

Copyright work details

Title: News Recommendation System Using Sparse Matrix Implementation

1. Preface

Online news has grown as a result of the development of the internet, which has made online media, the primary medium for delivering news and information from most agencies because news is critical for internet users to be up to date on the latest and most relevant news. Several news agencies are using social media to spread their online news to encourage people to read it.

The news will appear according to users' personalized news recommendations once they log in to the site. Nevertheless, the site can provide personalized recommendations to users if it has records of their browsing history, and they have logged in. It may be problematic if readers are anonymous, and the site needs to store more information about their browsing history. News agencies will find it challenging to increase their daily readership if they have to compete with other agencies for personalised recommendations. Consequently, an anonymous user visiting the news site occasionally requires a personal recommendation system so that the website can recommend the news they prefer to read by implementing CSR Sparse Matrix-Vector Multiplication and Proximity Processing as Sparse Matrix applied methods. We will also implement the Bayesian framework for user interest with a modification to its term. With limited knowledge of users' search histories, personalized news recommendations can still be provided to readers.

2. Objectives of the proposed work:

1. To provide users with the ability to engage, to increase traffic and reach of the news website among its competitors.
2. To provide a multi-functional news website by adding various features, which users can interact with.
3. To provide a personalized News Recommendation System Using Sparse Matrix Implementation.



A handwritten signature in blue ink, appearing to read 'S. Pruthi', located at the bottom right of the page.

3. Concept Deployed:

The project's name is News-O-Mania (News Recommendation System Using Sparse Matrix Implementation). It allows you to read the news quickly and efficiently. Essentially, it is a web application that includes various news categories. Users can browse through a variety of categories, such as General, Technology, Business, Entertainment, etc. In this application, data is retrieved from third-party APIs and then populated onto the front end of the application. Now, for creating the news application, we have divided the structure into three modules. The component Structure is shown in FIG 1.

