



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Implementation of News Classification System Based on Area

Prachi Prajapati¹, Sakshi Tagalpallewar², Ruchita Bacchuwar³, Krishnanjali Shinde⁴

BE Students, Department of Information Technology, Genba Sopanrao Moze College of Engineering, Pune, India^{1,2,3}

Assistant Professor, Department of Information Technology, Genba Sopanrao Moze College of Engineering, Pune, India⁴

ABSTRACT: In 21st century the internet is filled with loads of news articles, there is a pressing need to classify news according to the requirements of an individual. People are generally more interested what is going on, in their immediate surroundings. News has a vital role in the society. Most people read news every day to keep up with the latest information and trends. The information could be anything, from technology, disaster, politics, even the affair of the celebrities. After they absorb the information and understand it, it will be used by the people as a reference to their ideology and decision making. With the help of technology advancements, news disseminates relatively quick across the globe. Using the internet, people can send information from another side of the world in under a second. Because of this, almost any kind of information such as knowledge, idea, entertainment, and news from the people can easily spread to the community. With the development of the web, and ton of internet sites that provide similar information and data. So, users find it hard to decide that of those websites will offer the specified information inside the foremost valuable and effective way.

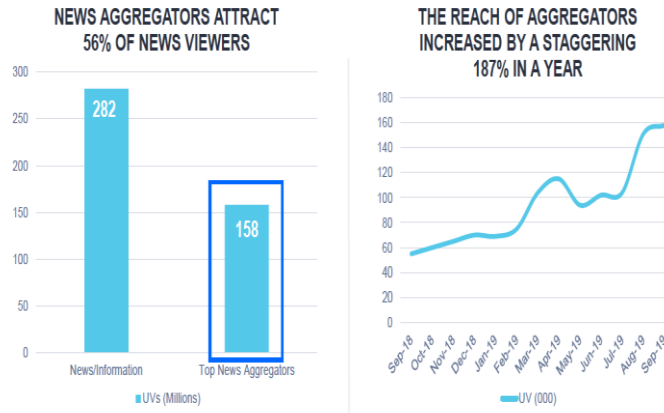
In the previous few years, the globe had unimaginable and huge growth within the rate of reports that's published. Individuals sleep in a time jam-packed with info, data, and news. So, today news has a vital part and position inside the community. As people read the news daily to stay up with the foremost recent data and inputs. The information is also regarding technology, sports, weather, food, and celebrities or many different fields simply think about all the various websites that are out there and also the info they provide. Visiting each single one that you're interested in will be boring and longtask. Most of the people don't have enough time to try and do it, but there are solutions to it. That's where news aggregators comes in.

I. INTRODUCTION

News aggregator is just an online software which collects new stories and events round the world from numerous sources tied one place. News aggregator plays a really necessary role in reducing time consumption, as all of the news that might be explored through quite one web site are placed solely in a single location. Also, summarizing this aggregate content fully can save reader's time. In this paper, we tried to solve this issue by classifying the news articles on basis of cities and providing the entity with the gathering of town specific news. We've developed our own net crawler for content extraction from the hypertext mark-up language pages of reports articles. Random Forests, Naive Thomas Bayes and SVM classifiers are used for classification of reports articles. Results exhibit that machine learning techniques will be controlled to realize our goal and thus demand more

analysis to enhance the efficiency of resolution this issue. News aggregator websites enable users to look at news and updates from various sources at one convenient location. They fetch the info, organize them in tags / classes, and show it within the right order for easier consumptions. Some well famous news aggregation websites are FEEDLY, Google news, ALLTOP, News360, Panda, TECHMEME, Flipboard, Pocket, etc.

News aggregators as gateway to news consumption.



Graph 1 TERMS USED:

Web scrapping:

Web scrapping also called as web harvesting or web data extraction is process of collecting or extracting information or data from different websites. It will directly access the World Wide Web using hypertext transfer protocol or web browser. This technique can be implemented manually as well as automated process .For automated process in to make use of bot or web crawler. Web crawler is additionally referred to as spiderbot it's an internet bot that consistently browses the world wide web usually for net indexing purpose.

News API:

An application programming interface (API) is a connection between computers or between computer programs. It is a type of software interface , offering a service to other pieces of software . A document or standard that describes how to build such a connection or interface is called an API specification .

RSS feeds:

RSS stands for Really simple Syndication and it's may be an easy, standardized content distribution method that may assist you keep up-to-date together with your favorite newscasts, blogs, websites, and social media channels. Rather than visiting sites to search out new posts or subscribing to sites to receive notification of latest posts, one will scan RSS posts through RSS post column in news websites and then reading in RSS reader.



Figure 1

II. IMPLEMENTATION

- 1) Collect data from various of news websites like Hindustan times, India today, India news etc.
- 2) For data collection use RSS (really simple syndication) and news API (application programming Interface).
- 3) create website using html, CSS, PHP and Jquery for user interface
- 4) MySQL database is used to store user sign in information
- 5) Once user enter the city or state name the news articles related to that particular area gets displayed

