Personalized Buzzfeed

Submitted September 2021, in fulfillment of
the conditions for the award of the degree **MSc Advanced Computer Science.**

Sittukala Saravana Alagappan
199034381

Supervised by
Panneerselvam, John K. (Dr.) and Ulidowski, Irek (Dr.)

Department of Informatics
University of Leicester

I hereby declare that this dissertation is all my own work, except as indicated in the text:

Date : 20/08/2021

Word Count : 8538

Abstract

Personalized Buzzfeed is a simple user-friendly news feed desktop application with responsive designs which is developed for the users to login and read news around the globe. The application is designed considering three major features which includes Personalization, Visualization and Recommendation. Personalization is more important in any web application to keep the users around. But Visualization helps them to feel interactive and keep engaged. Additionally Recommendation helps users to understand their preferences much better and explore the application with the suggestions. In this project, Personalization is approached with few ideas from both Contextual and Behavioral aspects and Visualization is represented using amcharts chart and map projections with Recommendations using the Apache Mahout Recommendation Model which is a machine-learning framework. So, the application is designed to develop all these features at its possibly the best first version which still can refined and enhanced in the future to a business product.

Acknowledgements

I would like to thank my Supervisors Dr.John Panneerselvam and Dr.Ulidowski, Irek for providing guidance and support throught the project. The ideas and approach provided by Professor John was very helpful to develop the application according to the software development cycle. And the inputs from Professor Irek during the meeting gave a clear picture of the how the project report should come up and also to do some enhancements in the application. And a sincere thanks to Dr. John Drake who organized our master project conducting up-to-date guidance sessions and cleared our queries throughout the dissertation.

Also, I would like to extend my acknowledge to my parents and my sister who constantly supported to achieve my goals throughout the university.

DECLARATION

All sentences or passages quoted in this report, or computer code of any form whatsoever used and/or submitted at any stages, which are taken from other people's work have been specifically acknowledged by clear citation of the source, specifying author, work, date and page(s). Any part of my own written work, or software coding, which is substantially based upon other people's work, is duly accompanied by clear citation of the source, specifying author, work, date and page(s). I understand that failure to do this amount to plagiarism and will be considered grounds for failure in this module and the degree examination as a whole.

Name: Sittukala Saravanan

Date: 20/08/2021

Contents

List of Figures

Chapter 1

Introduction

Personalized Buzzfeed is a news aggregator that gathers news all around the world and present it to the users considering and analysing users personal interests making it more personalized application. This application enables the users to explore all around the global news with a click anywhere on the globe thus giving the users enhanced visualization look and feel.

1.1 Aims and Objectives

The main Aim of this project is to gather worldwide news and provide them to users across the countries knowing their interests and likes. The three key objectives of the application are observed as Personalization, Visualization and Recommendation.

1. **Personalization**

   Personalized Buzzfeed app provides the users, news and information of their personal interests and likes and also stays up-to-date on their interests by analysing their day-to-day behaviour on the application. So It stays unique to individuals preferences anytime.

   Definition for Personalization

   The adaptivity to individual users or user groups in a website is called personaliza-

tion.

Web Personalization is defined as the process of tailoring the content of a website to suit the user's needs and interests by making use of their behavioral activities in the application and their contextual details[8].

2. Visualization

Visualization is apparently achieved in this application providing a more appealing UI for users enhancing the user-friendliness. Visualization is provided in a more sophisticated way to the users in the form of more **interactive charts** and **geographical map** which is well achieved using amcharts library.

3. Recommendation

Recommendation is key feature which helps the users to handle increasing number of items or articles in case of news feed with the personalized suggestions provided by the application based on their preference patterns. According to research there are many recent development of recommendation techniques understanding its vital role. So, this application too integrates one of the Machine learning framework called Apache Mahout for recommendation system[13].

1.2 Motivation

- The main motivation in developing the personalized buzzfeed application is to gather various news data across the countries for its end users. The application is developed keeping in mind of two major category of stakeholders of this system. The **one** being the general casual news readers who are interested in various category news data. The **second** user-group being journalists, news companies or any organization who are interested in knowing popular category-wise locality data to better study the statistics of news by country or category.

- The other main motivation of this application is that to get a better understanding of

how web personalization is being automated and achieved instead of one-size-fits-all approach. **Personalization** is now the **key-role** in promoting any kind of websites and business to its end users and knowing this standard way of automation would benefit in the long-run learning and development. This system has implemented well-formed personalization effectively achieved through two dimensions of web-personalization.

Chapter 2

Background and Proposed system

2.1 Background study and Traditional applications

The traditional systems currently in the market are providing various news data like popular ones and breaking news for its readers. The increasing number of mobile phones contribute to large number of news readers in mobile phones and tablets. To maintain this growing population towards news-reading and to meet their expectations there is an excessive need of adaptive personalization to suit individual taste and interests which is limited in traditional systems.

Several applications in the marketplace were reviewed before giving a kick start for the project. For Example, **BBC News app** personalization works in a way letting the user to customize the ares of interests and populate '**My News**' page accordingly. It works more towards user-customization but lacks adaptability and presents outdated content if there aren't any recent news which makes it bit awful [9].

Others work in a way aggregating news from various sources which are knows as 'News Aggregators'. One such example is **Flipboard** which works in a way allowing users to pick their favorite stories from various sources or categories and populate most relevant content. Additionally it has Machine learning algorithms which learns from users interactions like follow, flip and heart and populate contents based on the high interactions[3].

And there is an observation made on **Google News** personalization which works consid-

ering two approaches.

- Taking users past click logs
- Building user profiles from personal preferences and interests [12].

Also, according to recent research, about **60 percent** of the applications follow hybrid recommendation system [7].The hybrid recommendation system refers to the efficient combination of content based filtering and collaborative filtering approaches.This enhances the accuracy of the recommendations quite well[10].

Missing features on existing applications

The applications in the marketplace shows some lag in few features which are considered in our proposed system.

1. **Lack of Visual Evidence of Users Interests**

   While such applications already exists in the market which learns from users context details like interested categories and interactions, there is a no explicit and user-friendly visualization of what the users favorite zone and order of interests and **numeric comparisons** of user interests. There is no clear picture of what the users interests are after a many interactions and the system starts to overload the content making it difficult for users to understand what their major interests are over a period of time.

2. **Overloaded feeds**

   The recommendations and the personal stories are overloaded in the apps sometimes making it a noise. This in-turn affects personalization and leads to criticism if the algorithm is not considering best relationship factors between the users or the relevancy score.

3. **Minimal or No Scope on Visualization**

   The present news app provides more data with but less ideal for users who look for more friendly and appealing visual tools.

2.2 Proposed System

The proposed system works in a way fulfilling three major targets of the application.

1. Personalization
2. Visualization
3. Recommendation

This all-in-one system helps enhancing the news app one step further making it more user-friendly and easy to use.