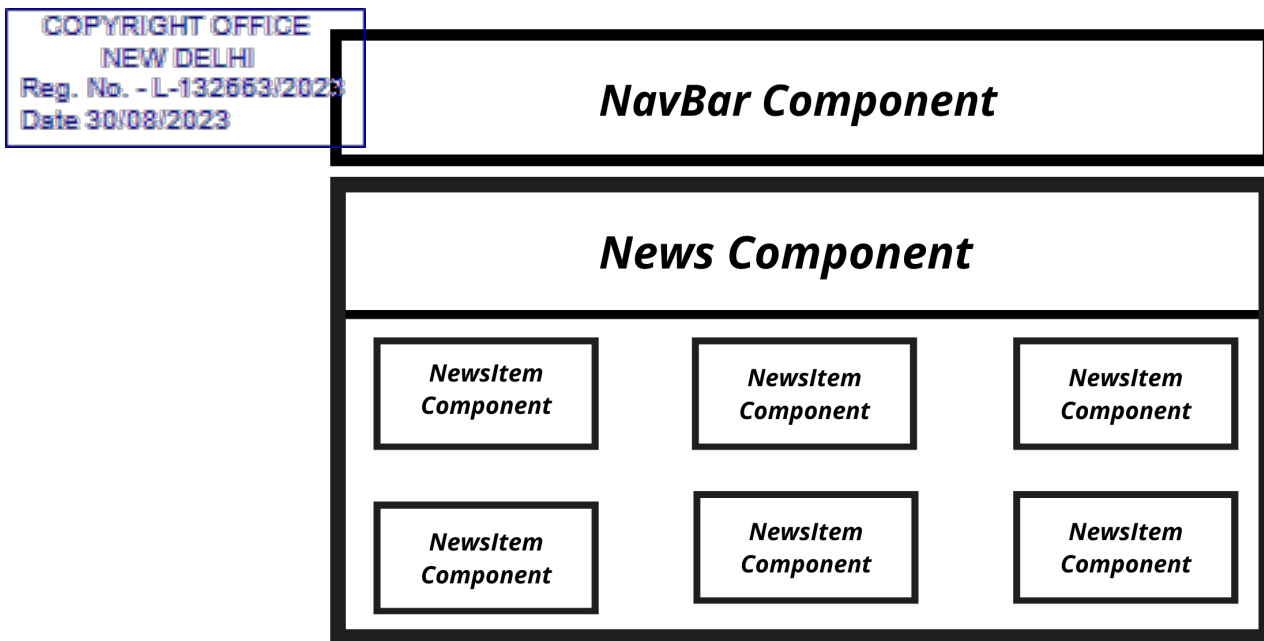


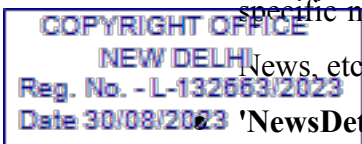
FIG 1: Component Structure of News-O-Mania



20/08/23

In the **First module**, we will design the component structure of the web application. It will be consisting of 'Navbar Component', 'News Component', 'NewsItem Component' & 'NewsDetail Component'.

- **'Navbar Component'** - This component is made for the navigation of the different categories of our application.
- **'News Component'** - It will be a big 'News Component', which will contain different 'NewsItem Components'.
- **'NewsItem Component'** - This is a component, which will be used for displaying specific news in the form of 'Cards' to users. Eg - Weather News, Politics News, Sports News, etc.



● **'NewsDetail Component'** - This component will redirect to the official source page of News. The Page, where the news has originated. It allows you to display news on a new page after clicking, NewsItem Component.

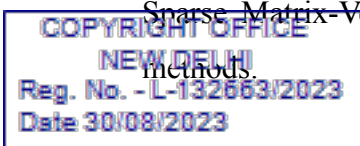
In the **Second module**, we will 'Fetch the API key from third party news API'. We will be using the News API, to render the news on our 'News-O-Mania' Application. A Unique API Key will be generated for every user who signed up on the third-party news API website.

In the **Final module**, earlier we planned to create 'Previous' & 'Next' buttons, to display the specified number of 'NewsItem' components on a single page, which would simply work as disabling the 'Previous' button if the page size is less than one and disabling the 'Next' button when the News API has produced all the news. But to enhance our application and make it modern looking, will be replacing the Previous and Next buttons with the 'Loading Spinner', and the Combination of 'Infinite Scroll' and displaying all the 'NewsItem' Components on a single page. Now, we will be also adding some unique and amazing features on every 'NewsItem' Components such as - Card Form like structure, Each card contains 'Author, Time, Date, Origin, Short Description & Image', Upvote and Dislike, Search & Filter, Sharing the news on different Social networks, Login/Sign Up in news application (Google or Facebook authorization), Save & Clean UI. And, At last, will be hiding the API key and doing the refactoring and on of code.



20 Rain

Now, we're ready with our application, by now it's time to add the 'Personalized News Recommendation' System to our application. As we already discussed, that is personalized recommendations could be provided to users if the site has records of users' browsing history, and users must have logged in to the news site. In the case of anonymous users, where the site does not have enough records of their browsing history, this could pose a problem. The number of daily readers on news agencies' sites will be difficult to increase, especially if they have to compete with other agencies for good personalised recommendations. As a result, anonymous users who only visit the news site occasionally need to receive a personalized recommendation system so that the website can recommend the news they prefer to read by implementing CSR Sparse Matrix-Vector Multiplication as well as proximity processing as sparse matrix applied methods.



The System Architecture of the proposed Methodology will now be outlined, as shown in FIG 2.

