IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

Web Content Acquisition in Business News Aggregation

Prof. Dr. P. S. Gayke[1], Sneha Panpat[2], Namrata Pimparkar[3], Vishakha Shinde[4], Varsha Kardile[5]
Asst. Prof., Department of Information Technology[1]
Students, Department of Information Technology[2,3,4,5]
Dr. Vithalrao Vikhe Patil College of Engineering, Ahilyanagar, Maharashtra
Savitribai Phule Pune University, Pune

Abstract: *In today's fast-moving digital world, people rely heavily on online news platforms to stay updated about current business trends and market changes. However, with a large number of news websites, it becomes difficult for users to find relevant and reliable business information quickly. To address this challenge, the proposed project titled "Web Content Acquisition in Business News Aggregation" aims to design and develop a Java-based web application that automatically collects, filters, and organizes business news from multiple online sources. The system uses web content acquisition techniques to gather business-related news articles and information from trusted websites. Users can personalize their experience by selecting preferred news categories such as finance, stock markets, startups, or international trade when they first register on the platform. The system applies content-based filtering methods to recommend relevant articles based on user interests. A built-in chatbot powered by Natural Language Processing (NLP) allows users to interact with the system easily and retrieve specific news. Additionally, sentiment analysis is used to identify the tone of news articles — whether positive, negative, or neutral — helping users better understand market sentiment. Overall, this project combines advanced techniques from machine learning and natural language processing to create an intelligent, user-friendly, and efficient business news aggregation platform. It not only saves time but also helps users access the most relevant and up-to-date business information for informed decision-making..*

Keywords: Web Content Acquisition, Business News Aggregation, NLP Chatbot, Sentiment Analysis, Content-Based Filtering, Java Web Application, etc

I. INTRODUCTION

In today's digital world, people depend heavily on online news platforms to stay informed about the latest developments in the business world. With a vast amount of business-related information available across the internet, it becomes difficult for users to find relevant and reliable news quickly. Many news websites display general updates without considering user interests. To solve this problem, our project titled "Web Content Acquisition in Business News Aggregation" aims to design and develop a Java-based web application that automatically collects, filters, and presents business news from different online sources in one place. This system helps users save time and stay updated with the latest business trends that match their preferences.

The proposed system allows users to create an account and select their preferred categories or topics, such as finance, startups, marketing, or global business. Once the user selects their interests, the system begins to gather related articles from various multimedia sources using web content acquisition techniques. The main focus is on business news articles collected from reliable websites. To make the system more interactive, a chatbot powered by Natural Language Processing (NLP) is included, which allows users to search or ask for news updates in a conversational manner. The system also performs sentiment analysis on each news article to understand its emotional tone — whether it conveys a

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

positive, negative, or neutral business outlook. This helps users quickly grasp the overall sentiment of the market or a specific topic.

Additionally, the system uses content-based filtering techniques to recommend personalized news content to users based on their previous reading patterns and preferences. This ensures that the user receives news that is most relevant to their interests. The platform provides a simple, user-friendly interface where users can easily browse, read, and analyze business news. By combining technologies like web scraping, NLP, sentiment analysis, and intelligent filtering, this project provides an efficient and smart way to consume business news. Overall, the system aims to create a personalized, interactive, and intelligent news aggregation platform that helps user's access important business information quickly and efficiently.

II. PROBLEM STATEMENT

In today's digital world, business news is published across many websites, apps, and platforms. Users often find it difficult to access relevant and trustworthy news quickly because information is scattered and repetitive. There is a need for a smart web-based system that can automatically collect, analyze, and filter business-related content from multiple online sources and present it to the user based on their interests and preferences.

III. LITERATURE SURVEY

In recent years, many researchers have focused on improving web content acquisition, multimedia data aggregation, and intelligent filtering techniques to make information retrieval faster and more accurate. These studies form the foundation for our proposed business news aggregation system. The existing literature highlights the use of machine learning, natural language processing (NLP), and hybrid system designs to collect and organize data from multiple online sources efficiently. Below are some of the significant works reviewed during the study of this project.

• Cem Tekin and Mihaela van der Schaar, et. al. This research introduces a contextual online learning model for multimedia content aggregation. The authors propose a learning-based method that dynamically adapts to user preferences by analyzing user behavior and contextual information in real-time. Their approach helps in predicting which type of multimedia content (articles, videos, or blogs) will be most relevant to users. This concept is useful for the proposed system since business news aggregation also requires real-time learning and adaptation to user preferences based on browsing history and selected topics.