1. Personalization

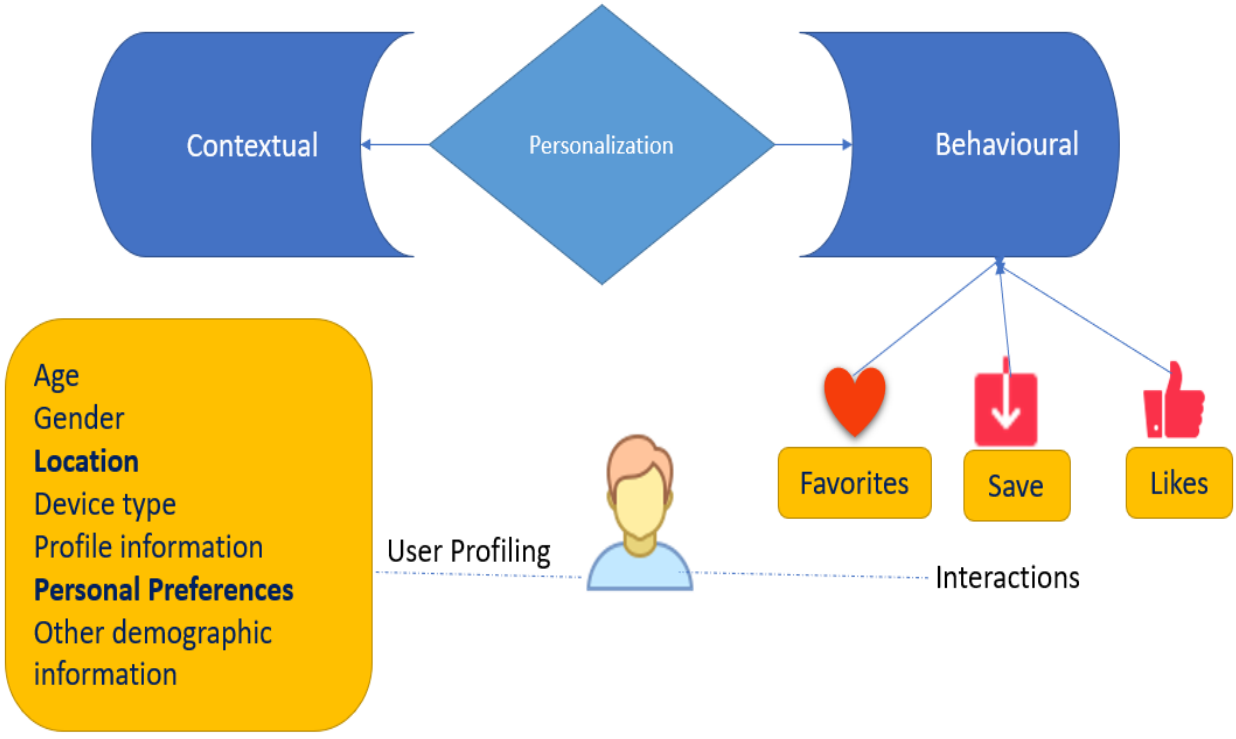

Figure 2.1: Personalization

Personalization is implemented in this Buzzfeed application with two possible approaches as shown in the figure.

(a) Contextual Personalization

Contextual personalization is the most widely adopted model where the user profiling is considered for customization and recommendation. The user profile is gathered and analysed thoroughly thus understanding individual selected preferences and several other factors expressed during the signup process for creating his/her profile.

It generally comprises several factors including:

- Age
- **Location**
- **Selected Preferences**
- Gender information
- Demographic information [4]. Personal Buzzfeed considered two of the above Location and Selected Preferences and works with the personal feed.

(b) Behavioral Personalization

Behavioral Personalization is measuring user's every interaction and moves in the application to predict their interests and personalize their feed with most likely items. This is considered more important as it determines the users trending mindset and gets updated with their moods to adapt to their current preferences. This is assessed in our application considering three major interactions for loading their personal feed.

- **Likes** with a thumbs-up icon
- **Favorites** with a heart icon
- **Save** with a download icon

How It All Works

Weight-Map Algorithm

For Personalization, the application works out an algorithm called '**Weight-Map**' Algorithm which takes all combination of Selected preferences from Contextual and other interaction based preferences and sorts the personal feed to present the users

with top likes news articles.

User Location

User location is considered separately and loaded initially with the Global data component and list down the news of his/her location with option to switch to other countries and any categories.

2. Visualization

Visualization is implemented using amcharts which is a library that was integrated into the frontend for accomplishing geographical maps and pie charts.

Why Amcharts?

Amcharts is chosen for integration as it has good projection of advanced charts and also supports plugins for maps with well-defined projections and compatible with the existing front-end technologies design and development [2].

Where in Buzzfeed?

This is implemented in two primary areas in our application.

- **Map View** The amcharts projects globe with all country geodata and initially it loads the user's location data on getting to this component in the frontend with possible option to see any kind of category the user likes to know in this location. The globe is clickable and allows to switch to any country and loads respective news based on selected category.
- **User charts** The amcharts makes it possible to project any kind of data having advanced chart types. Each user interests are mapped with respective preference weights associated for various category items in the news feed an they are represented in terms of pie-chrts to let users know explicitly what their interests and zones are.

3. Recommendation

Recommendation is another area where the users interests are analysed and compared with other similar users and suggested some categories which they might be

interested in. They are a kind of Machine learning algorithms which learns the users and provide relevant and optimal recommendations [15]. This is accomplished using **Apache Mahout Recommendation System**.

There are many types of recommendations possible like popularity-based recommendations, Content-based recommendations, Collaborative recommendation systems,etc

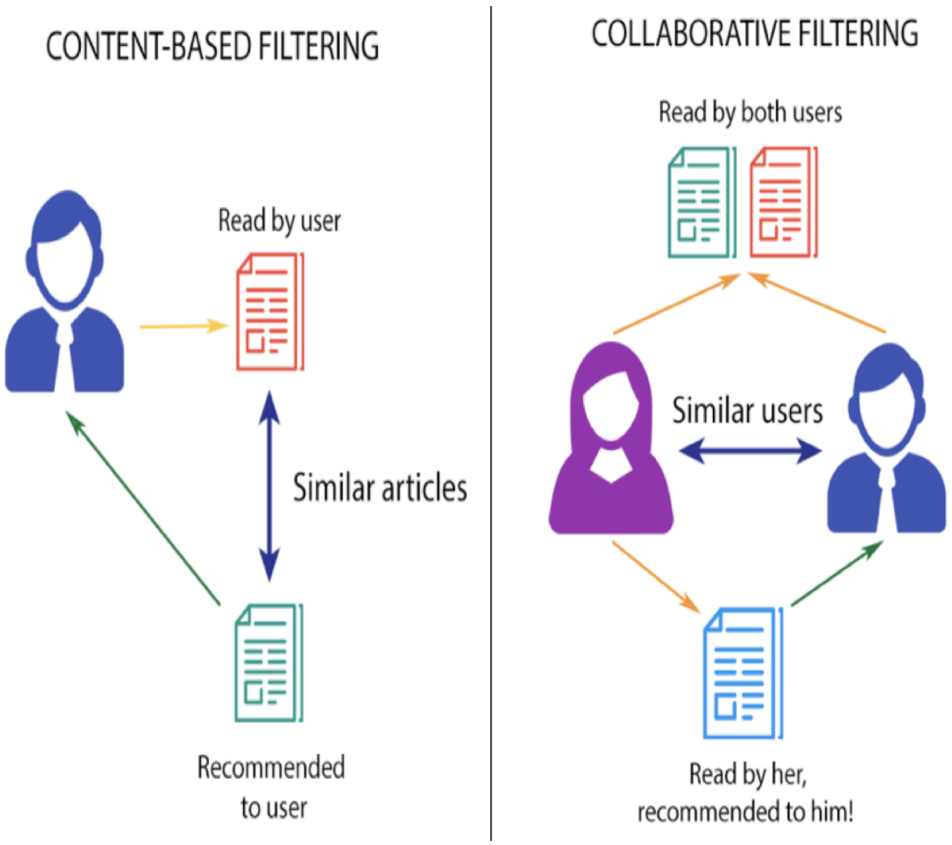

Figure 2.2: Recommendation Types

(a) **Content-based Recommendations:** Content-based Recommender system recommends the things or similar items which are most liked by the user. It observes the users like patterns and search history and recommends similar items to the users.

