

NEWSLY: FILTERING AND AGGREGATING NEWS

Prachi Yerojwar¹, Pratiksha Mesharam², Shubhangi Mesharam³, Anand Singh⁴, Sumedh patil⁵,

Dr. J.M.Bhattad⁶

B.E. Students^{1,2,3,4,5}, Professor⁶

Department of Electronics and Telecommunication Engineering

Priyadarshini College of Engineering, Nagpur, Maharashtra, India

ABSTRACT

With so many websites providing different news on different platforms it is difficult to visit every page. In this paper, we describe an online platform where we are filtering, aggregating news and providing a user-friendly experience using HTML, CSS, Javascript at the front end, Python, Django for the back end and also NLP. This will help in saving time and provide quality content.

Keywords: Agregator, filtering, news, website

I.INTRODUCTION

In this online world, sources of information are increasing at a rapid speed. According to the estimates, there are around two hundred million active websites out of one and a half billion websites on the whole world wide web. This is the amount of information we are looking at when we search for a topic. On a single topic when we search in a search engine we get a lot of result pages to visit and on an average about ten links on every page, this surely makes everyone's work to get information very tedious and time consuming. This is the reason we tried to bring forth a news aggregator. News aggregator is a web application which aggregates data from multiple websites and then presents the data at one location. This works as the very important start of a busy day where you want to just take a glimpse on the best quality news highlights. Another related program is Google news which provides you news depending on your search history and interesting topics.

II.METHODOLOGY

In the online news aggregator portal, for a reader visiting the website for the first time will see the homepage from there they are instructed on how to navigate through the website. Users will have to register themself with a username, email and a password. After the successful registration they will have to login into their profile after which they will be directed towards the home page from where they can navigate to their profile section or dashboard.

The profile section allows the user to modify his or her information or even add a profile picture. Dashboard represents all the different categories in which the aggregated news has been divided into. The news scraped is categorised as Technology, Economy, Sports, Politics, Lifestyle & Health and Entertainment. From there you are directed to the news section of the topics you are interested in.

