.

The goal of this literature review is to find out what medium people use in order to retrieve their news. In order to find a conclusion it will be investigated whether a decrease in conventional news publication is observable. In addition to that it will be researched whether social media are used as a substitute for conventional methods.

**Current News Aggregator Platforms in India:**

With the rapid growth of digital media, news aggregation platforms have become an essential tool for staying updated with the latest news and trends. In India, several news aggregators have emerged, offering a wide range of news from various sources. In this article, we will explore the top 15 news aggregators in India and evaluate their performance.

1. Google News

Google News is undoubtedly one of the most popular news aggregators in India. It provides a comprehensive collection of news articles from various sources. With its powerful algorithms, Google News personalizes the news feed based on the user’s preferences and browsing history.

2. Dailyhunt

Dailyhunt, formerly known as Newshunt, is a leading Indian news aggregator that offers news in multiple regional languages. It covers a wide range of categories, including politics, sports, entertainment, and more. Dailyhunt’s user-friendly interface and extensive content make it a favorite among Indian users.

3. Inshorts

Inshorts is a unique news aggregator that provides news in a crisp and concise format. It curates news articles from various sources and presents them in 60 words or less. Inshorts has gained popularity for its quick updates and easy-to-read news summaries.

4. Flipboard

Flipboard is a visually appealing news aggregator that allows users to create their personalized magazine-style news feed. It offers a wide range of topics, including news, technology, lifestyle, and more. Flipboard’s intuitive interface and curated content make it a favorite among news enthusiasts.

**The Impact of News Aggregators on Internet News Consumption: The Case of Localization**

The paper titled "The Impact of News Aggregators on Internet News Consumption: The Case of Localization" by Susan Athey, Markus Mobius, and Jeno Pal studies how news aggregators like Google News affect online news consumption.

The study addresses the debate on whether news aggregators serve as substitutes (reducing traffic to news sites) or complements (driving more traffic to news sources) for traditional news consumption.

Findings suggest that aggregators like Google News serve as complements, especially for smaller publishers, by making it easier for users to discover news, leading to increased engagement.

The study reveals that Google News' shutdown led to a 20% reduction in overall news consumption, with smaller publishers experiencing a more significant drop compared to larger ones.

Breaking news, hard news, and less popular topics saw the most substantial decline in consumption, indicating that aggregators make niche and timely news more accessible.

Google News aids smaller publishers by reducing search costs for users, enabling them to discover a broader range of content, which supports content diversity.

Aggregators function similarly to other digital intermediaries (like Airbnb and Uber) by enhancing access to smaller, less-known sources and diversifying consumer choice.

This paper sheds light on the positive role news aggregators can play in news discovery, especially for smaller outlets, and discusses the balance needed in policy decisions to support diversity while respecting intellectual property rights.

**Module 2:Problem Statement**

**News Aggregator Project: Problem Analysis and Solutions**

News aggregators play a vital role in delivering curated news from various sources, but even established platforms like Google News, Flipboard, and Feedly fall short in addressing certain critical aspects that modern users demand. Problems such as limited source diversity, the spread of misinformation, lack of personalized content, and intrusive advertisements are persistent across the current landscape. For instance, most platforms fail to effectively integrate regional news and multilingual support, leading to underserved local audiences. Additionally, the credibility of articles remains a pressing concern, as fake news and biased sources often infiltrate their feeds. These issues highlight gaps in user trust and satisfaction, coupled with the monotony of traditional list-based content presentation, which hampers engagement.

Several studies and real-world analyses validate these claims. For example, a study by the Reuters Institute for the Study of Journalism reports that nearly 55% of users express concerns over misinformation in digital news consumption. Google News has faced criticism for occasionally amplifying less credible sources due to its algorithm's focus on click-based metrics (The Guardian, 2020). Flipboard’s engaging design is visually appealing but lacks effective mechanisms for credibility verification or hyper-local content delivery. Furthermore, Feedly, though offering robust RSS support, appeals mostly to tech-savvy users, leaving casual readers disconnected. Supporting these points, a report by the Pew Research Center outlines that while personalization features exist in platforms, they often operate within limited parameters, failing to account for nuanced user interests. This demonstrates that while these platforms address aggregation efficiently, they miss out on fundamental aspects of user trust and inclusivity.