(b) **Collaborative Filtering:** Collaborative filtering is the highly beneficial and effective methods of recommendation system being implemented and provided

high success rates in many companies. For example, Amazon recommendation system which suggests on selecting a product, a pair of other products being bought together or similar products bought by other users which has increased sales by 29%. Other similar success use cases with Collaborative filtering are Netflix movie rentals by 60% and Google news click-through rates increased by 30.9% [11].

Reason for Opting Collaborative Filtering

So, similar seeing increase in success rates on recommendations with collaborative filtering, this application too incorporates the Collaborative recommendation system comparing similar interests of users and recommend news articles.

Why Apache Mahout?

Apache Mahout is a popular Apache-licensed open-source library for scalable machine learning algorithms [19]. Mahout provides most useful frameworks for Collaborative Filtering implementation. It has in-built framework, functionalities and APIs for machine learning algorithms to achieve the appropriate recommendations to users.

Chapter 3

Requirements Analysis

3.1 Top Level Requirements:

1. **Gather News Feed:** The application must gather news articles around the world for the users. This is planned to be accomplished through querying News Api which has categories and country fields and updating database with sample data to make it reliable for application.

2. **Personalization:** The application is mainly focused on delivering news based on user personalization aspects. For achieving this requirement, the application targets two main concepts[4] which are Contextual and Behavioral Personalization. Application is expected to develop Contextual personalization by building location-specific news and user-preferred categories news data in their personal feed. Also, in terms of behavioral the application is targeted to achieve personalization based on user interactions which includes the likes, save or add to favorites. Additionally, the application has to develop a way to recommend news articles to the users for which Mahout Recommendation system is considered for analysis.

3. **Visualization:** The application is aimed to have more interactive views for the user where the amcharts map for the location filter and charts for projecting user category-specific interest rate are desired to be achieved.

3.2 Detailed Requirements

Essential Requirements

1. **User Authentication:** The user registration should be feasible and the registered users should be able to login with a valid token.

2. **News Feed:** The application has to gather news from a well-defined api which has all essential fields of our news model.

3. **Database:** The news api has to be queried and updated to database with sample data for reliability of application especially during presentation to avoid any issues.

4. **Personalization:** The main feature of personalization to be well-designed and also updating for changing user's interests.

Recommended Requirements

1. **User Charts:** The application should be designed to project user category chart which shows percentage of interests for various news categories.

2. **User Profile:** There should have options for users to view their profile anytime in the application view.

3. **Visualization:** The application has to have a globe view to help users to choose any country to view the location-specific data and filter through categories if applicable.

Optional Requirements:

1. **Category Newsfeed:** The application has to load separate category data into various view for easy access.

2. **Saved/ Favorite feed:** The application could have the option to show users saved and favorite feeds for them to revisit anytime.

3. **Update Profile:** The option to update the profile details for the users could be better.

4. **Recommendation System:** The application is desired to have the recommendation of news articles for users based on their interest pattern.

Chapter 4

Technical Specification

Specifications	Software	Version	Description
Back-end Framework	Java	8	Used for efficient back-end.
	Spring boot	2.0	Provides supporting framework with in-built dependencies.
Front-end Framework	Angular	12.0	Front-end framework which aids component-based application design.
	HTML	5	Used to build individual components.
	CSS	4	Better styling.
	Bootstrap	4.5	Provides built-in designs and development
	JavaScript	ES6	Validation support
Query Language	SQL	Language support	Performing queries from database
Database Application	H2 - RDBMS	1.4	To store the application data and ability for having either in-memory or remote accessible database through url.
Repository	JPA	2.0	Access data from database and perform CRUD operations on the data
Web Services	Rest APIs		To post Http requests and receive http responses for transferring data.
Testing	Postman	5.5	To test API end-point.
Version Control	SVN Repository		To manage the code and documents during the entire project without any conflicts.

Chapter 5

Design

The application comprises the system design, architecture adopted, approach followed and the database used with a brief justification on why and how they are incorporated.

5.1 Approach

The Approach followed for developing this application from scratch is a step-by-step practise which are as follows:

- Wireframe Diagrams
- H2 Database Installation
- Initial Springboot setup
- Installed Angular modules and implemented design
- Agile approach

Let's see how it evolved one-by-one and the reason for adopting the model.

1. **Wireframe Diagrams**

   Wireframe diagrams act as a key step to begin any web development and get consent from the clients regarding the requirements on how it might look on screen. So it is considered more vital for any development process. Wireframes provides the layout

for the final product and it acts as a rough sketch for the application[17]. So, I have completed the requirements analysis phase and then designed the wireframe diagrams for getting approval from the Professor for proceeding with the initial designs and ideas which further evolved on development.

So, here are some of the screenshots of wireframe diagrams in the intial phase of design.

