Google News Aggregator: A Web Application made with React

Aryan Srivastava[A], Sahil Kumar[B], Sourab Sagar[C], Divij Jain[D] , Anjani Gupta[E]

Author Affiliations:
Dr. Akhilesh Das Gupta Institute of Technology & Management
FC-26, Panduk Shila Marg, Zero Pusta Rd, Shastri Park, Shahdara, New Delhi, Delhi 110053

Author Emails:
[A]sriar1111@gmail.com , [B]sahilkhichy01@gmail.com,[C]sourabhsagar17@gmail.com ,
[D]divijjain55@gmail.com, [E]anjaniaggarwal.06@gmail.com

ABSTRACT

In today's society, we are aware that there are several news websites that cover a wide range of topics, just a few of which are of interest to us. A news aggregator can save us a lot of time, and we can fine-tune it to present only the items that we are interested in with a few modifications and filtrations. A news aggregator is a very handy application for quickly accessing daily news. By scraping certain well-known websites and serving the scraped content via React on the web or in any app, we can create a news aggregator application. It's a web application that gathers information from several websites. The articles are then displayed on the desired pages. The news is an important part of the day's beginning. There are several well-known news websites available online. Consider browsing a variety of news websites every day. The time spent reading those websites and gaining information is equally important. It can provide you an advantage over others who don't. We've now made this task incredibly simple for everyone! You may read news from multiple websites using a news aggregator. The news aggregator then collects the articles and displays them on our user-interface. You may view the news of your choosing with a few clicks.

Keywords: News Aggregator, React, Live News API

INTRODUCTION

Information denotes power in modern society, and accessing it has gotten easier because of the advent of the Internet. The need for programmes that can condense data and provide the ability to group and filter it is apparent, given the vast volume of information available online.

The main objective of a news aggregator is to collect news from numerous online sources, such as headlines, blog articles, and social media updates, and to save time by not having to visit each site separately. Our goal is to create a news aggregator that combines the results of news web services with APIs for getting articles [4].

Because each website has its own structure, a solution based on web crawling and HyperText Markup Language (HTML) parsing algorithms is insufficient. Furthermore, even minor changes to the website's Document Object Model (DOM) might have disastrous effects for the processed results. Another well-known means of obtaining articles is through the use of specialised APIs provided by news sites [4].

Some of the few unique features of the application are listed below:

1. Information can be centralised. It cuts down on the time it takes to visit websites on a regular basis. Subscribers may easily browse feeds without having to visit each page, and then click on articles that interest them to learn more.
2. The graphical interface is smooth and simple to use. Despite the fact that a few existing news aggregators can provide far more complex interfaces and facilities, Advanced settings might be challenging to utilise, especially for non-technical people. The created aggregator provides a solution to this challenge by utilising a simple yet comprehensive interface. At the same time, it provides a responsive web interface, available on both large screens and mobile devices.
3. Although most aggregators offer spam and ad filtering for articles, more and more are beginning to incorporate their own advertising in order to generate revenue. The built-in aggregator does not display advertisements to the user.
4. Filtering and sorting of content based on user-defined criteria.

RELATED WORK

First, we'll look at news aggregators from the standpoint of the websites that provide the stories. Columbia University created Newsblaster [5], a news aggregator. This technique is limited to downloading articles from 17 pre-selected news websites and cannot be expanded. Similarly, MediaGist [6] extracts news from only one English news website and four other language news websites (Czech, Italian, French and German).

Hamborg et al. [7] utilises Scrapy, which is a Python framework for extracting HTML news headlines from news sources. The main goal is to extract several web page elements such as the title, news text, lead paragraph, publishing data, news author, and their accompanying image.

Feedly [8] lets you examine RSS feeds in text or a customised format that looks like an e-book reader. RSS feeds can be organised into categories with this aggregator. It also provides news summaries, which cuts down on reading time.

Flipboard [9] is another well-known RSS feed management programme. Flipboard, like Feedly, allows you to organise your feeds into categories, as well as update and adjust them.
The European Media Monitor (EMM) [10] is a large project run by the Joint Research Centre of the European Commission. EMM monitors social media in real time across Europe, collecting between 80 and 100.000 news stories per day in up to 50 languages. Over 2,000 hand-selected web sources and 20 commercial news providers were used to compile the statistics. EMM crawls the sources once every five minutes to ensure that the articles are current.