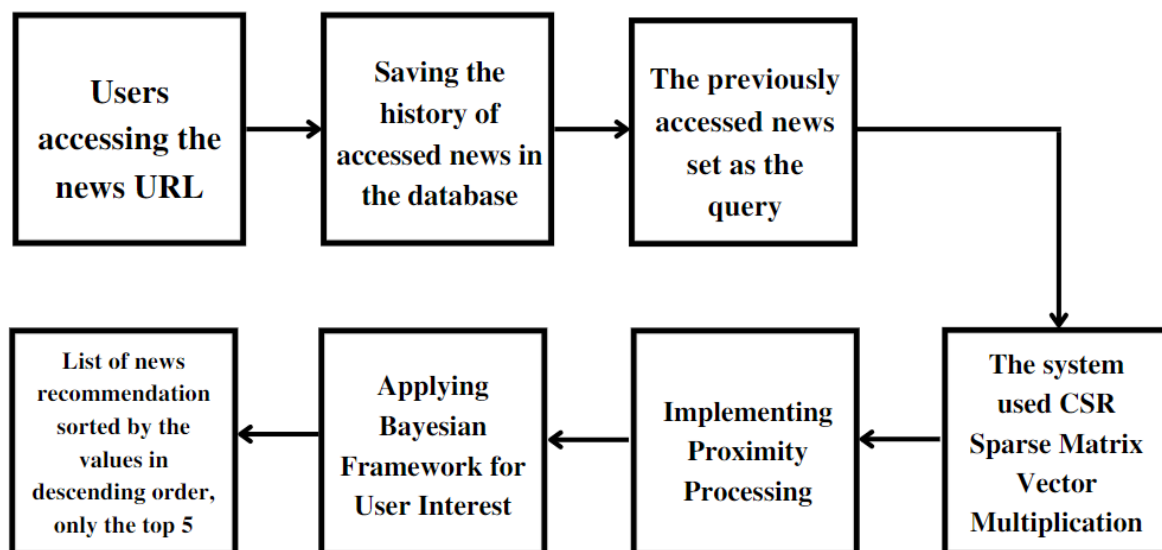


FIG 2: News-O-Mania System Architecture

Thus Recommending the News is divided into four phases - Elimination, CSR SpMV, Proximity, and Personalized News Recommendation.



20/08/23

In the **Elimination** phase, it aims to get rid of unnecessary words or characters. Therefore, primarily it focuses on, replacing characters other than (a-z, A-Z, 0-9 and spaces). Also, the process works as in the raw document, detects the character, then the process converts to lowercase, performing stemming, removes stop words and in the end, we get the terms from the document. The flowchart of the elimination phase is shown in FIG 3.