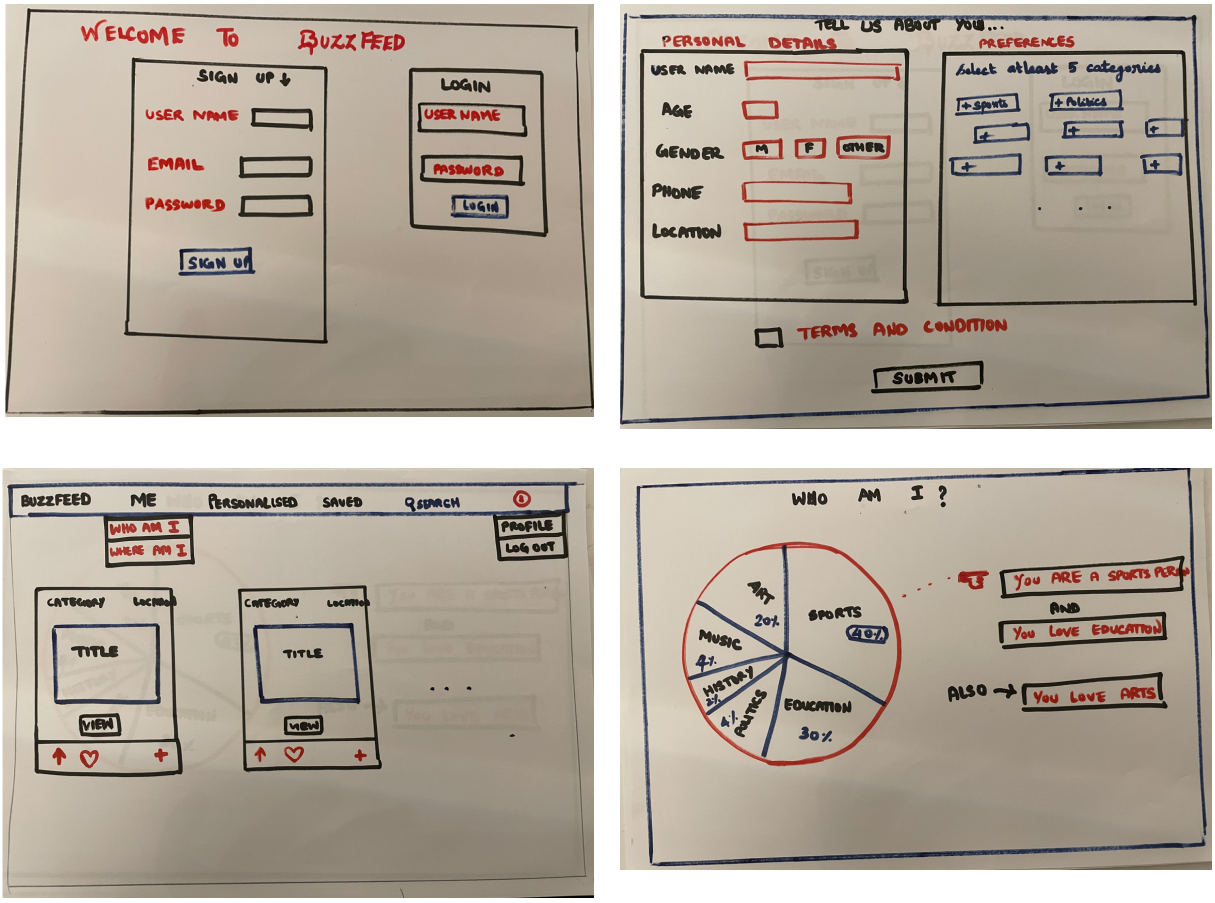

Figure 5.1: Wireframe diagrams

2. **H2 Database**

H2 is an open-source database written in java and hence it makes a very good combination with java springboot.

Features of H2 Database

H2 Database has the features compared against other databases which makes it

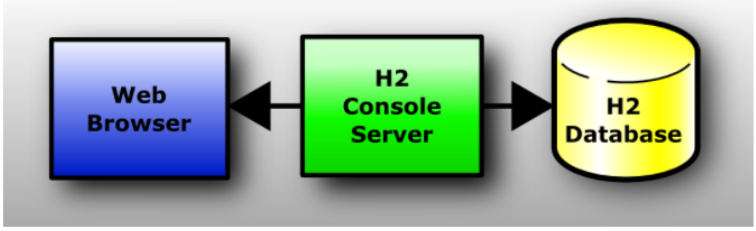

Figure 5.2: H2 Database

more preferable to use. The major reasons H2 database is used in our application

Features

	H2	Derby	HSQLDB	MySQL	PostgreSQL
Pure Java	Yes	Yes	Yes	No	No
Memory Mode	Yes	Yes	Yes	No	No
Encrypted Database	Yes	Yes	Yes	No	No
ODBC Driver	Yes	No	No	Yes	Yes
Fulltext Search	Yes	No	No	Yes	Yes
Multi Version Concurrency	Yes	No	Yes	Yes	Yes
Footprint (embedded)	~2 MB	~3 MB	~1.5 MB	—	—
Footprint (client)	~500 KB	~600 KB	~1.5 MB	~1 MB	~700 KB

Figure 5.3: H2 Features

are:

- **Java**: It is a Java open source database compatible with Springboot.
- **Embedded Persistent Database:** It supports several modes like embedded, in-memory and in both the modes it has persistent and in-memory databases. So, if we don't require to persist any changes to the database we go for in-memory or otherwise embedded database. This application works with Em-

bedded Persistent database as there might be requirement for users to anytime update their details in their profile.

- **Encrypted security:** It has Encrypted security which makes the data secured.
- **H2 Console:** Additionally it has an option of H2 console which provides browser based access to the database.

3. **Springboot Setup**

   Springboot provides all the java application framework ready to start the development process. It comes with all the packages as separate dependencies which can be added anytime into the pom.xml which inturn builds the Maven and makes the dependencies available to use.

   Once the project is created with Java springboot the next step is to create Controller, Repositories and the Service class which makes the application run with respect to the Multi-layered Architecture.

4. **Angular modules**

   Angular npm and node.js are installed and the project is started from scratch angular CLI command. Angular is chosen because it supports code reusability, framework that supports component based development and supports many enhanced integration for appealing front-end UI components that can be used.

   Once the angular modules are installed the designs are implemented having wireframe diagrams as a reference.

5. **Agile approach**

   The application is developed following agile standards which works very well with short-term standard software development. It is a results-focused development process which works adapting to the rapidly changing requirements.

There are many agile methodologies in software development process out of which the following are adopted in our application development cycle.

- **Working product:** The sprints are planned weekly and anytime the application is delivered or made available with a minimum viable functionalities ready to deliver.
- **Client Collaborations:** Meetings happened with the Professor and developer throughout the development.
- **Responding to Change:** The changing or updated requirements from Professors are considered and developed regularly to develop a complete product.

5.2 Spring Boot Architecture

Springboot works in a layered architecture consisting of four different layers communicating directly above or below them. This can be run easily with the embedded tomcat server.

- **Presentation layer:** This is the front-end view where the application presents the features to the users and handles the http requests transferring to the business layer.
- **Business layer:** This layer performs the business logic and handles the authentication and validation of application.
- **Data Access layer:** This layer plays a key role in handling data to and from the databases which has a repository to do so.
- **Database layer:** This is where the application data resides and all the operations like CRUD are performed[20].

Architecture Flow:

A Springboot application has a controller which process the http requests received from the client and it then interacts with the service layer where it does all the business logic for the application and data access or update using the repositories of the model [18].