Second, some news aggregators concentrate on analysing rather than gathering news pieces. For example, NewsBird [11] uses multilingual news datasets that contain over 1.5 million articles gathered from about 4.000 news outlets in over 100 countries from October to November of the year 2014. Non-English articles are automatically translated into English using automated translation techniques; however, due to the deficiency of the translation, this does not result in high-quality articles.

Finally, individualised news aggregators must be considered. PNS [12] provides additional flexibility to users by taking into consideration the user's personal information, which is collected during registration, as well as the user's news search history.

ARCHITECTURE

We all are familiar with how newspapers have shaped the centuries. Every morning, it is a very frequent identification through which we learn about various news stories from many categories that are popular throughout the world. A digital newspaper is the same as the news application. The only change is that now news is not just restricted to news on paper, but rather we now receive it on our mobile or device screen. It might be the next stage in the digital world.
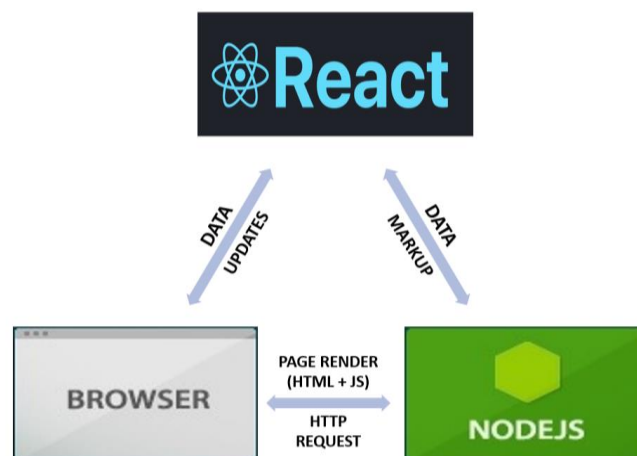

Figure 1: React.JS Server Rendering

You must have npm installed on your PC as a prerequisite. It is necessary to react. The npm package manager is a JavaScript package manager. To get started with npm, go to Node.js and download the most recent version.

React App: React is a front-end JavaScript library for creating user interfaces or UI components that is free and open-source. It is managed by Facebook and a community of independent developers and designers. It's a well-known framework for developing single-page apps.

Getting Started with React JS: To launch our first react application, we'll need to open our working directory and a suitable IDE. Pycharm is used in this project. It's an excellent React IDE. It gives us several important React extensions that make writing code easier.
Enter the command "npm create-react-app React-Live-google-news-API-Project" into the terminal. It will install a number of files in your directory when you press enter. It may take some time depending on the specifications of your computer system.We must execute npm start after installation. After pressing enter, a development server will start on port 3000 and display a default page.

Folder Structure of React app :

Use of React : We use react to avoid code repitition. We declare a components folder inside which we divide the entire html page using some components.
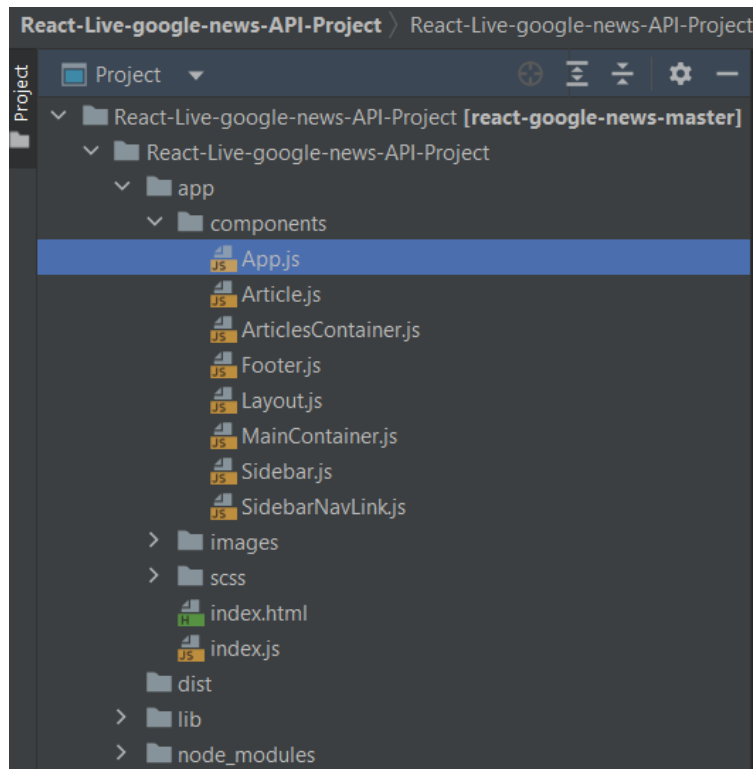
Figure 2: Folder Structure of the Live Google News API app