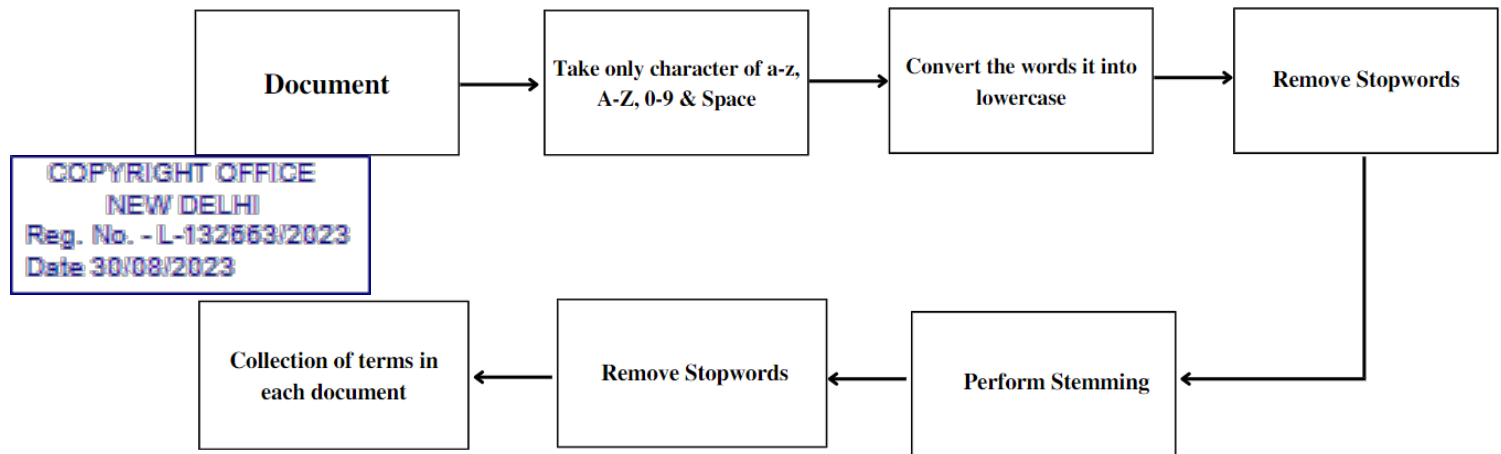


FIG 3: Elimination Phase Flowchart

In the **CSR SpMV** phase, here it stands for 'Compressed sparse row, Sparse matrix - Vector Multiplication'. It works as in the previous phase, when we get the terms from the document, and convert them to the term frequency (tf). Then, identify unique terms in each document to get the document frequency (df). Then, Calculate the inverse document frequency (idf). Finishing this phase by query formation and obtaining the relevant result of CSR SpMV.

The flowchart of CSR SpMV is shown in FIG 4.



20/08/23

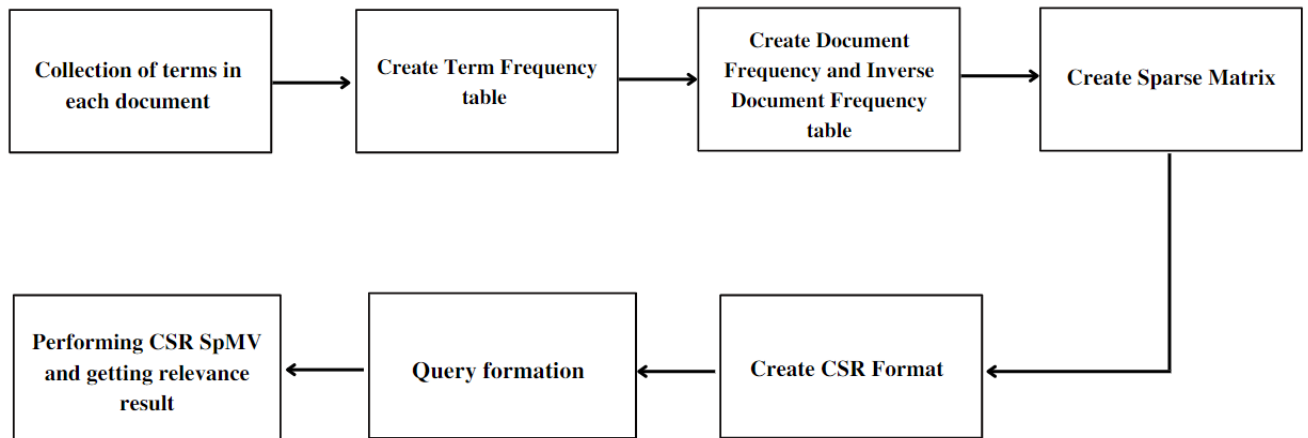


FIG 4: CSR SpMV flowchart

In the **Proximity Processing** phase, it is the continuation of CSR SpMV or we can say that it follows the same approach but with some modifications or additions in the certain steps. Adds the 'offset column' with 'term frequency'. Then, at the time of applying the CSR, it adds additional two functions 'offset_vector' and 'offset_marker'. And, then when performing 'query formation', it also provides two vectors 'v1' & 'v2'. So at last, we'll get the relevant result of 'proximity processing'. The main purpose of this phase is to record the position of each term in the document and do the reordering of sentences.

The proximity Processing flowchart is shown in FIG 5.



