Synopsis

On

News Recommendation Portal

Submitted for partial fulfillment of Award of

BACHELOR OF TECHNOLOGY IN (COMPUTER SCIENCE & ENGINEERING)

Subject Name: Minor Project Subject
Code: CSD0603

Submitted by

Tejeshwer Singh Sachdeva [BETN1CS19066]
Sunny Singh [BETN1CS19063]
Sankalp Pandey [BETN1CS19054]

Guided by

Neeraj Goyal (Assistant Professor)

School of Computer Science and Applications ITM University Gwalior, M. P.

Session Jan - June 2022

Table Of Content

Project Objective

The objective of this project is to develop a News Recommendation website which will provide daily news to the user. The news displayed on the portal to the user would be based on the interest selected by the user after registration on the portal, these interests can be then modified through the user page. The major focus of this project is on segregating the news from various news broadcasting platforms on one portal. But that's not it the portal will be designed using such methodology that the user's obtain a personalized news based on their preference, after the initial build the portal would also be modified to track user activity based on likes and bookmarks so that next time the user feed has improved news recommendation. The user choses their interest points and the news are then fetched and filtered out according to their interest areas.

The portal also would be made to accommodate voice search, and user activity tracker via machine learning algorithms.

Broad Area of Development/Research

A lot of work has been done in the last decade about news recommendations using different publishing platforms. To formalize news domain and approaches adopted, datasets used, challenges encountered, measures applied, etc. the study pursued and the literature is keenly evaluated. To quantify the objectives, quantitative analysis was conducted in the news recommendation domain and designed the following questions:
1) What are the different challenges facing by the news recommendations domain?
2) How to classify the news recommendations systems based on a specific domain?
3) Up to what extent the common recommendation approaches (CB, CF or Hybrid) are adopted?
4) How do authors provide information about the Efficiency of algorithms or techniques applied?
6) What is the e-news sources used for the news recommendation studies?

Integrating Voice Search

The news recommendation portal when fulfils the basic idea of rendering daily news articles based on user's preference will then focus on providing voice search functionality to the user's.

For implementing voice search the development team came across a software named "ALAN AI". Alan AI is a revolutionary software that enables us to add voice capability to any application. It allows users to control absolutely everything in the app using just their voice. This feature in our opinion would be very beneficial for user's with lower literacy rate.

Apart from that Alan AI is also easy to be integrated which makes it more easier for developer team to build the working version of the new feature.

Implementing Machine Learning Algorithms for final version

Once done with the above mentioned feature the next goal we'll take up is implementing machine learning algorithms, so that our news feeds could be much more organized based on the user's preference. We previously had an API that does the news fetching stuff for us, but let's say over a period of time we find out that user is more interested in one category specifically, so using our ML algorithm we'll try to display that particular topic more in the user's feed.

Now how do we plan to implement it? Here are a few points on how we plan to implement this feature:

- Finding readers with similar interest
- Topic modeling
- Making recommendations
- Evaluating our recommendations

Introduction

A lot of people like to view and analyse news from various news sources. At instance people are looking only for the pinnacle news tales in their categories of interest. Hence are more likely to subscribe to RSS feeds from different news sources. Therefore, the user shave to scan through all the top news stories in order to read stories of their likes and interests. Just like, a person interested in sports related top news category has to go through all the top news stories analysing news from different channels. We identified this need of bringing together news from different sources and categorizing them and presenting them to the users as a

single news feed. Many users tend to subscribe to RSS feeds of their interest in order to get updated with the latest news. However, many times this information is scattered across various news sources and spans more than one domain. Our system provides RSS feeds that presents all the news items from various news sources and groups them into categories with the main objectives such as:

1. Providing the User with personalized news they like by analysing the User's click behaviour.
2. Processing RSS feeds (representing news channels) and obtaining a single, well-categorized output feed.
3. Analysing and searching keywords entered by user to find a particular news article.

To make an application which responds quickly to users' action and preference and provide them better results. Its major concern is to save time and find and categorize top news as per user's interest. Our project aims at processing multiple RSS feeds (four news channels) and obtaining a single, well-categorized output feed.

Once the initial version is implemented we then move ahead, with implementing the voice control feature we mention in our area of development above. In this we try to make our app functional using Alan AI software which would allow user's to navigate through the app using just voice.