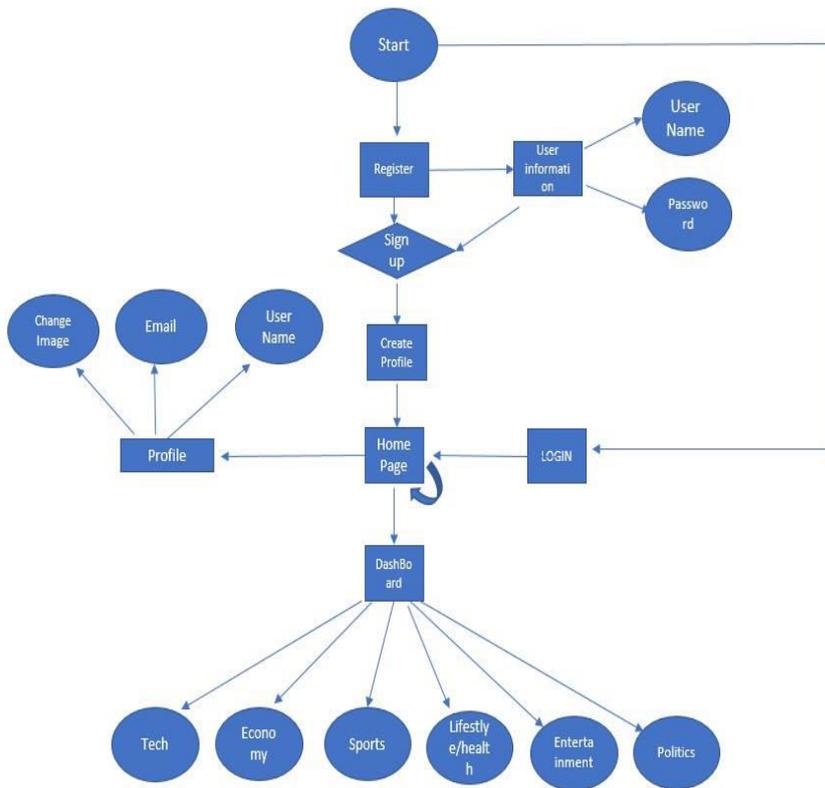


Figure 1. Entity-Relationship Model

III. WORKFLOW

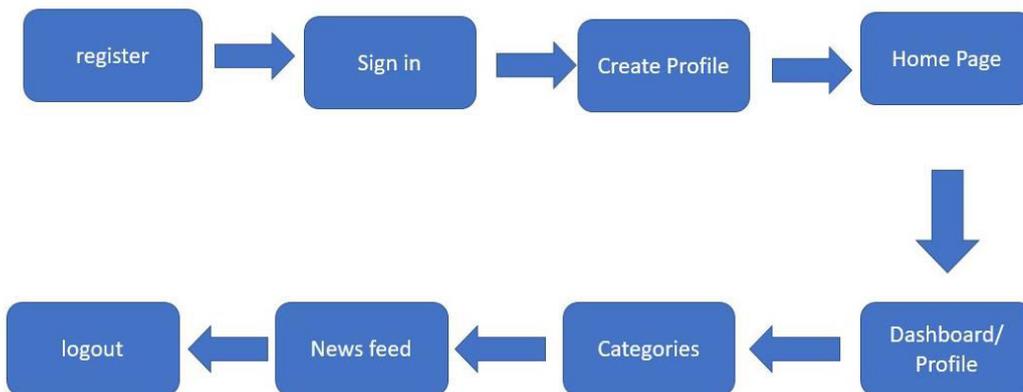


Figure 2. Workflow

(1) PROCESS FLOW

(A) REGISTRATION AND LOGIN PAGE

Visitors visiting the portal for the first time need to register into the portal with their username, email id and password. After registration they will be directed to the login page to login into their profile.