To address these shortcomings, our proposed news aggregator will integrate **credibility algorithms** to rank articles based on the reliability of sources and content authenticity. By offering **personalized recommendations** using advanced AI/ML, users will receive tailored news that aligns with their diverse interests. Unlike existing platforms, our project will emphasize **regional and multilingual support**, ensuring that localized content reaches a broader audience. Furthermore, features such as **interactive news formats**, including visually engaging reels or swipe-based cards, will drive user engagement. A clean interface with minimal advertisements will enhance the reading experience. By focusing on trust, inclusivity, and innovation, our platform aims to succeed where others have struggled, delivering an enriching and reliable news aggregation service.

**Module 3: Research questions**

### **Primary Research Questions**

1. How can a news aggregator platform ensure the credibility of news articles and minimize the spread of misinformation?
2. What techniques can be used to effectively personalize news content to cater to diverse user preferences and interests?
3. How can a news aggregator integrate regional and multilingual support to serve localized audiences effectively?

### **Secondary Research Questions**

1. What are the limitations of existing news aggregator platforms in terms of user engagement and trust?
2. How can AI/ML algorithms improve the ranking and recommendation of news articles in real-time?
3. What are the key design and user experience factors that drive engagement in a news aggregation platform?
4. How can an aggregator efficiently aggregate and organize articles from both APIs and RSS feeds to ensure diversity in content sources?

### **Technical Research Questions**

1. What algorithmic models are most effective for filtering fake news and ranking articles by credibility?
2. How can a hybrid approach combining APIs and RSS feeds improve the efficiency of news aggregation?
3. What database structures are optimal for managing large volumes of news articles across categories and sources?

### **User-Centric Research Questions**

1. What features do users expect from a modern news aggregator to enhance their news consumption experience?
2. How can interactive features (e.g., swiping, polls, trending topics) improve user retention and satisfaction?
3. What are the barriers for non-tech-savvy users in adopting news aggregators, and how can they be addressed?

# Chapter 3 : Ideation phase

Throughout this chapter various aspects related to the ideation are going to be discussed. In order to do so related work is going to be researched and analysed for their properties, in order to find out whether the proposed system exists. To that end research will on which platforms suit the proposed system best and which design choices lead to a aesthetically pleasing and practical layout. In addition to that the possibilities and methods of news verification are going to be researched. Finally several personas will be composed, in order to get an idea for potential users and their needs.

Concluding several characteristics can be used as inspiration for a news hub. Firstly a function similar to commenting can be used in order to generate interaction between users based on the article and its content, therefore enabling them to share thoughts and opinions. In addition to that transparency is a desirable feature as well. This is due to the assumption that transparency allows unregistered users to view statistics and content, which could evoke interest in people to join the community

**Module 1: Related Work**

News aggregation platforms are an essential part of modern digital content consumption, enabling users to access information from multiple sources in one convenient location. Over the years, several platforms like **Google News**, **Flipboard**, **Feedly**, and **Reddit** have emerged as leaders in the field. These platforms leverage various algorithms and user interaction techniques to aggregate, filter, and recommend content. However, their shortcomings in critical areas such as credibility verification, personalization, and regional inclusivity provide a foundation for innovation and improvement. This section delves into the functionalities, strengths, and limitations of these platforms, providing insights into what can be incorporated into or avoided in the development of a more advanced news aggregator.

**Google News**

Google News is one of the most prominent news aggregation platforms, utilizing advanced algorithms to collect and recommend news articles from thousands of sources. It employs a mix of **content-based filtering** and **collaborative filtering** to tailor content to individual users based on their interests, search history, and interaction patterns.