The reason this feature is trying to be embedded is because there are many user's who consume stuff by navigating on their devices using just voice commands, and in this era of IOT where things are controlled by gestures, voice, senses, etc, why not there be a portal that serves users quality journalism(the 4^{th} pillar of democracy) by just using their voice commands.

Literature Survey

This section gives an overview of existing technologies, their methodologies. We present a comparative study on different approaches.

People usually want to collect more information about a news. Gathering all these News helps users to aware of the current reality. Web blogs are full of un-indexed and unprocessed text that reflects the heterogeneity. It is not easy to walk through a lot of news and read it carefully. Sometime news is directly talked about the product and sometime reviews are explicitly mentioned. Thus, there is a need to collect and process different news sources so that it can be used in decision making processes.

News Recommendation System is basically a web application which requires user to log in or register. As the user logs into the application, he/she is required to select the category of news of his/her choice. After selecting the category, all live news of that category are fetched from different electronic news portal through RSS feeds. User can even manually search the news of his/her choice by typing the keyword. On the server-side, administrator uploads the news articles which are stored in the database. Thus, user can even watch the news when there is no internet connection. Hence, News Recommendation System is a true replacement to the old recommender system which used only data sets to recommend the news.

There are some topics that work under the umbrella of PNR and have attracted the researchers recently. In this subsection, few of these topics are presented in some details with related articles.

This work by Dr. M.Durairaj and K.MuthuKumarv was carried out in June2014. In this work,the first section discusses different News Recommendation System and its functionalities as well as the technologies involved in these systems. The second section discusses about extensively different topic analysis models and technologies involved to develop these models. The third section discusses the advantages and disadvantages of these systems as described by the respective authors. Finally, the paper is concluded with the suggestions and recommendations for building an effective news recommendation system based on the observations made from the extensive reviews and study. The observation made out from this literature study is that the news recommendation is challenging due to the rapid evolution of topic sand preferences. The comparison and analysis are carried out on different techniques and methods used for mining the news recommendation models.

Disadvantages-The proposed system does not have good efficiency in terms of user keywords and search results.

This work considers 3 kinds of recommender systems: collaborative filtering at the level of news items, content -based recommendation where we recommend items with similar topics to what was read, and a hybrid where collaborative filtering is applied at the level of topics.

Collaborative filtering recommends site to a user based on users with similar tastes, while content-based techniques create recommendations by analysing the content of the items. Collaborative recommendation compares reading histories in order to extract reading behaviour patterns. It recommends news items that other readers with similar reading histories have read. Readers are in different stages at a point of time, and news feeds are generated on basis of transition probability from one stage to another. In conclusion, it demonstrated that personalized recommendations using collaborative filtering can be useful even for individual newspaper sites with limited amounts of data about their users. Disadvantages- The content-based and hybrid recommendations have surprisingly poor performance.

In order to get real time updates of general Web news topics, interests and preferences of a user, a keyword knowledge base is maintained. The non-news content irrelevant to the news Web page is filtered out. Topics consist of keywords that are extracted using lexical chains which signifies semantic relations between words. Text summarization, collaborative filtering and recommendation of online news have received much attention in web and artificial intelligence, focusing on finding relevant and interesting news and also summarizing concise content. The Personalized News Filtering and Summarization system (PNFS) consists of two phases which are Personalized Web News Filtering and Web News Summarization. The purpose of keyword extraction is two- fold. First, it gives a concise form of the news to the user that saves the reading time. Second, the extracted keywords are also used to build a user interest model. In this work, it has presented the recommendation and summarization components of the personalized news filtering and summarization (PNFS) system. For their commendation component, it has designed a content-based news recommender that automatically obtains Word- Wide-Web (WWW) that is online news from the Google News portal and recommends news to users as per their preference. Disadvantages-The Google news in Personalized News Filtering and Summarization system (PNFS) provides general news avoiding missing important news.

