

# News Aggregation using Web Scraping News Portals

Mr. Mayur Bhujbal<sup>1</sup>, Ms. Bhakti Bibawanekar<sup>2</sup>, Dr. Pratibha Deshmukh<sup>3</sup>

Bharati Vidyapeeth's Institute of Management and Information Technology, Navi Mumbai, India  
mayurbhujbal3@gmail.com, bibawanekarbhakti@gmail.com, pratibha.deshmukh@bharativedyapeeth.edu

**Abstract:** *In today's fast-paced world, staying informed through reliable and high-quality news is important. However, with an overwhelming number of online publishers and platforms, accessing information from multiple sources can be time-consuming and challenging. To simplify this process, news aggregators have emerged as a solution. These platforms allow users to customize their news consumption by selecting preferred websites and receiving curated articles from those sources in one centralized location. This saves valuable time and effort but also streamlines information gathering in our daily routines. To develop an influential online news aggregator, web scraping for data structuring is crucial. Web scraping involves analyzing a website's HTML structure to extract the desired data. By understanding the underlying structure of web pages, developers can fetch relevant information such as article titles, summaries, authors, and publication dates. People often struggle to find valuable news sources amid the vast sea of information available online. News aggregators address this issue by providing personalized news feeds based on individual interests. Users serve as valuable resources for knowledge tailored to specific preferences. While news aggregators are widely recognized for their popularity and utility, there is still room for improvement in their role in software development. Enhancements could include refining algorithms for more accurate content curation. In addition, Users could include implementing better user interfaces for enhanced user experiences and exploring innovative features like sentiment analysis and topic clustering. By continually iterating and improving upon existing news aggregator platforms, developers can ensure users have access to timely, relevant, and trustworthy news conveniently and efficiently.*

**Keywords:** Web scraping, Online news portals, Information gathering, News Aggregation

## REFERENCES

- [1]. K. Sundaramoorthy, R. Durga and S. Nagadarshini, "NewsOne — An Aggregation System for News Using Web Scraping Method," 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), 2017, pp. 136-140, doi: 10.1109/ICTACC.2017.43.
- [2]. E. N. SARR, O. SALL and A. DIALLO, "FactExtract: Automatic Collection and Aggregation of Articles and Journalistic Factual Claims from Online Newspaper," 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2018, pp. 336-341, doi: 10.1109/SNAMS.2018.8554421.
- [3]. T. Tona Landu et al., "Machine Learning Algorithm for Text Categorization of News Articles from Senegalese Online News Websites," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), 2022, pp. 1-8, doi: 10.23919/CISTI54924.2022.9820408.
- [4]. Authors, Joel Christian, Sharada Valiveti & Swati Jain, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Chharodi, Ahmedabad, 382481, Gujarat, India
- [5]. M. I. Akrianto, A. D. Hartanto and A. Priadana, "The Best Parameters to Select Instagram Account for Endorsement using Web Scraping," 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 2019.