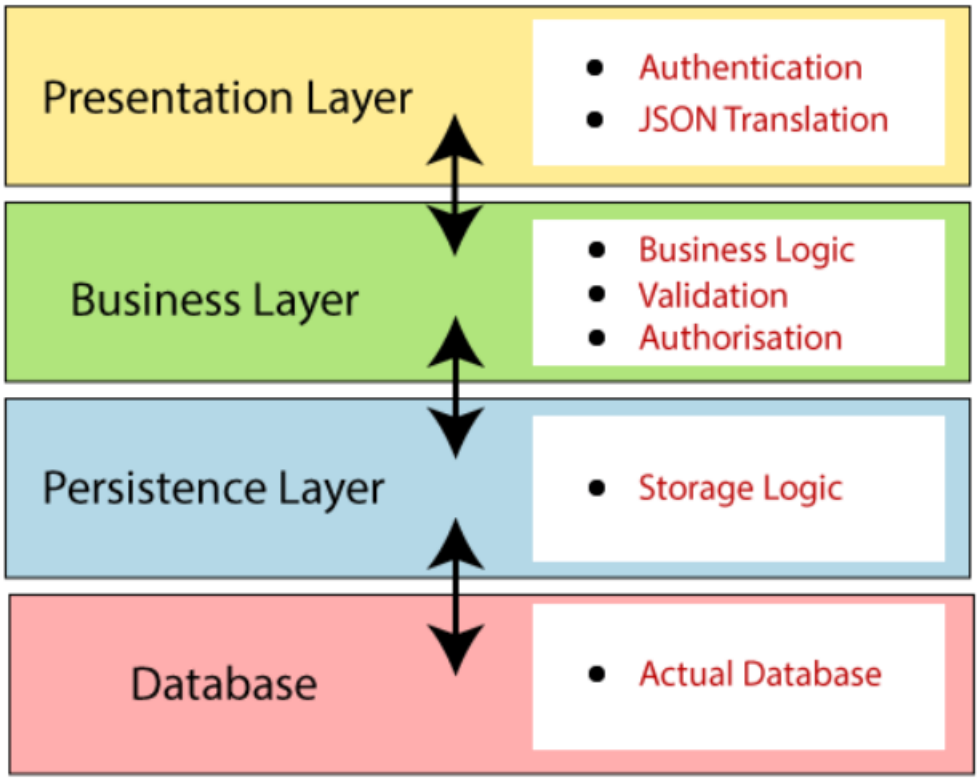

Figure 5.4: Springboot Architecture

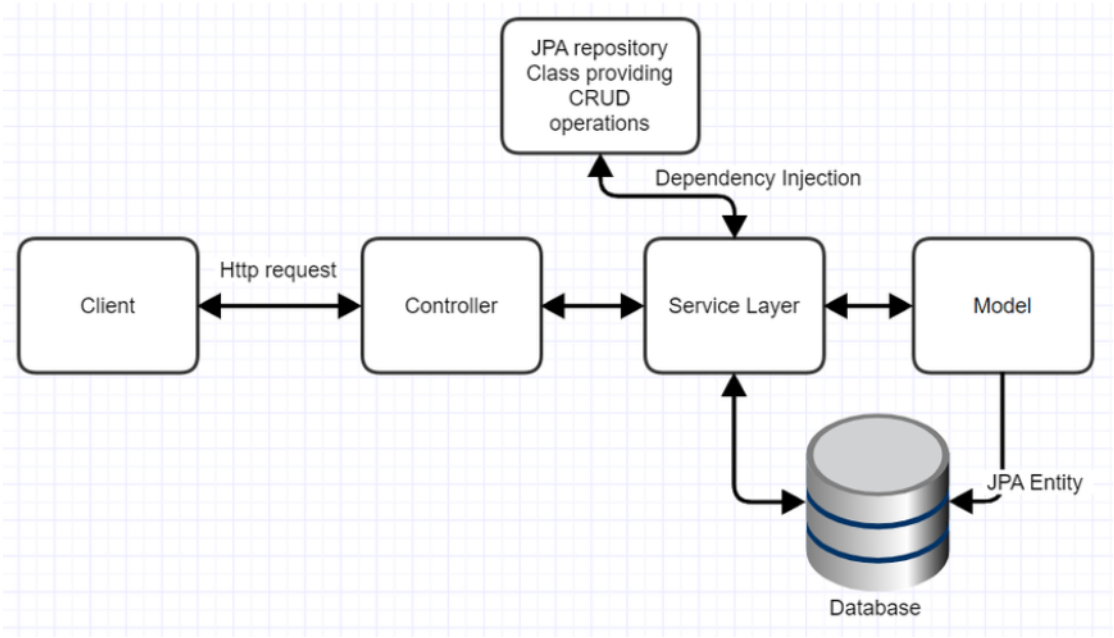

Figure 5.5: Architecture Flow

How It works in our Buzzfeed: Sample architecture flow in our application is represented as follows:

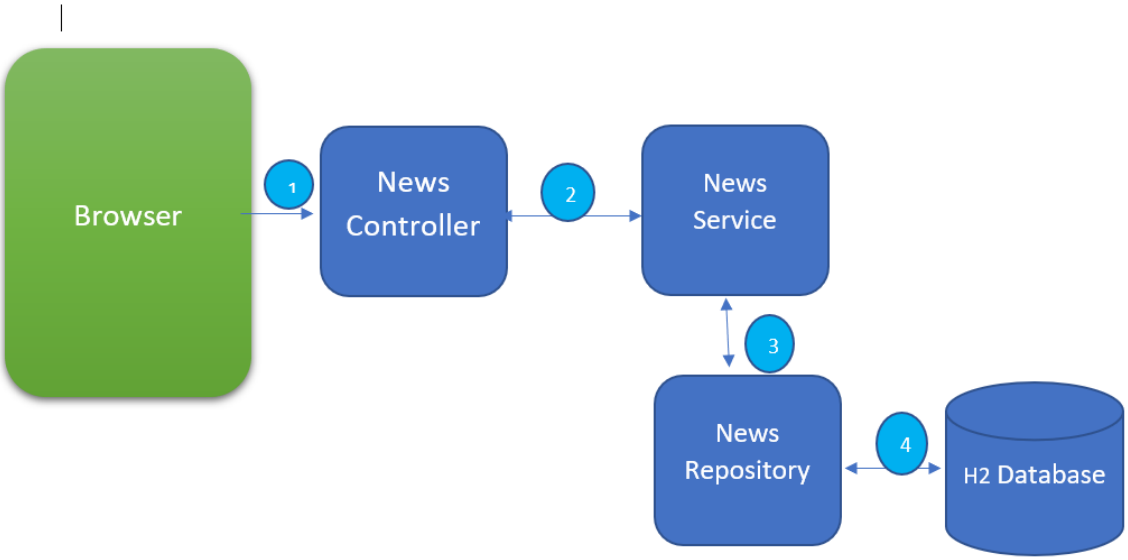

Figure 5.6: NewsFeed Server Flow

1. Springboot News Controller receives the http requests from the frontend.

2. The News Controller then process the requests to the News Service class which handles all the business logic and actions.

3. The News Service class then process the data from the database or updates the data to the database through the database access layer called New Repository.

4. The H2 Database stores the data and handles the data storage and updates if any. It is embedded Persistent database in our application so it can effectively persist data in case of changes and it is immediately available.

5.3 Object Oriented Design

5.3.1 Use Case Diagram

Use Case diagrams is the visual representation of system requirements. It gives a glimpse of what the system is intended to do to the stakeholders and the users of the system.

Here the main actor is the user in various forms and an external Mahout recommendation system which interacts with the system. The below diagram gives an idea of what the system does in brief.