20/08/23

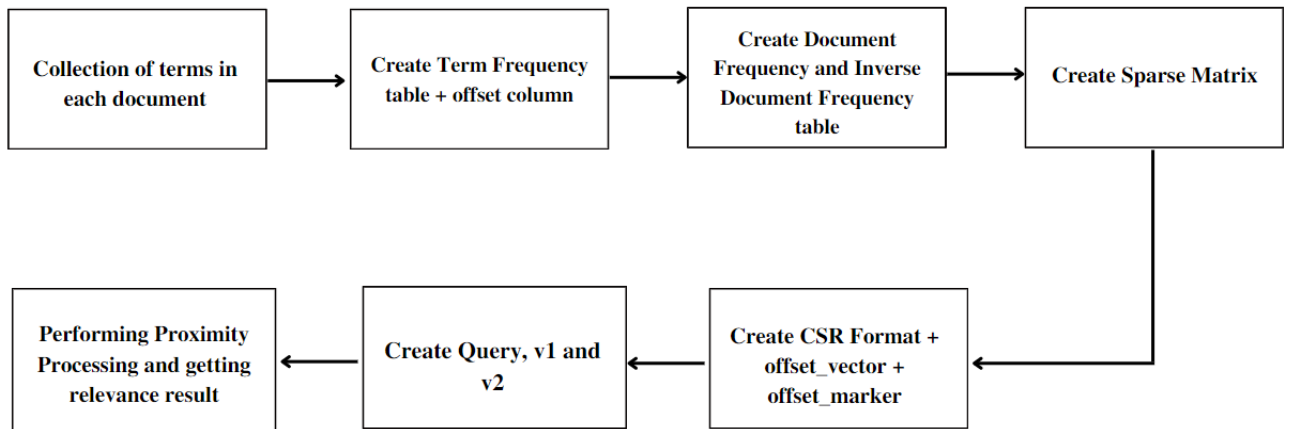


FIG 5: Proximity Processing Flowchart

COPYRIGHT OFFICE
NEW DELHI
Reg. No. - L-132663/2023
Date 30/08/2023

The **Personalized News Recommendation phase**, provides different news recommendations for every different newsreader, according to their preferences. It is developed from 'Naive Bayes' to predict each reader's news interest, called 'Bayesian Framework User Interest'.

$$p(\text{category} = c_i | \text{click}) = \frac{p^0(\text{category} = c_i) \times \left[\sum_t \left[N^t \times \frac{p^t(\text{category} = c_i | \text{click})}{p^t(\text{category} = c_i)} \right] \right]}{\sum_t N^t}$$

Here, it denotes:

$p^t(\text{category} = c_i | \text{click})$ --> is the probability of category access of i in t time.

$p^0(\text{category} = c_i)$ --> indicated the probability of last minute category access to total clicks in all categories.

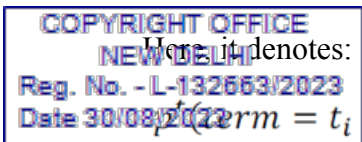
for the total click of t time.



20/08/23

As a result of this research, the category is being converted into terms, and click is being replaced by the term frequency to rate the news relevant to the user. This decision was made because the recommendation system used content-based recommendation that focuses on the details of each recommended document.

$$p(\text{term} = t_i | \text{freq}) = \frac{p^0(\text{term} = t_i) \times \left[\sum_t \left[N^t \times \frac{p^t(\text{term} = t_i | \text{freq})}{p^t(\text{term} = t_i)} \right] \right]}{\sum_t N^t}$$



Here, it denotes:

$p(\text{term} = t_i | \text{freq})$ --> is the term probability of freq in t time.

$p^0(\text{term} = t_i)$ --> is the term probability of the last access to total freq in all categories.

N^t --> is for the total freq of t time.

Now, It's time for the real-time testing of all the processes. The testing was conducted in two sections. The first section was to explain the recommendation result in detail using one anonymous user. The second one was to explain the quality of the recommendation system.

Let's take a test case scenario for a single anonymous user.