• Fengwei An and Hans J¨urgen Mattausch, et. al. This paper focuses on improving multimedia processing efficiency using a flexible hardware - software co-design approach for K-means clustering. The study demonstrates how clustering techniques can be optimized for real-time multimedia applications. The concept of clustering from this research is relevant to news aggregation, where clustering can be applied to group similar business articles or detect trending topics across multiple sources, improving the organization and accessibility of aggregated content.

• V. Pichiyan, et al. This paper presents a method that integrates Natural Language Processing (NLP) with web scraping techniques to improve the extraction of structured information from unstructured web data. By using NLP for entity recognition and sentence understanding, the system can identify meaningful content more accurately. This research directly relates to the proposed project since NLP-based web scraping is used to gather and interpret business news articles from various platforms efficiently, ensuring high data quality and contextual relevance.

• R. Kotsakis et al, In this study, the authors developed a web framework designed to collect and aggregate multilingual web content efficiently. Their approach supports content acquisition from different languages and formats, providing a unified data representation. This concept is valuable for expanding the scope of the proposed system, allowing future integration of global business news sources in multiple languages. The paper also emphasizes modularity and scalability, which align with the design goals of our proposed business news aggregation system.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

IV. SYSTEM OVERVIEW

The proposed system is a web-based Business News Aggregation application developed using Java technology. The main goal of the system is to collect, organize, and present business-related news content from multiple online sources in a single platform. These sources include business news websites, online articles, and multimedia content such as images and videos related to business topics. When a user signs up or accesses the system for the first time, they are asked to select their preferred business categories or topics, such as stock markets, startups, finance, economy, or technology. Based on these preferences, the system personalizes the news feed for each user.

The system uses content-based filtering techniques to analyze news articles and match them with user interests. Natural Language Processing (NLP) techniques are applied to process the textual content of news articles and to power an integrated chatbot. This chatbot allows users to interact with the system by asking questions, searching for specific business news, or getting summaries and recommendations. Sentiment analysis is also performed on news articles to determine whether the content has a positive, negative, or neutral impact, which helps users quickly understand market trends and public opinion.

Overall, the system provides a personalized, intelligent, and user-friendly news aggregation experience. By combining multimedia content aggregation, NLP-based chatbot interaction, sentiment analysis, and personalized content filtering, the application helps users stay informed about business news in a more efficient and meaningful way. This approach reduces information overload and delivers relevant, timely, and insightful business updates through a single web platform.

V. PROPOSED SYSTEM

The proposed system is a web-based Business News Aggregation application developed using Java technology. The system is designed to collect business news and related content from multiple online platforms, such as business news websites and multimedia sources. Instead of visiting different websites, users can access all relevant business news in one place. When a user registers or opens the system for the first time, they are asked to select their preferred business categories or topics, such as finance, stock market, startups, economy, or corporate news. Based on these preferences, the system customizes the news feed for each user.

The system uses content-based filtering techniques to analyze news articles and recommend content that matches the user's selected interests. Natural Language Processing (NLP) techniques are applied to process the text of news articles and to support a chatbot feature. This chatbot allows users to search for business news, ask questions, and receive instant responses or summaries related to their interests. Additionally, sentiment analysis is performed on the aggregated news content to identify whether the news carries a positive, negative, or neutral tone, helping users understand market trends and public sentiment more easily.

Overall, the proposed system aims to provide a personalized, intelligent, and user-friendly news aggregation platform. By combining multimedia content aggregation, NLP-based chatbot interaction, sentiment analysis, and personalized content filtering, the system improves the way users consume business news. This approach saves time, reduces information overload, and delivers relevant and meaningful business updates through a single web-based application.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

