INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
*(Volume 8, Issue 1 - V8I1-1418)*
Available online at: https://www.ijariit.com

# News aggregator web application

*Kartik Bhatnagar*
bhatnagararchana19@gmail.com
*HMR Institute of Technology and Management, New Delhi, Delhi*

*Arya Tomar*
aryatomr@gmail.com
*HMR Institute of Technology and Management, New Delhi, Delhi*

*Dr. Shalini Goel*
profshalinigoel1803@gmail.com
*HMR Institute of Technology and Management, New Delhi, Delhi*

## ABSTRACT

***News Aggregator is a web app that collects and displays all of the latest news stories and events happening around the world in one place. It helps reduce time consumption. One other thing to do is to provide a few lines of text to summarize the article. This paper presents a system that collects news from various electronic publishers and distributors.***

***Keywords***: *News Aggregator, Web App, Python*

## 1. INTRODUCTION

In recent years, the number of users and the amount of material, data, and services available on the World Wide Web has grown at a self-feeding rate. More content makes the Web more appealing to more users, who create more content in turn.

Web 2.0 technology and the ever-increasing potential for user-generated media appear to be hastening this spiral effect. The news sector, for example, appears to be going entirely online and attempting to keep up with innovations in Web publication. The majority of news publishers have begun to offer electronic copiesof their content, which are sometimes far more structured than traditional print editions. A variety of intermediate services, such as topical news portals, have alsoemerged, which aggregate and re-distribute information from multiple sources.

In this way, the end user has obtained access to a massive amount of data, which, aside from its obvious benefits, also brings with it the problem of information overload. For the non-expert user, the process of discovering useful information among all that is available is as intimidating and frustrating as looking for a needle in a haystack. As a result, if we want to promote this exciting development, we'll need to come up with better and easier ways to obtain important data.

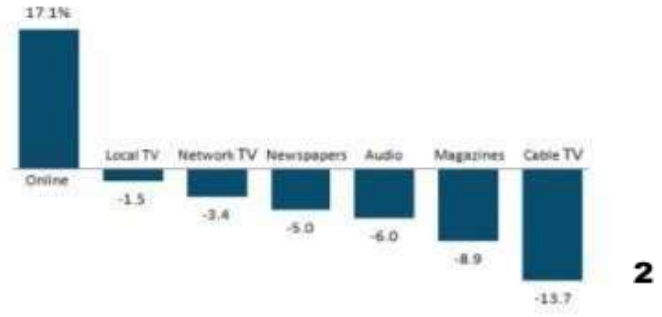



Fig. 1. Online feeds Develops Quickly contrast to Others

According to an Outsell study (2009), 57 percent of feeds media customers use computerised sites, and they are 31% more likely to go to an aggregator than to a daily newspaper location. 8% of the time, or other news sites 18% of the population Feeds aggregator combinesnews data, and regularly

Academic Report briefs it in a good format and design for the reader, from various sites, newspapers,and agencies. News aggregators are typically listed in categories like "Websites Every Engineer Should Visit." Despite the benefits of providing people with a wealth of information via the internet, we will face a new problem: information overload. There will be an excessive amount of information in front of the user, which may not be relevant to his needs. The proposed system can solve this issue. News Aggregator organises feeds by subject, as it appears to be a gateway that connects several feed websites. It might be a website that compiles data and news from a variety of sources and presents it all in one place. By aggregating content based on viewing history, this service makes it easier for readers to find and consume news. News aggregation is a method of obtaining information from a variety of sources. Using news aggregation is one of the best ways to stay on top ofthe news and topics you want. They offer convenience and time-saving features

## 2. NEWS SERVICE

The writers focused on gathering news in three main parts: the first and second steps are data gathering and extracting articles from websites and saving them inthe database; and the third and fourth steps are data gathering and extracting articles from websites and saving them in the database. The final phase is visualisation, which allows the user to see the important article.

The authors wanted to collect content from a variety of sources, including articles and news headlines from blogs and websites.

The authors wanted to use a web-based interface to process information from a certain source and then change it according to news categories and personalised web views. They explain how they use HTML to do the content scanner. The firstphase is wrapping, which entails matching the URL address of new items from the source with the appropriate category for the news, storing the address in the database for each category pair, and combining it with the appropriate wrapper.

The second phase in the wrapping process is to acquire information from the new items that will be used for retrieving and indexing the article; for each article, they obtain the first sentence and transfer it to the appropriate HTML page.