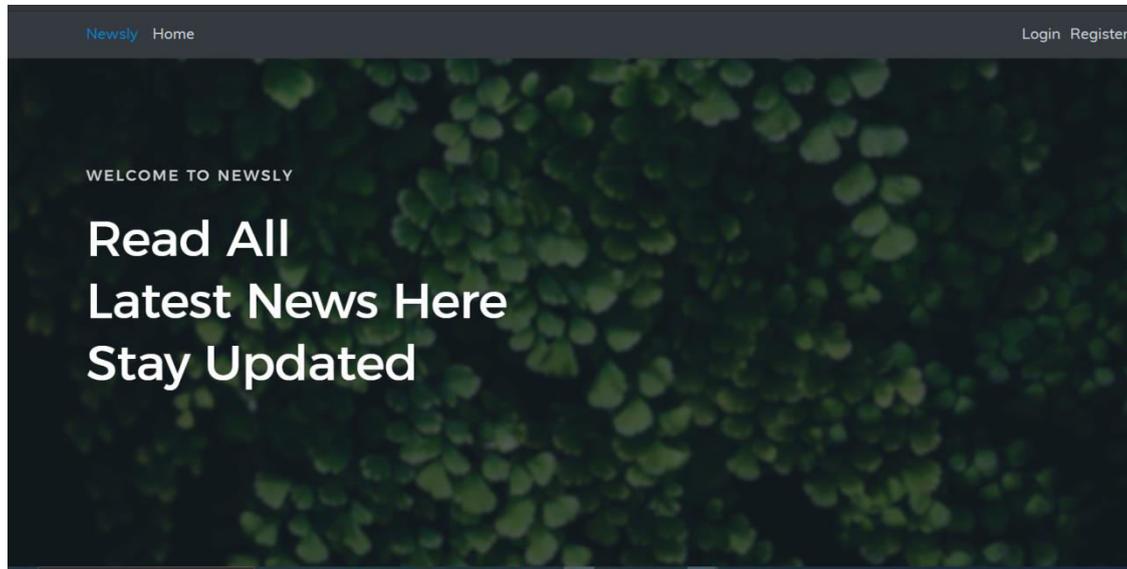


Figure 3. Home Page

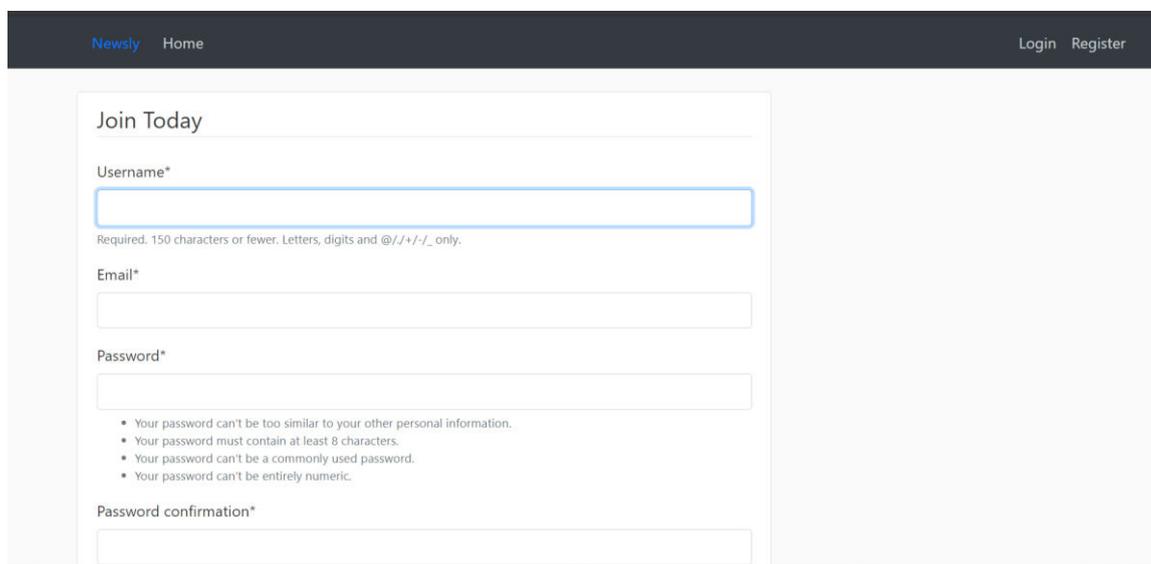
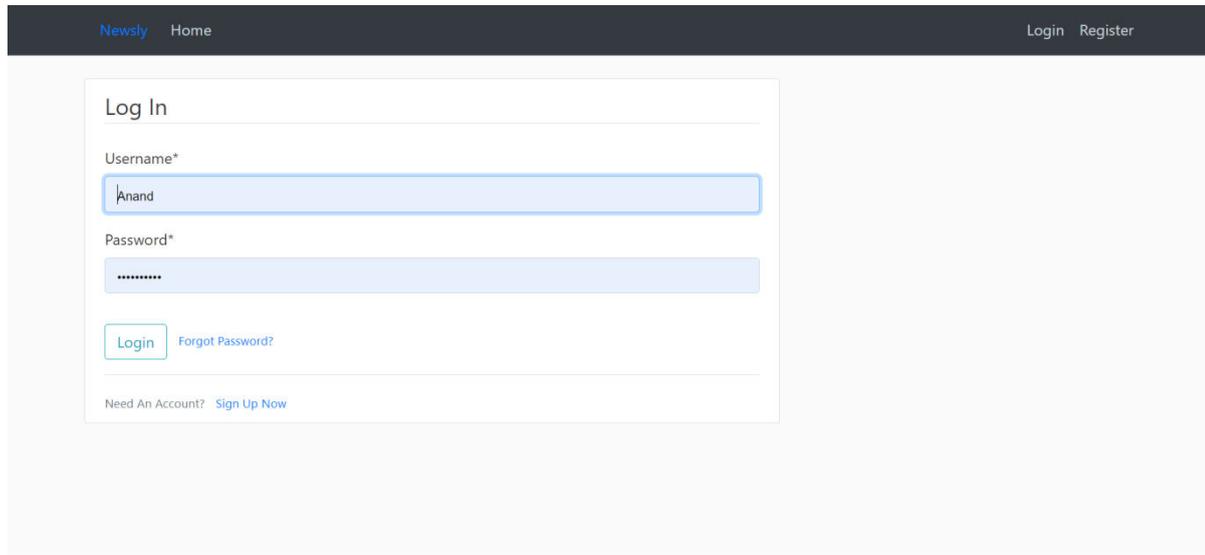