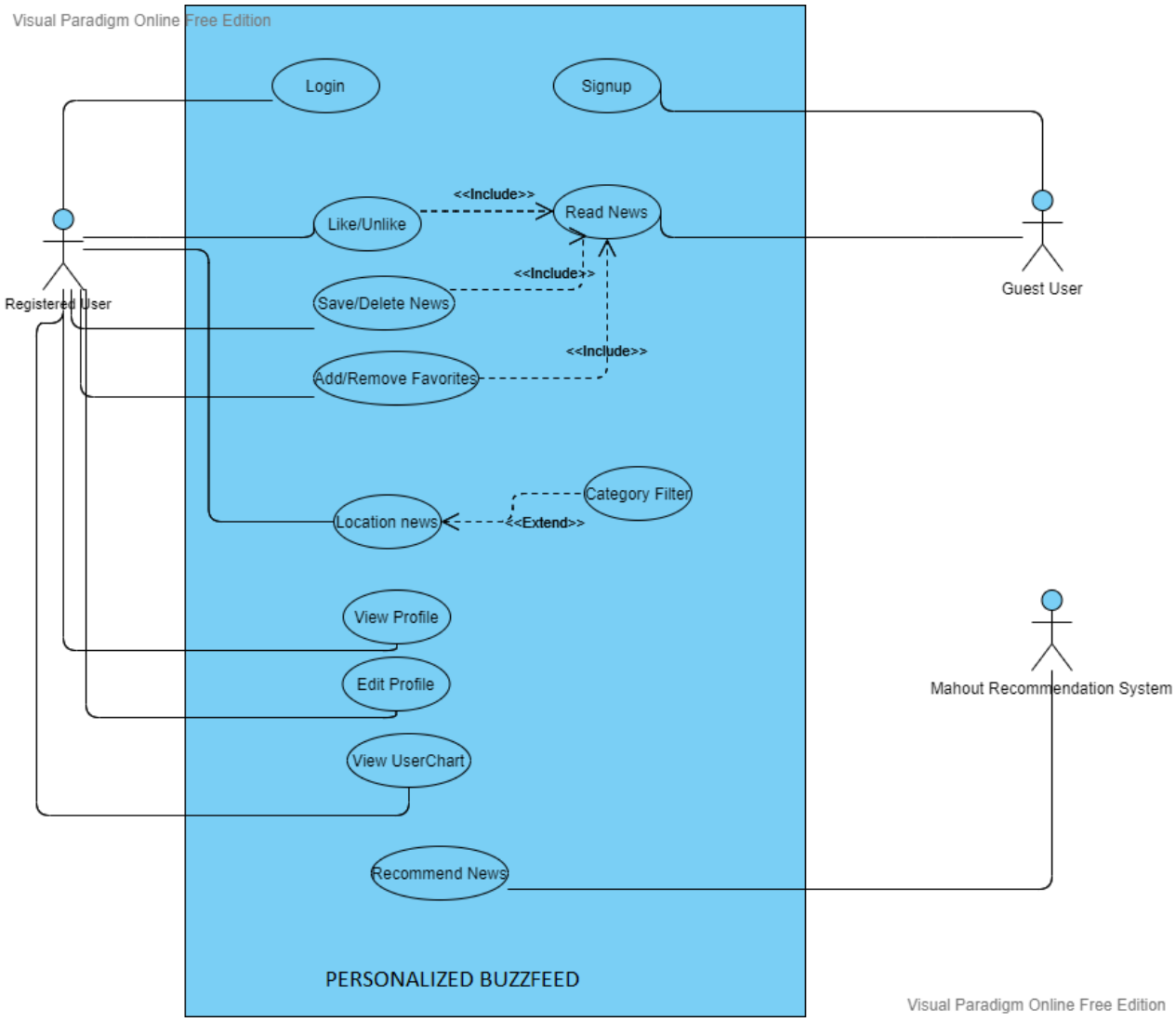

Figure 5.7: Use Case Diagram

5.3.2 Class Diagram

Below Class diagram represents the static view of the application and it has attributes and operations which represents the structural and behavioral features of the application.

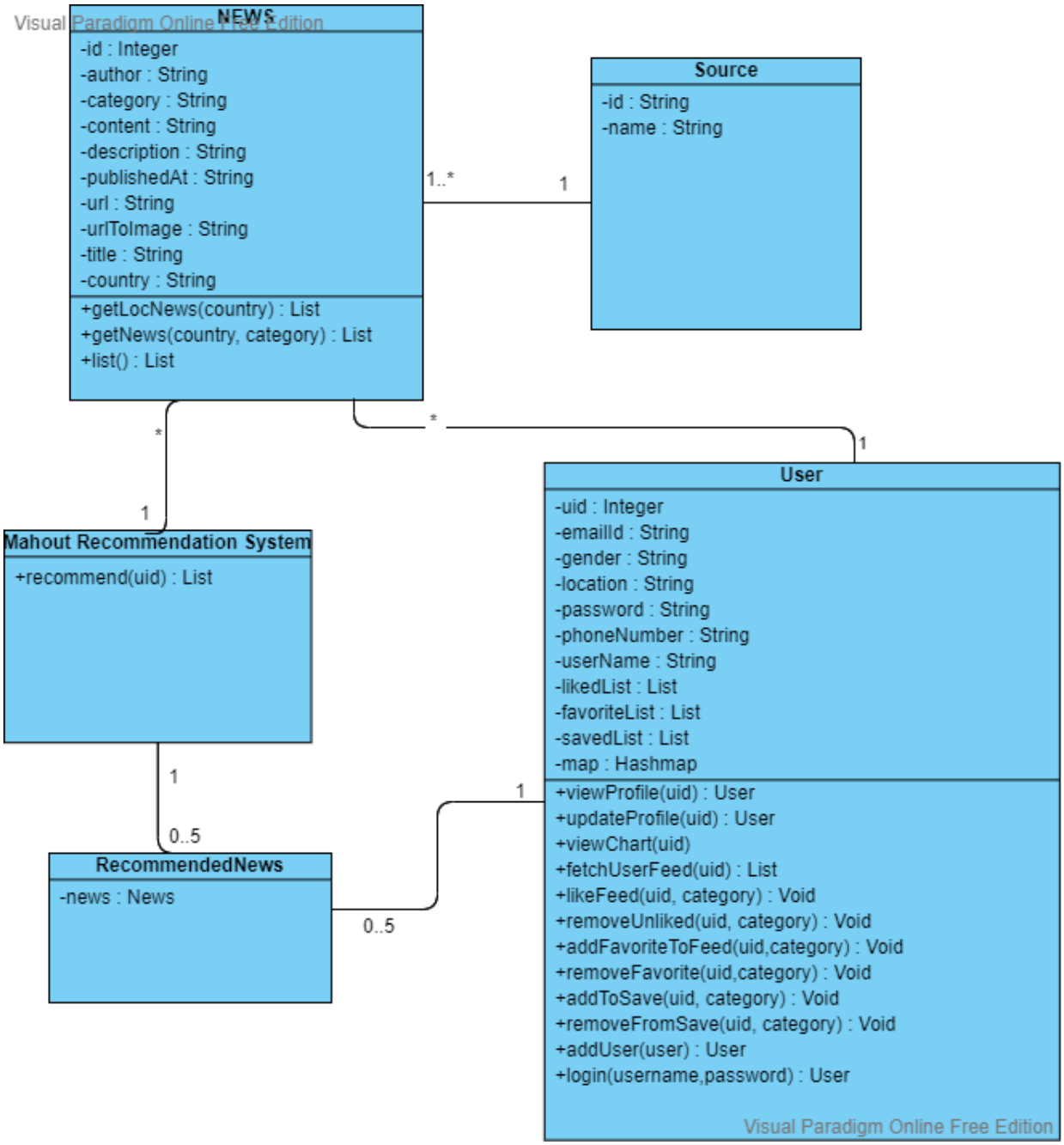

Figure 5.8: Class Diagram

5.4 Database Design

5.5 Application Design and Workflow

5.5.1 News Dataset

The first and the foremost activity for this application is to gather the right dataset that is required to accomplish the product. There are many datasets and third party service

APIs that provides the news data that are required but there were certain limitations in the necessary fields that out news model needs.

NEWS API

There is a third party rest API service called - 'NEWS API' which provides live news articles throughout the globe. It is accessible through - https://newsapi.org/

The application sticks to the database data for news data even when we get the live news without any restrictions from NEWS API service is to stay in some safe zone for the final project presentation and avoid any dependencies in order to avoid any case of unexpected issues of loss of data at the time of presentation [16].

Why NEWS API:

The News Api is chosen as the news dataset for our project as we require few major fields like category, url, url Image and ccountry details to present the news to any news readers in our application. And most importantly there is a requirement to classify the articles based on country and category and these fields can be made available with the news api request. Hence we decided to go with News API service.

Let's see the structure of news api request and response and how it can be made useful for us.

NEWS API Request:

The NEWS API request can be obtained from https://newsapi.org/v2/top-headlines which will result in all country top news that can be processed.

The is carried out for few countries and categories from the NEWS API and then processed for the response.