Only the 'app' and 'components' folders need to be dealt with here. and scss on occasion. 'app' contains all of the subfolders and files required for the application, such as pictures, components, and scss.

The 'components' folder contains the main files, the building blocks of the application like app.js, article.js, ArticlesContainer.js, SidebarNavLink.js.

App.js : App.js is the file that integrates and specifies all of the components. After that the app.js is also incorporated into the index js.

React is used to prevent repetition in code. We create a components folder and use various components to partition the full HTML page.

In index.js the id name root is derived from the index.html file in JavaScript. It is the single html file in the app and the primary file.

News Api: You'll need an API key to connect to the News API and get articles. You can receive one by going to the News API and clicking the "Get API key" button. Before you can acquire the key, you must first create an account.

News API also works on REST principles [15]. In order to obtain the information, one must generate a GET request on the article resource. The response fields used in the aggregator are the following ones:

1. Title - the title of the article;
2. Description - a short description of the article;
3. URL - a link to the article website;
4. URL to image - a link to an image related to the article;
5. Published at - date when the article was published.

Because certain fields in the API may be empty, the aggregator additionally employs Newspaper library and URL Meta API, like in the case of RSS feeds, to ensure that each field is appropriately shown to the user.

Front End

The application leverages AngularJS [17], a JavaScript-based framework, to implement client functions. W3CSS [13], an HTML, CSS, and Bootstrap framework, is used to create the graphical interface.

1. HTML – The HyperText Markup Language is the standard markup language for documents designed to be displayed in a web application. It's a method of defining the structure of a website. It may also be used to designate whether your content belongs in a paragraph, list, header, link, image, multimedia player, form, or any of the other existing components, or even a new element you create. To organise content a series of headings, paragraphs, lists, or graphics and tables are used.

2. CSS - Cascading Style Sheets (CSS) is a style sheet language that describes the styling of a document expressed in a markup language like HTML. Every web page would be boring plain text and graphics that flowed straight down the page if CSS wasn't used. You may adjust the layout of your website and add colour and background pictures with CSS, making your web sites feel like pieces of art! CSS is in charge of the webpage's overall design . Anyone may modify the colour, text, style, fonts, and spacing using CSS.

3. Bootstrap - It is a free and open-source tool for developing responsive web pages and online apps. It's well-known for being used to develop responsive, mobile, and desktop websites utilising HTML, CSS, and the JavaScript framework.. It resolves compatibility difficulties. Bootstrap allows us to create faster and easier websites.

4. AngularJS - It was chosen because it makes web application development easier. AngularJS apps have a straightforward layout that is ideal for both small and larger projects. This is the best option for any online site, particularly if it's a single-page application (SPA). AngularJS works nicely with any CSS framework, allowing users to develop fluid interfaces that can be accessible from both large and small screen devices. If a big number of articles must be presented all at once, the page load performance will be slow due to the SPA technique.

It presents less problems than cross-browser compatibility. It's a consistent framework with CSS-based compatibility patches that all browsers support. A grid system that is both simple and effective.
News aggregator: In computing, a news aggregator, also known as a feed aggregator, news reader, RSS reader, or simply an aggregator, is a client programme or web application that procures data/information such as online newspapers, blogs, podcasts, and video blogs (vlogs), and stores it all in one location for processing. By employing the notion of web scraping, an aggregator gathers together and allows a user to browse news in one location. We may take the data of practically any website or channel and present it anywhere we want, in any form we want.