First, the anonymous user visited an article entitled 'How This Startup's Product Is Helping Restaurants Be More Efficient About Delivery,' and the article title will be set as the query. However, before it is set to be a query, there will be character removal using remove stopword i.e "transforming the query into startup product restaur more effici deliveries". Now, we will do the 'Consistency of Stemming Results', For example - the word 'restaurant' after the stemming process became 'restaur'. At this moment, the recommendation process will go through the CSRSnMV The result can be seen in FIG 6.



20/08/23

3 Reasons Product Managers Are the Secret to Unstoppable Startup Growth (6.38 = 6.38)
How Detailed Should Your Startup's Product Road Map Be? (Infographic) (6.38 = 6.38)
What Does It Take to Get Your Product on QVC? (3.72 = 3.72)
7 Must-Have Attributes of a Member of a Super Startup Team (2.66 = 2.66)
5 Ways to Manage an Outsourced Team on a Startup Business Budget (2.66 = 2.66)

COPYRIGHT OFFICE
NEW DELHI
Reg. No. - L-132663/2023
Date 30/08/2023

FIG 6: SpMV result

Now, the process continues to proximity processing. After the proximity, the recommendation results began to change in order. It can be seen the first place has changed where the top article recommendation of "3 Reasons Product Managers Are The Secret to Unstoppable Startup Growth" is being replaced by "How detailed should your startup's product roadmap be?"

The reason for this to happen is because the latter contains the word "startup product" sequentially following the query that is "startup product restaurant more efficient deliveries". The proximity result can be seen in FIG 7.



2023/08/30

How Detailed Should Your Startup's Product Road Map Be? (Infographic) (6.38 + 1 + = 7.38)
3 Reasons Product Managers Are the Secret to Unstoppable Startup Growth (6.38 + 0 + = 6.38)
What Does It Take to Get Your Product on QVC? (3.72 + 0 + = 3.72)
7 Must-Have Attributes of a Member of a Super Startup Team (2.66 + 0 + = 2.66)
5 Ways to Manage an Outsourced Team on a Startup Business Budget (2.66 + 0 + = 2.66)

FIG 7: Proximity Result

After proximity, the next step is the Bayesian process. For the first-time user, Bayesian access has not been enabled yet. It will work once the user accesses the next article. Bayesian Results can be seen in FIG 8.

How Detailed Should Your Startup's Product Road Map Be? (Infographic) (6.38 + 1 + = 7.38)
3 Reasons Product Managers Are the Secret to Unstoppable Startup Growth (6.38 + 0 + = 6.38)
What Does It Take to Get Your Product on QVC? (3.72 + 0 + = 3.72)
7 Must-Have Attributes of a Member of a Super Startup Team (2.66 + 0 + = 2.66)
5 Ways to Manage an Outsourced Team on a Startup Business Budget (2.66 + 0 + = 2.66)

FIG 8: Bayesian Result



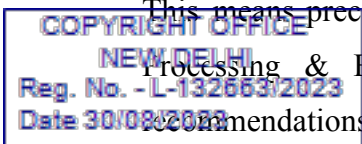
2023

As such, another section focused on improving the quality of recommendation system by displaying the news recommendation system for multiple anonymous users. It displays recommendations using some kinds of measurement, precision value and recall. Precision calculates the ability of the recommendation system to reject irrelevant documents. Recall calculates the ability of the system to find relevant documents. Ideally, a good recommendation system should have a precision value close to 100 and a recall value well above 0.

Precision values always show a perfect result at 100%. And, the recall doesn't give any 0 values.

This means precision, and recall with the combination of Elimination, CSR SpMV, Proximity Processing & Bayesian Framework for user interest can give the best result used in recommendations for multiple anonymous users.

At last, we have built a big news web application, with a hut of amazing features and the best-personalized news recommendation system for anonymous users.



Paarth Kothari, Atul Rane, Vansh Chitlangia, Dharmanshu Adlak
y: Mrs. Deepali Gohil

A handwritten signature in blue ink, appearing to read 'Paarth'.