Request Parameters:

The request parameters we have chosen for our application includes the following:

- **APIKEY:** This is more important and unique for individual users and allows access to the news api.
- **country:** The two-letter country code for whichever countries we need news.

- **category:** The category news we may be interested in can be selected from the available set of categories list: Business, technology, entertainment, health, sport,general and science.

NEWS API Response:

The NEWS API Response data has many fields which makes it more flexible for our application.

- **Title:** The headline for the news.
- **description:** Descriptive information about the article to know more details.
- **URL:** To navigate to the respective url for the news publications to go through the whole article.
- **URL Image:** To have the respective images with the news article that makes more sensible.
- **Source:** The valid source information for the article who has published the news.
- **Published At:** The published time and date for the article by the source.

  The response does not have country and category fields directly available but appended to the items in the database.

5.5.2 News Feed

The next phase of the application is to present the news to the guest users to view the articles but to like or add them to favorites they need to have a valid login credentials and logged into the application.

NEWS FEED Design: News articles is now available on the database. The only operation we do here is to just fetch or read the news for a guest user. The Guest user access is controlled to view just view and read the news and restricted from interactions with the article as they interactions cannot be monitored for personalization.

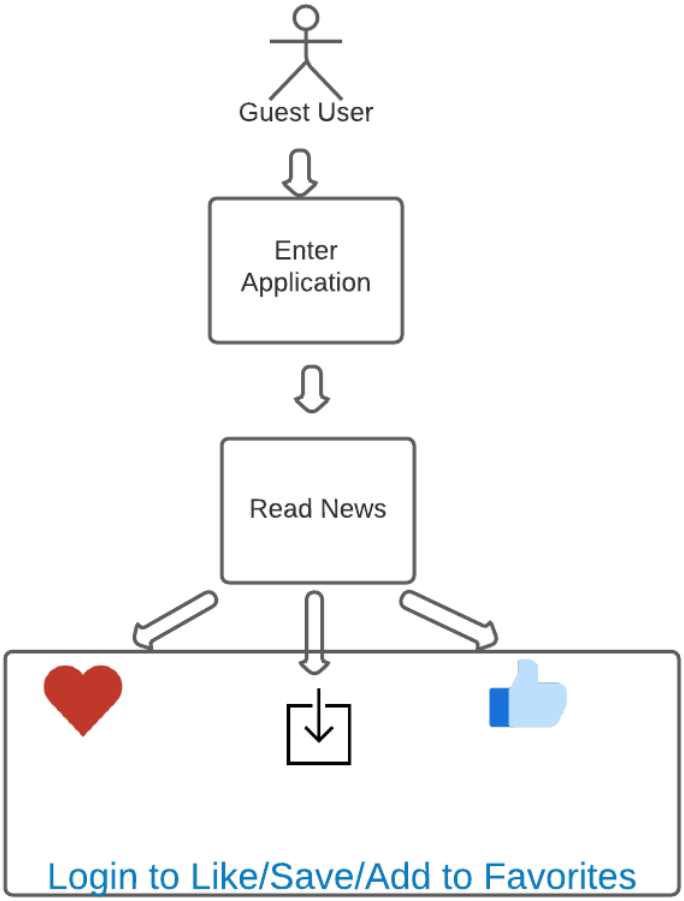

Figure 5.9: Guest User

5.5.3 Registration and Login

1. Add User Profile

- **User Registration:** The users are requested to give some basics information about them to derive their interests from their context details for setting up their initial preferences. This user details page is primarily important for 'User Profiling'.
- **User Login:** Users login module works with JWT authentication mechanism to allow users to login and view and interact with the application.

Spring Security and JWT Authentication This module is designed inline

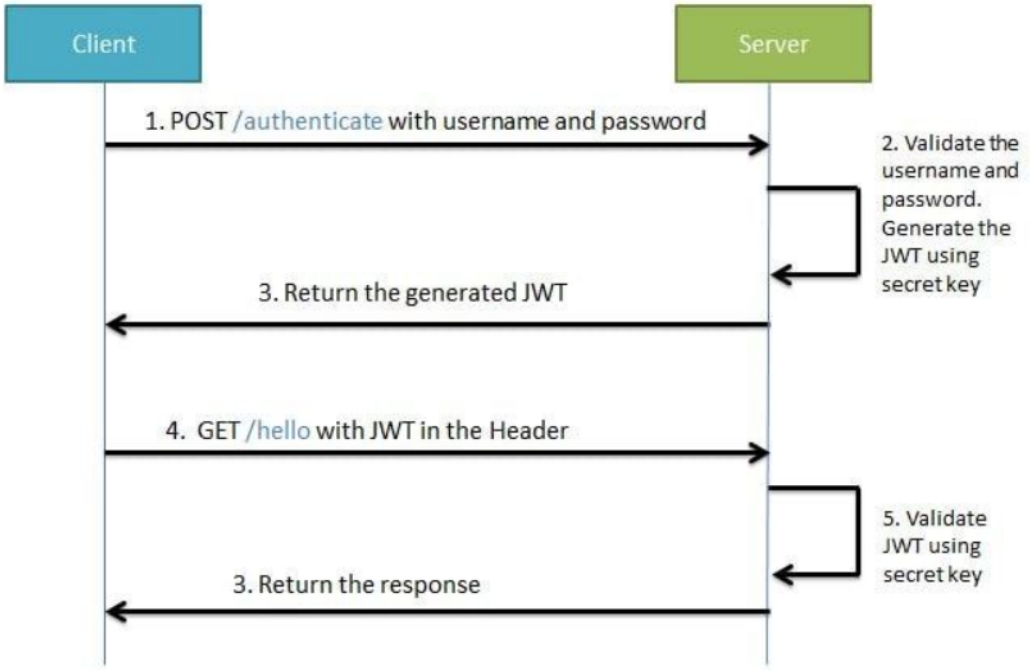

Figure 5.10: JWT Authentication

with the Spring Security concepts which authenticates the user with their login credentials. On valid Authentication with the username and password, the server responds to the client with a valid **JWT token** generated with the user details. Upon receiving the JWT token, the client has to send all the http requests with the authentication header in bearer token form otherwise it the request would be rejected by the server.

This way it secures all the http requests and response in Spring Security application [5].

Spring Security Architecture Authentication manager:

Spring Security works with Authentication manager processing the http requests and the response. **Authentication Providers:**

Authentication manager has multiple Authentication providers which process different types of authentication services. This is an interface which has a function called 'authenticate' which authenticates the user upon request.