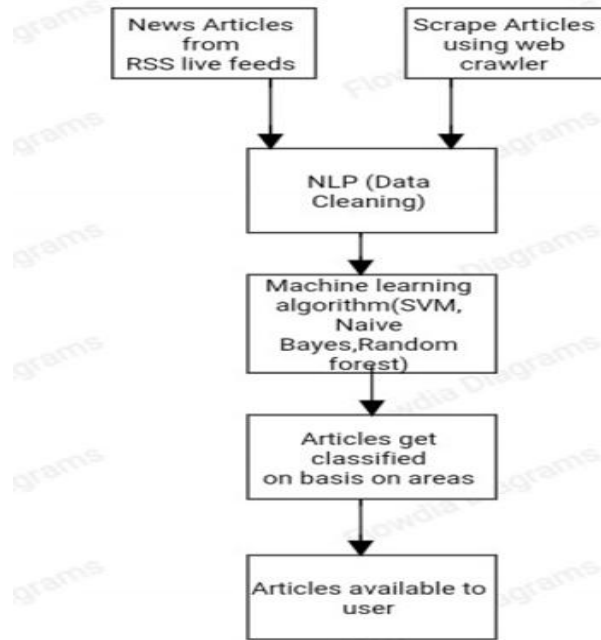


Figure 2 - Flowchart of system

III.RESULTS

Below are the results of code implementation.

Figure 3 of shows homepage of news website .It has various subsections like latest news, entertainment, sports, coronavirus, health, politics etc.



Figure 3: Homepage of website

Figure 4 is search page where users need to enter the name of place they want to get news about.

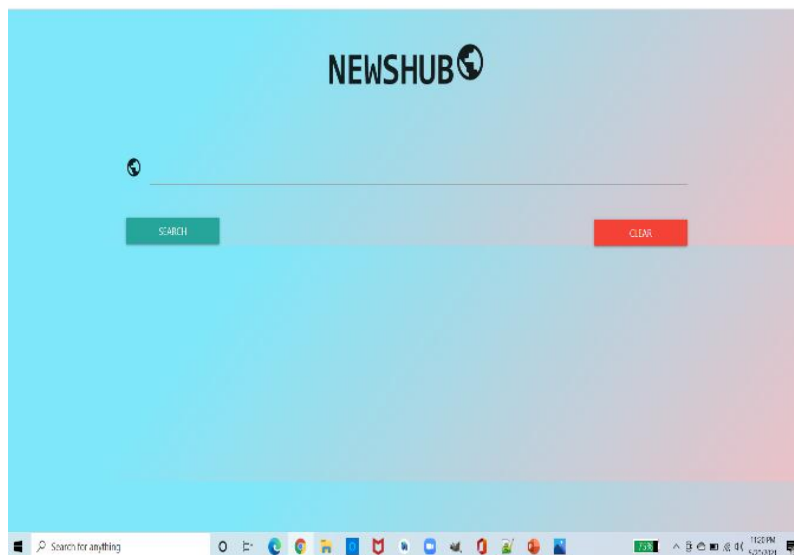


Figure 4: Search Engine

Figure 5 and 6 indicates that once user enters place name and click on search button news articles related to that area are displayed on screen.



Figure 5: News results of Mumbai



Figure 6: News results of Pune

Figure 7 and 8 shows that users can also search news articles related to fields like entertainment, sports, technology, health, politics etc.



Figure 7: Entertainment News



Figure 8: Sports News

ADVANTAGES OF SYSTEM:

- Proposed system will classify news articles which are scraped using web crawlers on basis of areas like city, state, country using machine learning algorithms.
- System will also classify news according to genres like sports, business, politics, entertainment, fashion, etc.
- News articles are scraped from RSS live feeds and news API of 50+ news websites
- Easy navigation and simple to use website.
- **FUTURE SCOPE:**
 - We can create a live chat window for enhancing interaction between same interest users.
 - Live blogs and post by different authors and publishers .
 - Newspapers of different publications can be made available to readers to read online on website as well as offline by downloading it.



IV. CONCLUSION

In this implementation paper we have specified the design and implementation process of news aggregator system that classifies news on basis of area (locations). It provides news based on particular area like city, state, country. The main motto behind initiating this website is to help users to get access to most relevant news articles from various news website at one place. It saves user's time and efforts.

The Goals mentioned in the future scope can be implemented in future .

REFERENCES

1. Vighnesh Rao, Jayant Sachdev, "A machine learning approach to classify news articles based on location". IEEE Xplore Compliant - Part Number: CFP17M19-ART, ISBN: 978-1-5386-1959-9.
2. Zaiyinh Wang, Baha Song, "Research on hot news classification algorithm based on deep learning". 2019 IEEE 3rd Information technology, Networking, Electronics and Automated Control Conference.
3. D. Shen, Z. Chen, Q. Yang, S.H. Zeng, B. Zhang, Y. Lu, and W. Ma, Web-page classification through summarization, in Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2004, pp. 242249
4. S. Chowdhury and M. Landoni, "News aggregator services: user expectations and experience", Online Information Review, 30(2), pp.100-115, 2006
5. Scrapinghub, Ltd., "Scrapy," GitHub repository, 6 December 2016. [Online]. Available: <https://github.com/scrapy/scrapy>. [Accessed 27 March 2017].
6. "An automatic e-news article content extraction and classification" by Dilini Dandeniya. 2018 IEEE 18th International conference in advance ICT for Emerging reasons (ICTer).
7. "A self-pruning classification model for news" by Leonidas Akritidis, Athanasios Feygas, Panayiotis Bozanis, Miltiadis Alamanitis . 2019 IEEE 10th International Conference on Information, Intelligence, Systems and Applications (IISA).
8. Nepali news classification using Naïve Bayes, Support Vector Machines and Neural Networks (IEEE) . Tej Bahadur Shahi, Ashok kumar Pant.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details