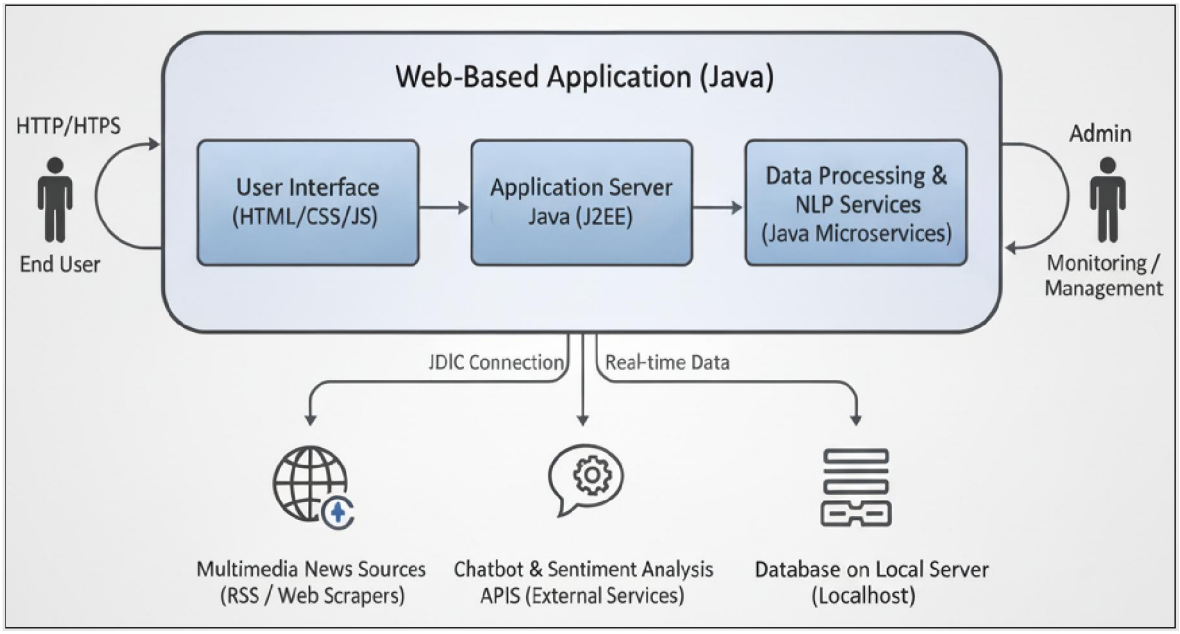

Fig.1: System Architecture Design

System design is an important stage of software development that defines the overall architecture, data flow, and interactions between different components of the system. It provides a clear understanding of how the system works internally and how each module communicates with others to achieve the desired functionality. The design also ensures that the system is efficient, maintainable, and easy to expand in the future.

VI. CONCLUSON

In this project, a web-based Business News Aggregation system is designed and developed using Java technology to provide users with a centralized platform for accessing business news from multiple sources. The system effectively collects and organizes business news articles from different multimedia platforms, making it easier for users to stay updated without visiting multiple websites. By allowing users to select their preferred news categories during registration, the system delivers personalized content that matches individual interests.

The integration of NLP-based chatbot support, sentiment analysis, and content-based filtering enhances the overall user experience by making news discovery more interactive and meaningful. Sentiment analysis helps users understand the overall tone of business news, while content-based filtering ensures relevant article recommendations. Overall, the proposed system offers an efficient, intelligent, and user-friendly solution for business news acquisition and aggregation, demonstrating the practical application of modern web and data processing techniques in real-world information systems.

ACKNOWLEGEMENT

We would like to sincerely thank the researchers and publishers for making their valuable resources available. We are also grateful to our guide for their constant support and guidance, and to the reviewers for their insightful suggestions. Finally, we all thank to the college authorities for providing the necessary infrastructure and support throughout the course of this project.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, March 2026
Impact Factor: 8.2

REFERENCES

[1] Cem Tekin, and Mihaela van der Schaar, "Contextual Online Learning for Multimedia Content Aggregation", IEEE Transactions on Multimedia, April-2024.
[2] Fengwei An, Hans Jürgen Mattausch, "K-means clustering algorithm for multimedia applications with flexible HW/SW co-design", Journal of Systems Architecture 59 (2013) 155–164.
[3] "Web Scraping using Natural Language Processing (NLP)", by V. Pichiyan et al., 2023. (ACM / Elsevier) — shows how NLP techniques improve scraping of structured data.
[4] "A web framework for information aggregation and multilingual content collection", R. Kotsakis et al., 2023. — describes a framework to collect and aggregate multi-language web content.
[5] "From Data to Insight: Transforming Online Job Postings into Labor-Market Intelligence", G. Tzimas et al., 2024. Information 15(8). DOI: 10.3390/info15080496 — outlines how to extract structured info from job portals using NLP.
[6] "Technical Job Recommendation System Using APIs and Hybrid filtering", N. Kumar et al., 2022 — uses hybrid recommender techniques in job domain.
[7] "Web Scraping Approaches and Their Performance on Modern Websites", Ajay Sudhir Bale et al., 2022 — a comparative study of scraping methods.
[8] "Web Scraping Techniques and Applications: A Literature Review", Chaimaa Lotfi et al., 2023 — reviews modern web scraping methods across domains.