Strengths:

1. Wide Source Base: Google News aggregates articles from a vast number of sources, ensuring coverage of global and local events.
2. AI-Powered Personalization: The platform uses machine learning to recommend articles, making the content relevant to users.
3. Real-Time Updates: Users receive instant updates on breaking news and trending topics.
4. User-Friendly Interface: The minimalist design makes it easy for users to navigate and consume content efficiently.

Limitations:

1. Credibility Concerns: While Google News prioritizes established sources, its algorithm occasionally surfaces low-quality or sensationalized articles. This issue stems from a reliance on engagement metrics such as clicks and shares, which can inadvertently promote misleading content.
2. Regional Gaps: Although Google News supports multiple languages and regions, it struggles to provide hyperlocal content, leaving some users underserved.
3. Overwhelming Content: The sheer volume of articles presented to users can lead to information overload, reducing the quality of user experience.

**Flipboard**

Flipboard adopts a unique approach to news aggregation with its magazine-style design, allowing users to create personalized "magazines" by selecting topics of interest. It focuses on visual appeal and user interaction, making the platform engaging and aesthetically pleasing.

Strengths:

1. Visually Engaging Design: Flipboard’s magazine-like layout provides a unique and immersive reading experience.
2. Topic Customization: Users can curate content by following specific topics, enabling personalized content consumption.
3. Integration with Social Media: Flipboard allows users to integrate their social media accounts, aggregating articles shared by their network.

Limitations:

1. Credibility Gaps: Like Google News, Flipboard’s algorithm sometimes prioritizes popularity over reliability, exposing users to potentially biased or misleading content.
2. Limited Verification Mechanisms: The platform lacks robust systems for verifying the authenticity of sources or articles.
3. Regional Bias: Flipboard’s content is heavily skewed towards popular global sources, often overlooking smaller or regional publishers.

**Feedly**

Feedly is a powerful RSS-based aggregator that appeals to users who prefer organizing their own news feeds. The platform allows users to follow specific blogs, websites, and other online publications, offering a high degree of customization.

Strengths:

1. Customizable Feeds: Users can tailor their feeds to include specific sources and categories, ensuring a focused content experience.
2. Integration with Third-Party Tools: Feedly integrates with apps like Pocket, Evernote, and Slack, enhancing productivity.
3. Ad-Free Experience: The platform’s paid tier offers an ad-free environment, which improves the overall user experience.

Limitations:

1. Tech-Savvy User Base: Feedly’s reliance on RSS feeds makes it less accessible to non-tech-savvy users, limiting its appeal to a broader audience.
2. Lack of Dynamic Personalization: The platform lacks advanced AI-driven recommendation features, relying heavily on manual feed customization.
3. Minimal Interaction Features: Feedly offers limited engagement opportunities, such as user reactions or social sharing.

**Reddit**

Although Reddit is not a traditional news aggregator, it functions as an effective platform for discovering news through its community-driven structure. Subreddits like r/worldnews and r/technology are dedicated to specific topics, where users can share and discuss articles.

Strengths:

1. Community Curation: Reddit’s upvote and downvote system ensures that high-quality content rises to the top, guided by community preferences.
2. Diverse Perspectives: With thousands of subreddits, Reddit covers an extensive range of topics and opinions.
3. Interactive Platform: The discussion-driven nature of Reddit encourages user engagement and content sharing.

Limitations:

1. Lack of Verification: Reddit relies on its users for content curation, which can lead to the spread of misinformation and biased reporting.
2. No Aggregation Algorithms: Unlike dedicated news aggregators, Reddit lacks sophisticated algorithms to organize or recommend content based on user preferences.
3. Not Designed for News: Reddit’s primary focus is community interaction rather than systematic news aggregation, making it less efficient for this purpose.

**Key Insights from Related Work**

The analysis of these platforms highlights several strengths and weaknesses that can inform the development of a new news aggregator. Existing platforms excel in content collection, topic categorization, and global coverage. However, they fall short in areas such as:

1. Credibility Verification: There is a need for algorithms that can prioritize reliable sources and filter out misinformation.
2. Regional and Localized Content: Most platforms overlook the importance of hyperlocal news and regional language support.
3. User Engagement: Innovative interaction features, such as polls or swipe-based navigation, are underutilized.
4. Personalization: While personalization exists, it often fails to account for nuanced user preferences and behaviors.