Figure 4. Registration Page

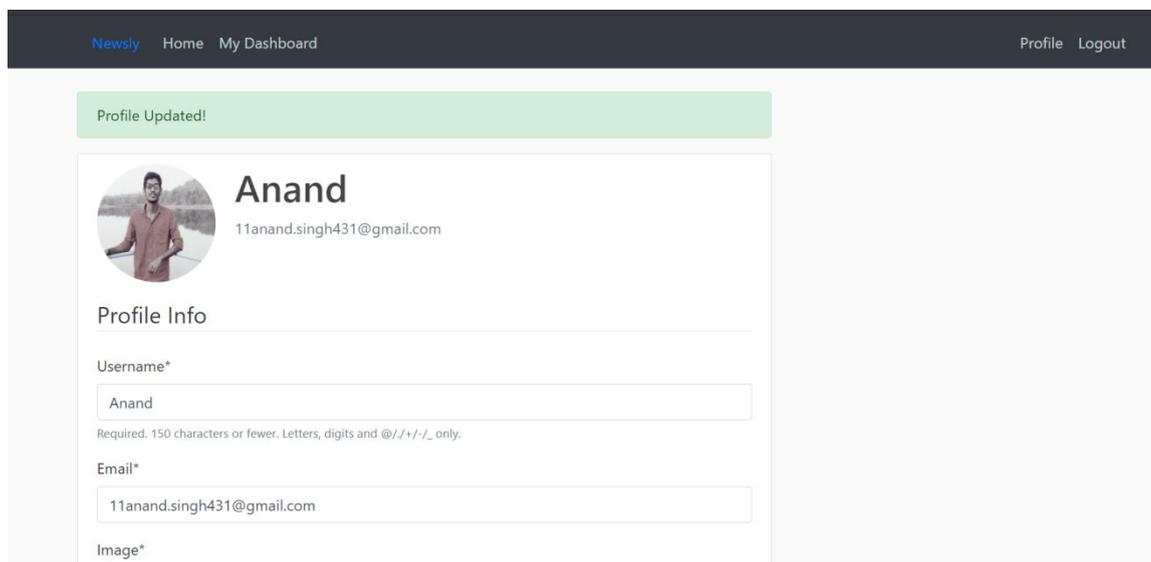


The screenshot shows a login page with a dark header containing 'Newsly Home' on the left and 'Login Register' on the right. The main content area features a 'Log In' form with a title 'Log In' and a subtitle 'Log In'. The form includes a 'Username*' field with the value 'Anand', a 'Password*' field with masked characters '*****', a 'Login' button, and a 'Forgot Password?' link. Below the form, there is a link 'Need An Account? Sign Up Now'.

Figure 5. Login Page

(B) PROFILE PAGE

After successful login into your account you are directed towards the home page from where you can go to either your Dashboard or your Profile. In your profile you can feed your personal details(Name,Email,Picture).



The screenshot shows a profile page with a dark header containing 'Newsly Home My Dashboard' on the left and 'Profile Logout' on the right. The main content area features a green notification bar 'Profile Updated!'. Below this, there is a profile card for 'Anand' with a profile picture and email '11anand.singh431@gmail.com'. The card includes a 'Profile Info' section with a 'Username*' field containing 'Anand' and an 'Email*' field containing '11anand.singh431@gmail.com'. Below the email field, there is a note: 'Required: 150 characters or fewer. Letters, digits and @/./+/-/_ only.' and an 'Image*' field.