RESULT

The resulting programme provides users an extensive range of options for implementation of the features discussed earlier. Material Design [14] which is a primitive visual language of Google is used for this online app, incorporated in its interface to optimise user experience (see Figure 3).
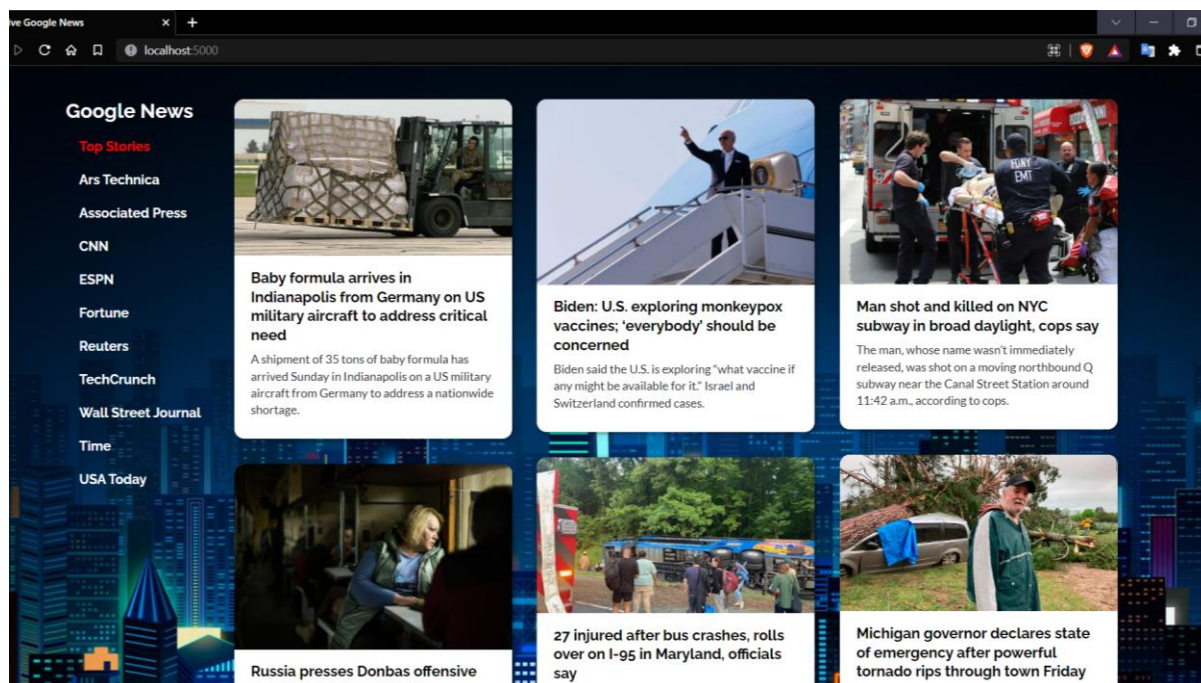

Figure 3: Preview of the Live News API Web Application in Desktop View

The news is represented in the form of cards, which are card elements defined in CFramework. Every news article is displayed with a reference image: if the image URL is successfully fetched, else a predefined image, an article title, and a brief summary explanation. The articles are organised chronologically, with three items in each row, from the most recent at the top to the oldest at the bottom. In addition, if the user clicks on the article title, he or she will be sent to the source page, where they may learn more about it.
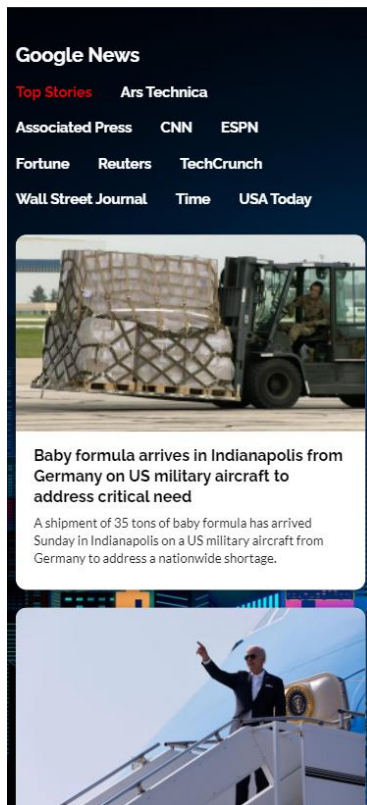
Figure 4: Preview of the Live News API Web Application in Mobile View
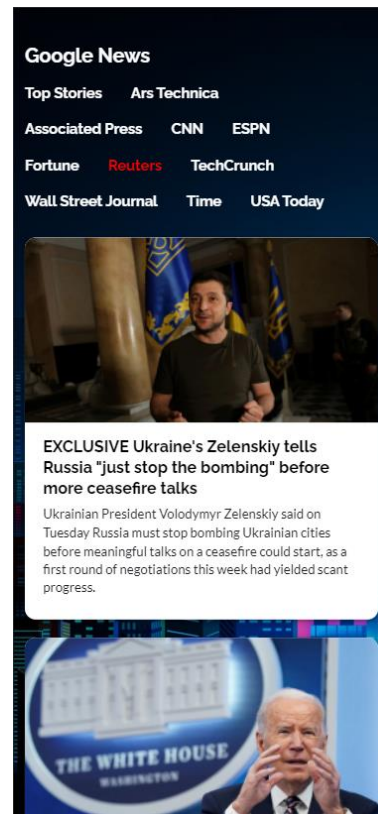