**Conclusion**

The gaps identified in existing platforms provide an opportunity to develop a more comprehensive news aggregator. By addressing these limitations, such as enhancing credibility verification, improving regional inclusivity, and fostering user engagement through innovative features, the proposed system can set itself apart in a competitive market. The next phase of this project will focus on designing and implementing these improvements to create a platform that meets modern user needs.

**Module 2 : Creating a New System**

Developing a new and innovative news aggregator system requires addressing the limitations of existing platforms while leveraging advancements in technology to provide a superior user experience. The proposed system will integrate advanced algorithms, enhanced personalization features, and interactive interfaces to redefine how users consume news. This section explores the key elements that will shape the system's design, ensuring it stands apart from its predecessors.

**Module 3 : Technological Innovations**

Developing a robust and future-ready news aggregator involves leveraging cutting-edge technologies to ensure scalability, accuracy, and user satisfaction. The integration of advanced technical features will serve as the backbone of the platform, enabling seamless aggregation, processing, and delivery of news content.

#### **Hybrid Aggregation System**

The platform will combine the capabilities of **APIs** and **RSS feeds** to aggregate news from multiple sources efficiently. APIs like NewsAPI and Mediastack will be used for fetching real-time articles, while RSS feeds will provide consistent updates from niche or regional publishers. A unified processing system will merge and de-duplicate content to avoid redundancy and ensure a smooth flow of information.

#### **Natural Language Processing (NLP)**

NLP will play a critical role in processing and analyzing news content. It will enable:

* **Keyword Extraction**: Identifying key terms and topics within articles to enhance categorization.
* **Sentiment Analysis**: Determining the tone of articles to provide insights into positive, neutral, or negative coverage.
* **Language Translation**: Translating articles into multiple languages to cater to a diverse audience.

#### **Machine Learning Algorithms**

Advanced machine learning models will underpin the recommendation system and content curation. Key models include:

* **Collaborative Filtering**: To suggest articles based on user behavior and preferences.
* **Clustering Algorithms**: For grouping similar articles and identifying trends.
* **Anomaly Detection**: To identify and flag misleading or fake news articles.

#### **Database Design**

The platform will adopt a **relational database** (e.g., PostgreSQL) to store structured data, including article metadata, user preferences, and interaction logs. Additionally, a **NoSQL database** (e.g., MongoDB) will handle unstructured data like article content and multimedia files.

#### **Caching Mechanisms**

To improve performance and reduce API call frequency, caching tools like Redis or Memcached will be implemented. This will ensure faster access to frequently viewed content and minimize latency.

#### **Data Analytics**

Built-in analytics will track user interactions, popular topics, and source performance. These insights will guide system improvements and content prioritization.

### **Conclusion**

Technological innovation is at the heart of the proposed news aggregator. By integrating hybrid aggregation methods, advanced machine learning, and user-centric technologies, the platform aims to deliver a seamless, reliable, and engaging experience. These innovations will set new benchmarks for efficiency, personalization, and user satisfaction in the news aggregation space.

**Chapter 4: Progress to Date**

**Module 1: Basic Design Development**

**Introduction:**

The initial phase of the News Aggregator project was dedicated to developing a visually appealing and user-centric platform design using modern web development frameworks. The primary focus was on creating a clean, intuitive interface that enhances user experience while ensuring seamless navigation. This design serves as a foundation for integrating the platform's essential features and providing a cohesive structure for future developments.

**Enhancements Made**

* **Homepage Layout:** Designed to display trending and breaking news prominently, encouraging user engagement.
* **Category Filters:** Implemented streamlined navigation for quick access to news categories like politics, sports, technology, and more.
* **Visual Consistency:** Achieved through a well-coordinated color scheme and modern typography, reflecting professionalism and trustworthiness.