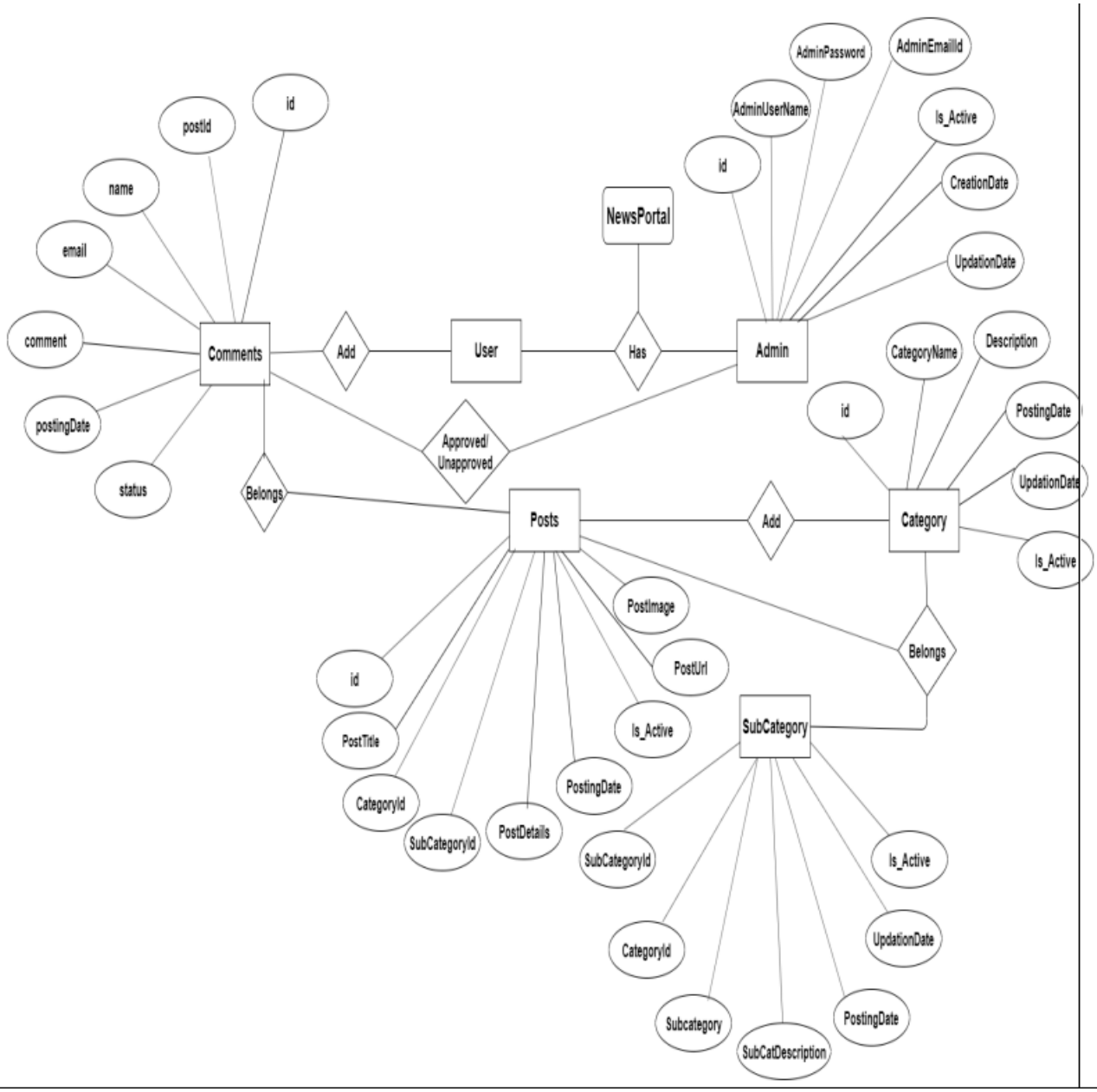
Scope / Limitations of The Project

It is often difficult for the public to quickly locate the news they are interested in.The personalized recommendation technology can dig out the user's interest points according to the user's behaviour habits, thereby recommending the news that may be of interest to the user. In this synopsis, improvements are made to the data pre-processing stage and the nearest neighbour collection stage of the collaborative filtering algorithm. In the data pre-processing stage, the user-item rating matrix is filled to alleviate its sparsity. The label factor and time factor are introduced to make the constructed user preference model have a better expression effect. Thereby improving the recommendation accuracy and recall rate of the recommendation and reducing the recommendation error.