- [6]. S. Ayyappan and S. Matilda, "Criminals and Missing Children Identification Using Face Recognition and Web Scrapping," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1- 5, doi: 10.1109/ICSCAN49426.2020.9262390.
- [7]. B. B. P. Maurya, A. Ray, A. Upadhyay, B. Gour and A. U. Khan, "Recursive Stock Price Prediction with Machine Learning and Web Scrapping for Specified Time Period," 2019 Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN), 2019, pp. 1-3, doi: 10.1109/WOCN45266.2019.8995080.
- [8]. S. K. Malik and S. Rizvi, "Information Extraction Using Web Usage Mining, Web Scrapping and Semantic Annotation," 2011 International Conference on Computational Intelligence and Communication Networks, 2011, pp. 465-469, doi: 10.1109/CICN.2011.97.
- [9]. A.-U. -N. Fatima, H. Ahmad, M. Ahmad, W. Ahmad and N. Faisal, "Extraction of Temporal Events' Frequency from Online News Channels," 2020 30th International Conference on Computer Theory and Applications (ICCTA), 2020, pp. 109-116, doi: 10.1109/ICCTA52020.2020.9477659.
- [10]. Jain and A. Kasbe, "Fake News Detection," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2018, pp. 1-5, doi: 10.1109/SCEECS.2018.8546944.
- [11]. Alrashed T, Almahmoud J, Zhang AX, Karger DR (2020) Scrapir: making web data apis accessible to end users. In: Proceedings of the 2020 CHI conference on human factors in computing systems.
- [12]. Đ. Petrović and I. Stanišević, "Web scrapping and storing data in a database, a case study of the used cars market," 2017 25th Telecommunication Forum (TELFOR), 2017, pp. 1-4, doi: 10.1109/TELFOR.2017.8249451.
- [13]. S. Mehak, R. Zafar, S. Aslam and S. M. Bhatti, "Exploiting Filtering approach with Web Scrapping for Smart Online Shopping: Penny Wise: A wise Tool for Online Shopping," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-5, doi: 10.1109/ICOMET.2019.8673399.
- [14]. D. Kurniawati and D. Triawan, "Increased information retrieval capabilities on e-commerce websites using scraping techniques," 2017 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, 2017, pp. 226-229, doi: 10.1109/SIET.2017.8304139.
- [15]. D. Mishra and N. Pujari, "Cross-domain query answering: Using Web scrapper and data integration," 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011), Allahabad, India, 2011, pp. 27-32, doi: 10.1109/ICCCT.2011.6075193.
- [16]. F. Lin and S. Chih Yang, "Web Scraping to Implement Tape Reading on Taiwan Stock Periodically with GUI," 2021 IEEE 3rd Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, 2021, pp. 112-116, doi: 10.1109/ECICE52819.2021.9645633.
- [17]. Deshmukh, P., Avinash, S., Gonsai, A.M., Sonawane, S.S., Khan, T. (2023). Ordering Services Modelling in Blockchain Platform for Food Supply Chain Management. In: Rathore, V.S., Tavares, J.M.R.S., Piuri, V., Surendiran, B. (eds) Emerging Trends in Expert Applications and Security. ICE-TEAS 2023. Lecture Notes in Networks and Systems, vol 681. Springer, Singapore. [https://doi.org/10.1007/978-981-99-1909-3\\_37](https://doi.org/10.1007/978-981-99-1909-3_37)
- [18]. G, M., Deshmukh, P., N. L., U. K. ., Macedo, V. D. J. ., K B, V. ., N, A. P. ., & Tiwari, A. K. . (2023). Resource Allocation Energy Efficient Algorithm for H-CRAN in 5G. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 118–126. <https://doi.org/10.17762/ijritcc.v11i3s.6172>
- [19]. K. S. Gill, A. Sharma, V. Anand, R. Gupta and P. Deshmukh, "Influence of Adam Optimizer with Sequential Convolutional Model for Detection of Tuberculosis," 2022 International Conference on Computational Modelling, Simulation and Optimization (ICCMO), Pathum Thani, Thailand, 2022, pp. 340-344, doi: 10.1109/ICCMO58359.2022.00072.
- [20]. Avinash, S., Deshmukh, P., Jamsandekar, P., Kumbhar, R.D., Kharade, J., Rajendran, R. (2023). Specifying the Virtual Reality Approach in Mobile Gaming Using Unity Game Engine. In: Rathore, V.S., Piuri, V., Babo, R., Ferreira, M.C. (eds) Emerging Trends in Expert Applications and Security. ICETEAS 2023.

Lecture Notes in Networks and Systems, vol 682. Springer, Singapore. [https://doi.org/10.1007/978-981-99-1946-8\\_26](https://doi.org/10.1007/978-981-99-1946-8_26)

- [21]. Sakhare, Mr & Deshmukh, Pratibha. (2022). Blockchain in Petroleum Industry. International Journal of Advanced Research in Science, Communication and Technology. 733-737. 10.48175/IJARSCT-5751.
- [22]. Prerana Patil and Dr. Pratibha Deshmukh (2022). Case Study of Indian Students Going to Abroad for Medical Studies. International Journal of Scientific Research in Engineering and Management (IJSREM).