Figure 6. Profile

(C) DASHBOARD

In the Dashboard you can select from the categories given as Tech, Economy, Sports, Lifestyle & Health, Entertainment and Politics.

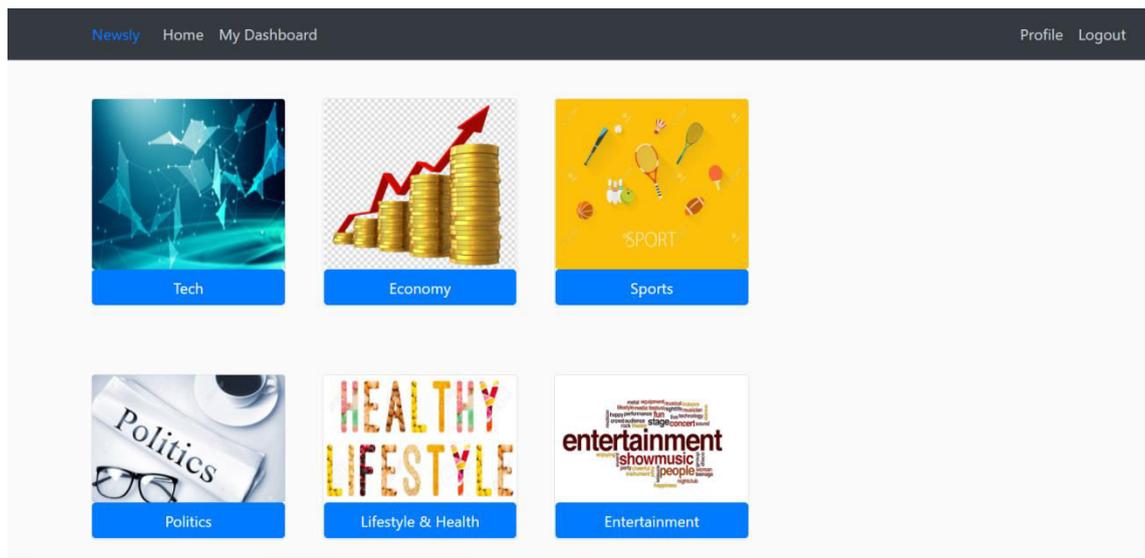


Figure 7. Dashboard

(D) NEWS

After selecting your interested topic in the news section you can read the news aggregated from trusted websites, based on the popularity and the fame and review of the people from past many years.