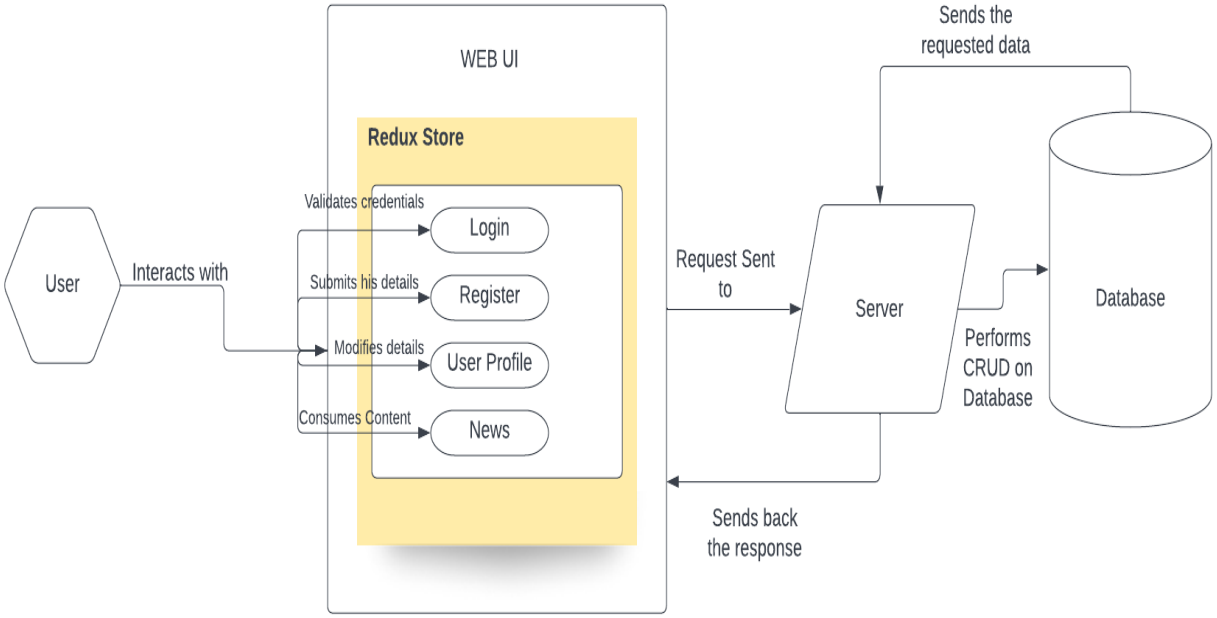
Requirements of The Project

- **ReactJS**
  React JS is front-end JavaScript library backed and managed by Facebook (now Meta). React is used to build Single-Page applications, it does so by dividing the whole web page in to different reusable UI components**.**
- **Redux**
  Redux is an open-source JavaScript library for managing and centralizing application state. It is most commonly used with Front-end libraries.
- **Node JS**
  Node JS is an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes.
- **Firebase**
  Firebase is a Google backed development software that enables developers to develop iOS, android, and web apps. It provides user SDK's (Software Development Kits) to track analytics, reports, and fix app crashes.
- **Material UI**
  Material-UI is simply a library that allows us to import and use different components to create a user interface in our React applications. This saves a significant amount of time since the developers do not need to write everything from scratch.
- **Styled Components**
  Styled-components is **a library built for React and React Native developers**. It allows you to use component-level styles in your applications. Styled-components leverage a mixture of JavaScript and CSS using a technique called CSS-in-JS.
- **React-Router-DOM**
  React Router DOM is an npm package that enables you to implement dynamic routing in a web app. It allows you to display pages and allow users to navigate them. It is a fully-featured client and server-side routing library for React.
- **Alan AI**
  Alan automatically trains on small existing datasets and generates models with phrases and intents for your app. As a result, you get an intelligent voice assistant that deeply understands the UI, workflow and business logic of your app.
- **Algorithms**
  To implement intelligent recommendations we make use of clustering algorithms to the user network present in our app, and then detect them based on similar topic preferences.