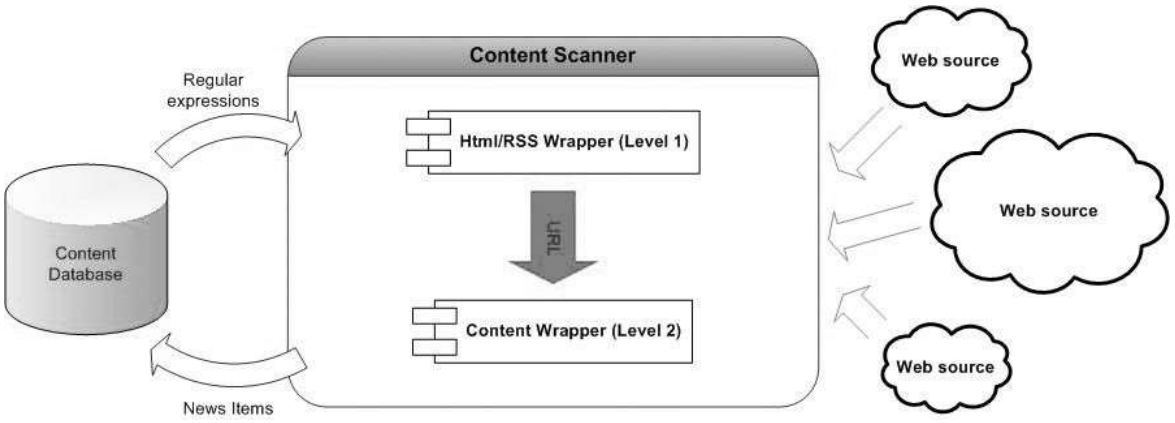
The authors wanted to compile news from a variety of sources, including websites, newspapers, magazines, and television, and put it all on one webpage.Because the contents and data are concise and summarised, it advances the quality of the outcomes. As a result, their work is centred on the Rich Site Summary fetcher, which allows them to retrieve Rich Site Summary data from certain websites at a specific time. In addition to TOI, they use web crawling toacquire more precise results. Web scraping is a technique for extracting large volumes of data from websites. The quality of the aggregator system is still an open field to be introduced, based on all of the above mentioned researches on the news aggregator.

## 3. PROPOSED SYSTEM

The basic structure of the proposed system, which would be able to collect internet news from cloud services and save user reading time, is discussed in thispart.

The content scanning module is in charge of locating and retrieving new items from a list of pre-specified sources, as well as storing basic indexing information in the database to enable for item personalization and retrieval.

The aggregation is done offline using a Web crawler that is summoned on a regular basis. The spider searches the database for a list of sources and their related URL addresses. It follows a two-stage method for each source: (a) identifying the addresses of new items in the source, (b) retrieving the items andextracting the information required for indexing them. For each of these two subprocesses an HTML wrapper is invoked, i.e., a small parsing module that identifies the required information within each Web page.




The initial level of wrapping (HTML/RSS wrapper) entails determining the URL addresses of new items in the source as well as per news category. As a result, an address is kept in the database for each source-category pair, along with a corresponding wrapper, which is a set of regular expressions. If the received page is an RSS article, the wrapper must parse the resulting XML file and identify the URL addresses of new items as well as any associated information. XML tags areused to mark the address of each time, each title, and other information in RSS documents. The address and title are the only pieces of information we have right now. If on the other hand, the source is not RSS, an HTML page is parsed by the wrapper and the addresses and titles of new items are extracted. So far, we are extracting only the first sentence of each article, parsing the corresponding HTMLpage. In future versions of the system, additional information, such as keywords from the content of the article will be extracted.

## 4. CONCLUSION

In this paper we have presented the News Aggregator which curates news fromvarious Web sources and provides access to it.

## 5. ACKNOWLEDGEMENTS

We would like to express our special thanks to our advisor Dr. Shalini Goel for supporting us and for her patience and motivation for our project. As her guidancehelped us in writing this paper.

## 6. REFERENCES

[1] Jagan, S., & Rajagopalan, S. P. (2015). A survey on web personalization of web usage mining. International Research Journal of Engineering and Technology (IRJET), 2(01), 6-12.

[2] Sigletos, G., Paliouras, G., Spyropoulos, C. D., Hatzopoulos, M., & Cohen, W. (2005). Combining Information Extraction Systems Using Voting and Stacked Generalization. Journal of Machine Learning Research, 6(11).

[3] Jeon, D. S., & Nasr, N. (2016). News aggregators and competition among newspapers on the internet. American Economic Journal: Microeconomics, 8(4), 91-114.

[4] Aggarwal, C. C., & Philip, S. Y. (2002, January). An automated system for web portal personalization. In VLDB'02: Proceedings of the 28th International Conference on Very Large Databases (pp. 1031-1040). Morgan Kaufmann.

[5] Wang, Q. (2020). Normalization and differentiation in Google News: A multi-method analysis of the world's largest news aggregator (Doctoral dissertation, Rutgers The State University of New Jersey, School of Graduate Studies).

[6] Cobos, T. L. (2017). New Scenarios in News Distribution: The Impact of News Aggregators Like Google News in The Media Outlets on the Web. Edited by Simone Tosoni, Nico Carpentier, Maria Francesca Murru, Richard Kilborn, Leif Kramp, Risto Kunelius, Anthony McNicholas, 95.

[7] Chiou, L., & Tucker, C. (2017). Content aggregation by platforms: The case of the news media. Journal of Economics & Management Strategy, 26(4), 782-805.

[8] Mohamed, A., Ibrahim, M., Yasser, M., Ayman, M., Gamil, M., & Hassan, W. (2020). News Aggregator and Efficient Summarization System. methods, 11(6).