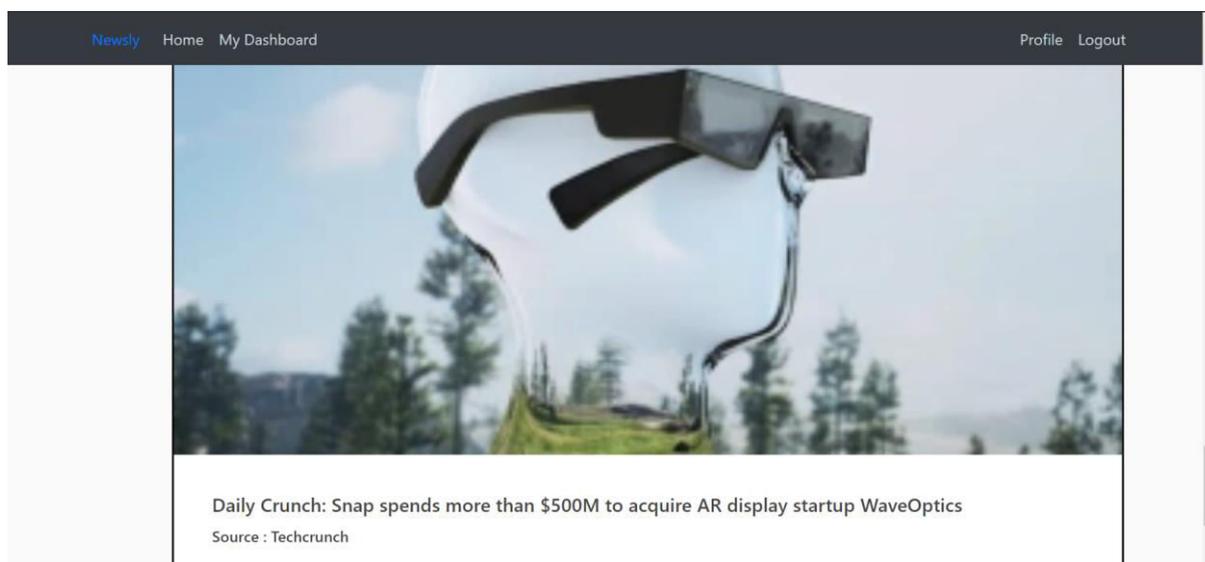


Figure 8. News

(2) HARDWARE REQUIREMENT

A desktop/laptop with a minimum configuration:

- Processor - Pentium(R)1.74 GHZ
- RAM - 1GB
- Hard disk 128GB

(3) SOFTWARE REQUIREMENT

- Operating System:- Windows, MAC OS, UBUNTU
- Code editor:- VS CODE, PYCHARM
- Browser:- Chrome/Mozilla Firefox

(4) TECHNOLOGY USED

(A) FRONTEND

HTML: It is the most basic building block of the web. It used to build web pages. It is supported by all the web browsers. It provides a means to create structured documents by denoting structures such as text, image, table, lists, links and other items.

CSS: It is used to add styles to the html web page, it defines where each element of the would be placed and how. It increases the page efficiency and makes the page load faster.

JAVASCRIPT: It is easy to learn and implement. It gives the programmer using it the ability to create rich interfaces. It is very fast because it often runs immediately within the client side browser.

(B) BACKEND

PYTHON: It is a high-level programming language which has an english like syntax which makes it easier to read the code. It is a free and open sourced language with vast library sources and portability.

DJANGO: It makes easier building web apps quickly and with less code. It takes care of the hassle of web development. It is secure and exceedingly scalable.

CRISPY FORM API: Django-crispy-forms, an application that helps to manage Django forms by allowing adjusting form properties such as send button, method or CSS classes.

IV. CONCLUSION

In this project, we have learnt that aggregators save time and hardwork of going through a large number of websites. During the development process we learnt about the various implementations of the Django framework in Python. We used Natural Language Processing in web scraping. We developed a platform where information will be filtered and provided according to the user requirement.

ACKNOWLEDGMENT

We would like to take the opportunity to express our gratitude to the college Priyadarshini College of Engineering, Higna, Nagpur which follows a goal of creating technocrats and brings it into reality.

We wish to express our sincere thanks to our guide DR. J. M. BHATTAD who continuously supervised our work and provided support at every point of our project on "NEWSLY WEB: FILTERING AND AGGREGATING NEWS".

REFERENCES

[1]Rabiyatou Diouf, Edouard Ngor Sarr, Ousmane Sall, Babiga Birregah, Mamadou Bousso, Sény Ndiaye Mbaye- "Web Scraping: State-of-the-Art and Areas of Application" - 2019 IEEE International Conference on Big Data (Big Data)

[2] Sumit Sakarkar, Vaibhav Chaudhari, Tanmay Gaurkar, Aditya veer, Mayura Kulkarni, SCET MIT Academy Of Engineering, Pune, India-"Web Personalisation based on User Interaction : Web Personalisation", Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)

[3] Shahid Khan, Chandresh Prasad, Irfan Khan, Prof. Tabrez Khan, Amaanulla Khatri Computer Engineering AIKTC Navi Mumbai, India- “Intelligent News Aggregator and Validator” - 2019 International Conference on Nascent Technologies in Engineering.

[4] K. Sundaramoorthy, R. Durga, S. Nagadarshini, Department of Information Technology, Agni College of Technology- “NewsOne — An Aggregation System for News Using Web Scraping Method”- 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)