E-R Diagram


AdminPassword
AdminEmailId
AdminUserName
Is_Active
id
CreationDate
NewsPortal
UpdationDate
id
postId
name
email
comment
postingDate
status
Comments
Add
User
Has
Admin
CategoryName
Description
id
PostingDate
UpdationDate
Approved/
Unapproved
Belongs
Posts
Add
Category
Is_Active
PostImage
Belongs
PostUrl
id
PostTitle
Is_Active
SubCategory
CategoryId
PostingDate
SubCategoryId
PostDetails
SubCategoryId
Is_Active
CategoryId
UpdationDate
Subcategory
PostingDate
SubCatDescription

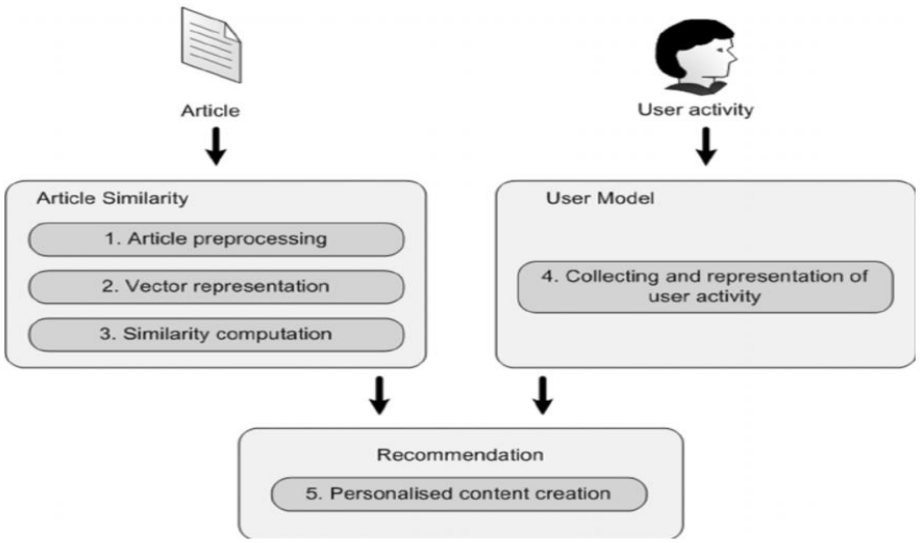
Data Flow Diagram



User:

- User can interact with the front-end of the portal.
- User can verify his credentials or create new account to get access to the portal
- User can choose topics of his choice from the provided template; news would then be rendered based on this information.
- User can then interact with the news rendered on the portal.
- User can change his details from the User Profile page.
- User can search news either in text format or voice format.

Proposed Model to achieve higer customer acquisition by implementing recommendation system algorithms, so that news is more relevant to the user.



Research / Development Methodology

News Recommender System (PNR) is a web application which requires users to login. The user logs in to the application and is required to select the category of news of personal choice. Hence after selecting the category, all the live news is fetched from different electronic news portal through RSS feeds relating to that category. Users can manually input search keywords to fetch news of his/her choice. To resolve the problems of existing work we propose a web application that will collect, parse, process, annotate and analyse different news from various RSS feeder channel which express either positive or negative information through crawling.

RSS Feeds: are used to fetch the news from heterogeneous news sites and portals which is stored in the database.

Web Application: The web application is the front-end interface for the User to access the system.

End Users: The Users that are accessing and using the system. Users can use the system as soon as they login.

Execution Steps of the portal:

Below mentioned is a brief workflow of the portal:

- **Getting User's Interest Points**

The initial step is to get user's interest area by allowing him/her to select from a variety of options, which will then be stored in the DB and used as and when needed.

- **Structuring the User's Data**

The User's data which in the initial prototype would include USER_NAME, USER_EMAIL, USER_PASSWORD, and the INTERESTED_TOPICS. In the next build the DB will be updated to focus on News Interaction ratio.

- **Fetching the API**

Once the User is logged in we then first fetch the interested categories from the DB, and then we make use of G-news API to fetch the data and render it strictly based on the User's categories.

- **Rendering the Components**

Now, once the data is available in the state of our React application, we then render the components built using ReactJS. The virtual DOM would then be rendered on the browser along with the data.

Build Step – 2 (Implementing Voice Functionality)

- **Integrating our app with Alan**
  The first step that the team would take in this build would be to integrate the Alan AI software with the currently existing build using the node package manager, after installing we provide the Alan API key and the root DOM node for it to get started
- **Specifying the commands that will trigger the voice controls**
  Once we successfully integrate our app with Alan AI we then specify the actions on our Virtual DOM that will trigger or enable the voice control feature. The user can then input commands using just their voice.
- **Acting on the inputs provided back**
  Then we send the users voice command to Alan AI SDK which then gives us a sequence of steps to follow to provide user his/her required results.