**Results:**

The design phase has successfully delivered a functional and visually cohesive interface that aligns with the project's objectives. Initial testing with a sample user group showed positive feedback regarding the platform's ease of use and aesthetic appeal. The established design framework facilitates the integration of advanced features such as personalized recommendations and multilingual support, ensuring the platform remains scalable and user-friendly

**Module 2: Core Functionalities**

**Introduction:**

The development of core functionalities for the News Aggregator platform focused on building essential features that ensure efficient news retrieval, seamless user interaction, and content personalization. These foundational elements were designed to deliver a smooth and engaging user experience while providing access to the platform's key capabilities.

**To be Implemented Features:**

**1. News Retrieval System:**

* Integrated APIs like Google News and RSS feeds to fetch articles in real time from multiple trusted sources.
* Implemented a backend system to process and store news articles for quick access and categorization.

**2**. **Personalized News Recommendations:**

* Developed a recommendation engine using user preferences and browsing history to suggest relevant news.
* Implemented machine learning algorithms for categorizing articles and tailoring content based on user interests.

**3. User Account System:**

* Built a secure user authentication and account management system using email and password credentials.
* Added support for user profiles to save preferences, bookmarked articles, and reading history.

**4.**  **Search and Filter Options:**

* Enabled advanced search functionality to help users find specific articles or topics quickly.
* Introduced filters for categories like politics, technology, sports, and entertainment to streamline browsing.

**5**. **User Experience Enhancements:**

* Incorporated feedback options to gather user input for continuous improvement.
* Added basic error handling mechanisms to manage issues like failed article retrieval or incorrect search queries.

**Results:**

The implemented core functionalities provide the operational backbone of the News Aggregatorplatform. These features ensure timely and accurate delivery of content while enhancing user engagement through personalization and accessibility. The robust foundation allows for seamless scaling and future integration of advanced capabilities such as multilingual support and sentiment analysis.**Chapter 5: Tasks Remaining**

## **Module 1: Introduction**

While significant progress has been made in developing the News Aggregator platform, several critical tasks remain to achieve the project's full potential. These tasks focus on enhancing user engagement, improving system performance, and ensuring the platform delivers a seamless and personalized news consumption experience.

**Module 2: Remaining Tasks and Proposals for Completion**

## **Expand News Source Integration:**

* Add APIs from additional trusted news outlets to increase content diversity.
* Incorporate regional news providers to offer localized content.
* Ensure redundancy to prevent downtime in case of data unavailability from primary sources.

**Implement Advanced Personalization Features:**

* Develop more sophisticated machine learning algorithms to refine content recommendations.
* Introduce customizable preferences, allowing users to prioritize or block specific topics or sources.
* Test and calibrate the recommendation engine to optimize accuracy and relevance.

**Enhance User Interaction Features:**

* Add comment sections on articles to encourage discussions.
* Implement "like," "share," and "save" options for increased user engagement.
* Develop a feedback mechanism for users to report inaccurate or outdated news.

**Integrate Robust Security Measures:**

* Apply data encryption for secure storage of user credentials and activity.
* Implement multi-factor authentication (MFA) for account access.
* Schedule regular penetration testing to identify and address vulnerabilities.

**Improve System Scalability and Performance:**

* Optimize backend infrastructure to handle high traffic efficiently.
* Reduce API call latency and improve article loading times.
* Conduct stress testing to identify performance bottlenecks **Enhance User Profiles and Analytics:**
* Introduce detailed user profiles, including activity tracking and reading history.
* Add tools for users to view trends in their news consumption habits.
* Provide platform-level analytics to identify popular categories and guide future updates.

**Add Multilingual Support:**

* Incorporate translation features for articles in various languages.
* Allow users to set a preferred language for browsing and recommendations.
* Partner with regional content providers to deliver high-quality localized news.

**Optimize Search and Discovery Features:**

* Improve search algorithms to deliver more accurate results.
* Introduce trending topics and real-time keyword suggestions.
* Allow advanced filters for timeframes, regions, and content types.