Figure 5: Preview of the Web Application with channel specific News

The sidebar, which are buttons on the left side of famous news websites such as CNN, Reuters, and others, may be used to access extra site features. When you click on those specific news channels, the news from those stations will be provided in the programme. When accessed on a mobile device, the website changes its orientation, making it responsive. The sidebar transforms into a top bar, with only one news item displayed in each row.(see Figure 4,5)

CONCLUSIONS AND FUTURE WORK

The acquired aggregator is a web application that indexes RSS and News API feeds from several news sites, and is the focus of this research. The software has elements that are unique to existing news aggregators, such as the ability to choose sources, organise information into a common, easy-to-read format, and access several news channels in one spot. At the same time, the programme addresses several flaws such as user privacy, excessive advertising, and spam. From the standpoint of application architecture, the aggregator is built with contemporary programming languages and technologies that allow for scalability and future evolution. It is possible that the application will be developed further.

1. Categorising news on the basis of their genres like sports, technology, business etc.
2. Users may follow the content of multiple channels directly in the aggregator interface thanks to cooperation with YouTube.
3. Integration with other services such as APIs providing currency exchange information, weather conditions, music charts, movies etc.

REFERENCES

01. Cosmin Grozea, Dumitru-Clementin Cercel, Cristian Onose, Stefan Trausan-Matu Cercel, 'Altas: News aggregation service' in 2017.
02. https://medium.com/@sayonislam.2000/making-a-news-web-application-using-react-js-54349bb3e1a2
03. https://www.boorje.com/react-news-app/
04. News API: https://newsapi.org/ [August 2017].
05. K. R. McKeown, R. Barzilay, D. Evans, V. Hatzivassiloglou, J. L. Klavans, A. Nenkova, et al., "Tracking and summarising news on a daily basis with Columbia's Newsblaster," in 2nd International Conference on Human Language Technology Research, Morgan Kaufmann Publishers Inc., pp. 280-285, 2002.
06. J. Steinberger, "MediaGist: A cross-lingual analyser of aggregated news and," in ACL, 2016.
07. F. Hamborg, N. Meuschke, C. Breitinger, and B. Gipp, "news-please: A Generic News Crawler and Extractor," in 15th International Symposium of Information Science, Berlin, 2017.
08. Feedly: https://feedly.com/ [August 2017].
09. Flipboard: https://flipboard.com/ [August 2017].
10. R. Steinberger, B. Pouliquen, and E. Van Der Goot, "An introduction to the europe media monitor family of applications," arXiv preprint arXiv:1309.5290, 2013.
11. F. Hamborg, N. Meuschke, and B. Gipp, "Matrix-based News Aggregation: Exploring Different News Perspectives," in Joint Conference on Digital Libraries (JCDL), pp. 1-10, 2017.
12. G. Paliouras, A. Mouzakitis, V. Moustakas, and C. Skouras, "PNS: A personalised news aggregator on the web," in Intelligent interactive systems in knowledge-based environments, Springer Berlin Heidelberg, pp. 175-197, 2008.
13. P. B. Darwin and P. Kozlowski, "AngularJS web application development," Packt Publ., 2013.
14. L. Richardson, M. Amundsen, and S. Ruby, "RESTful Web APIs: Services for a Changing World," O'Reilly Media, Inc., 2013.
15. W3CSS: https://www.w3schools.com/w3css/ [August 2017]
16. Material Design: https://material.io/ [August 2017].