Build Step – 3 (Implementing ML algorithms)

- **Finding Readers with similar interests**
  Identifying the past user activity and using content attributes as input, provide recommendations to users. When a user "U1" is reading an article "A1", articles recommended are computed using article attributes of similarity, freshness and correlation and these recommendations will be article A1 specific and not user U1 specific.
  User specific recommendations can be built using knowledge based recommendations or collaborative filtering approach, which is not covered in this article.

- **Preparing the team and data pipeline infrastructure**
  Using the analytics captured from web and mobile applications, create a data pipeline for audience data and page-level events data using a scalable data streaming service. Anonymization/Tokenization is required to ensure that PII (person identifiable information) data is privacy compliant and secure for data controllers (news publishers).

- **Validating the approach and binding use case to a logic**
  Use case: As a user, I would like to see engaging & relevant recommendations when I

am reading an article on a news site, so that I will read more articles in the same session.

Content based recommendation logic: We will compute the following 3 metrics to identify an article suitable for recommendation:

Similarity (S), **Correlation** (C), **Freshness** (F)

Final score for an article is computed as a function of these 3 metrics i.e.

Recommended_Article_Score = f (S +C+ F)

Assumption:

weightages for each metric in the final equation can be need-specific i.e.

If users like to read articles that are read by others with similar preferences on the platform, then set higher value for “Correlation metric”. If users want to read articles of similar category or topics, then set higher value for “Similarity metric”. If users want to read “fresh & trending” articles then set higher value for “Freshness metric”

Team Structure and Role

Sankalp Pandey -

- Designing The initial Structure of the news portal using Figma as a software tool.
- Converting The Design Made Through Figma into a Static Working Version using HTML and CSS as the markup Languages.
- Communicating and resolving the Design conflict arising during the prototype build of the news portal model.

Sunny Singh -

- Converting the static working version built using HTML and CSS into react components.
- Resolving the style conflict arise due to common stylesheet using styled components and Material UI.
- Getting and updating the data through our portal using API provided by the backend teammate.

Tejeshwer Singh Sachdeva -

- Structuring out the documents based (NoSQL) database according to the user's fields.
- Connecting the Database and Front-End to perform CRUD Operations through APIs.
- Converted the Local Component state in the Front-End to a global state using React-Redux Framework.
- Implementing the Voice search feature by connecting machine learning algorithm provided by firebase SDK.

Expected Outcome

The final result of the news recommendation website will allow the reader to read the news based on his/her interest. Readers can now leave comments, bookmark the news and oftentimes contribute their own opinions about the news.
The site offers the information about the public, political, social, sports, health, entertainment etc. Instant & latest news from all over the world & Entertainment News from all over the world of your choice.
The final version if built during the provided deadline would also allow user to navigate throughout the web application using just their voice commands, which would make the app and the we in general more accessible to a wider audience.
The final version would also contain algorithms that would track user's activities throughout the app and would then model the activity data accordingly such that next time the user visits the website the data rendered would be more correct according to the user's choice and real interests.

Below are some mock-ups of the current progress mad in the project until now:

News Portal

Login

Email

yourself@gmail.com

Password

Password

Forgot Password?

Sign IN

Don't have an account yet? **Register here**

Reference / Bibliography

- **https://alan.app/docs/client-api/web/web-api/** => *For integration Alan AI voice command enabler software*
- **https://blog.insightdatascience.com/news4u-recommend-stories-based-on-collaborative-reader-behavior-9b049b6724c4** => *For implementing the ML model for our news recommendation*
- **https://medium.com/@pradeepmurthy112/building-a-content-based-recommender-system-for-a-news-website-ec90f02ccc76** => For setting up the pipeline methodology for our API data
- **https://beta.reactjs.org/** => *to implement the project in React JS library*
- **https://firebase.google.com/docs/** => *Setting up firebase and using it in our project*
- **https://react-redux.js.org/** => *Setting up redux in our project*
- **https://newsapi.org/** => *the actual API from where data is being fetched*