User Details Service Class:

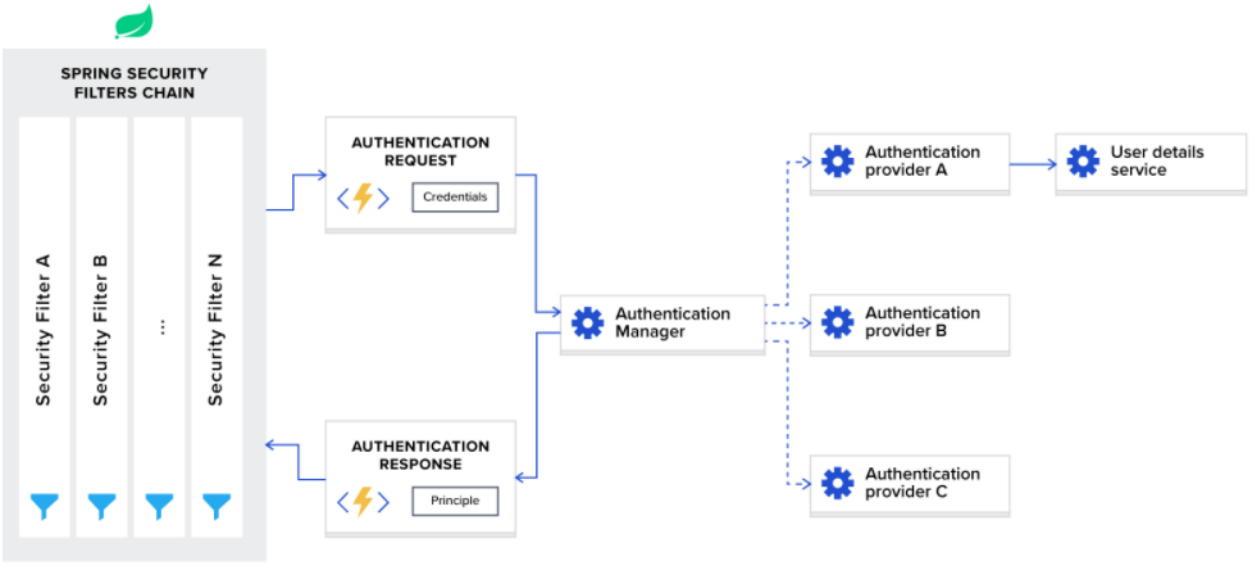

Figure 5.11: Spring Security

Additionally on valid authentication, it is necessary to generate JWT token with the username. Hence to achieve this we implemented **UserDetailsService** class in our application which loads user data with the username using the method loadByUsername. This helps in generating the JWT tokens with user claimed details.

2. **View User Profile:**

   Once the user is logged into the application, he lands in the user details page where he can view his/her profile details given during the registration process. This is processed as GET REST API call to the backend.

3. **Update User Profile :**

   The logged in user can also update his user profile from the client side anytime and view the updated details of his profile. This is processed as UPDATE REST API call to the backend.

5.5.4 User Personal Feed

User personal feed loads different content for each user based on user's likes and interests. It analyses individual's interests with the static and dynamic states and

decides the overall weightage for the preferences of categories for each user and loads them respectively.

Here is where personalization comes into picture. Personalization plays a key role in this application and configured in two different ways as mentioned.

User Category Map: Each user on signup is assigned a Hashmap which maps users preferred category and their respective weight. So this keeps track of the users most favorite category based on weight.

- **Contextual Personalization:** Achieved with the User Profiling. Here the selected preferences during the signup and the location are considered. The selected categories on signup will be added **+2.0** weights and the rest **0.0** . This map stores individual users category weightage. This is performed each time when a new user is created.

  The location preference is used in visualized personalization context which comes in a later view.

- **Behavioral Personalization:** This works understanding the users varying interests everytime and keeps it updated in the category map on change. It analyses the users likes, dislikes, favorites, saved news and understands the preferences implicitly and maps the preference weights accordingly.

  So, lets discuss how it judges the various interactions programmatically in brief with a pseudocode following the explanation.

  (a) **Likes:** The likes are identified with a filled thumbs-up icon and everytime the users like any article the respective category is identified and that category weight is increased by **+1.0** each time. This way it keeps the liked category into account and the user's category-map is updated.

  (b) **Dislikes:** The likes are identified with an unfilled thumbs-up icon and everytime the users like any article the respective category is identified and that category weight is decreased by **-1.0** each time.

This way it keeps the liked category into account and the user's category-map is updated. This is marked a lower weight as the likes and dislikes sometimes are random hits based on the story or the article and cannot completely assure the interest rate.

Like Zone: Also, the associated liked articles are mapped to a list and made available for users anytime to re-visit their liked ones.

(c) **Favorites:** The favorites are denoted by a filled heart icon and the user's favorite article's category is identified and the respective category weight is increased by **+3.0**.

(d) **Unfavorites:** The favorites are denoted by an unfilled heart icon and the unfavorite article's category is identified and the respective category weight is reduced by **-3.0**.

This is marked a higher weight compared to other interaction as the favorites explicitly states they are the most interested and remain almost the same.

Favorite Zone: Similarly users favorite articles are monitored and stored in a separate view for the users to revisit and go through their favorites anytime.

(e) **Saved Articles:** The users are allowed to save their news articles to make them available anytime for later read. This is a similar functionality that we have in many applications like Read Later and Watch later items. This helps the users to keep track and not miss the articles from reading at a later time. This is highlighted by a filled download icon and the weight associated with this option is **+2.0** for the respective category on save.

(f) **Delete Saved:** This allows the user to remove any article from saved list once they read through the article and there is no purpose having it in the saved items or read later items anymore. The respective added weight

while saving is decreased here to maintain the consistency of the category preferences

Read Later: As mentioned there is a separate view for read later items to make it easy for users to know where exactly the saved items are and help them read and delete anytime based on their interests.

Automated Personalization:

The above theoretically described personalization is achieved using a custom-designed algorithm called **'Weight-Map' Algorithm**.

Weight-Map Algorithm:

Weight Map Algorithm builds user Personalization programmatically based on above mentioned factors or user engagements. Each user engagement is counted and assigned respective weights as described above. The algorithm is named as 'Weight-Map' algorithm because a hash map is created for each user which maps users category based weights in the backend.

Similar Existing Algorithm: Facebook determines the personalization of feeds based on an algorithm called 'Edge Rank' Algorithm[6]. This is a rank-based approach considering three major factors: Affinity Score, Edge Rank and Time Decay. These factors ranks and decides the order of appearance of stories in the user's newsfeed.

The updated facebook algorithm works based on the likelihood of the user interatcions with the post. This is a complex algorithm which has to cross four systems.

1. **Inventory :** This is the initial phase where the facebook decides what all contents to load for any user based on his affinity which means the close relationship graph.

2. **Signal :** This phase has to approve the content based on various assessments like type of feed, whom and when the content was posted, the likes, shares,comments and all other factors and given **weightage**.

3. **Predictions:** This phase predicts how far the respective user might like the posts

based on his previous behaviour[1].

4. **Score:** The above two phases are combined like the predicted story and the respective weight for that story are combined and assigned a relevancy score which helps in fetching the news feed sorted in descending order.

Comparison - Similarities and Differences To Weight-Map Algorithm:

1. **Phase 1:**

   Weight-Map algorithm works in a simplified pattern taking all the contents at the first phase.

2. **Phase 2:**

   The second Signal phase works bit differently. Instead of evaluating each content, we assign weights for user interacted content's category everytime the user likes, saves or add the news to favorites. Actually this is the third phase of predictions according to facebook algorithm which is our only filtering mechanism and scoring method for assigning weights.

   For example, The user likes certain article, then the category weight for that article increases by +1.0 and the vice-versa.

3. **Phase 3:**

   This is the final phase where we filter based on weights or score assigned to categories interested for each user. Each user will have a category-weight map assigned in previous phase which is sorted in descending order and then based on more range of weights to load respective number of articles.

   For Example. category score more than 50, loads 20 articles from that category and for a lesser score the limits decreases. Then with the overall category weightage we provide the users with their personal feed.

Visual Representation of Algorithm: The below figure helps us to understand more clearly on how this algorithm works. **Pseudo-code of Weight-Map Algorithm:**

```
User register:
```

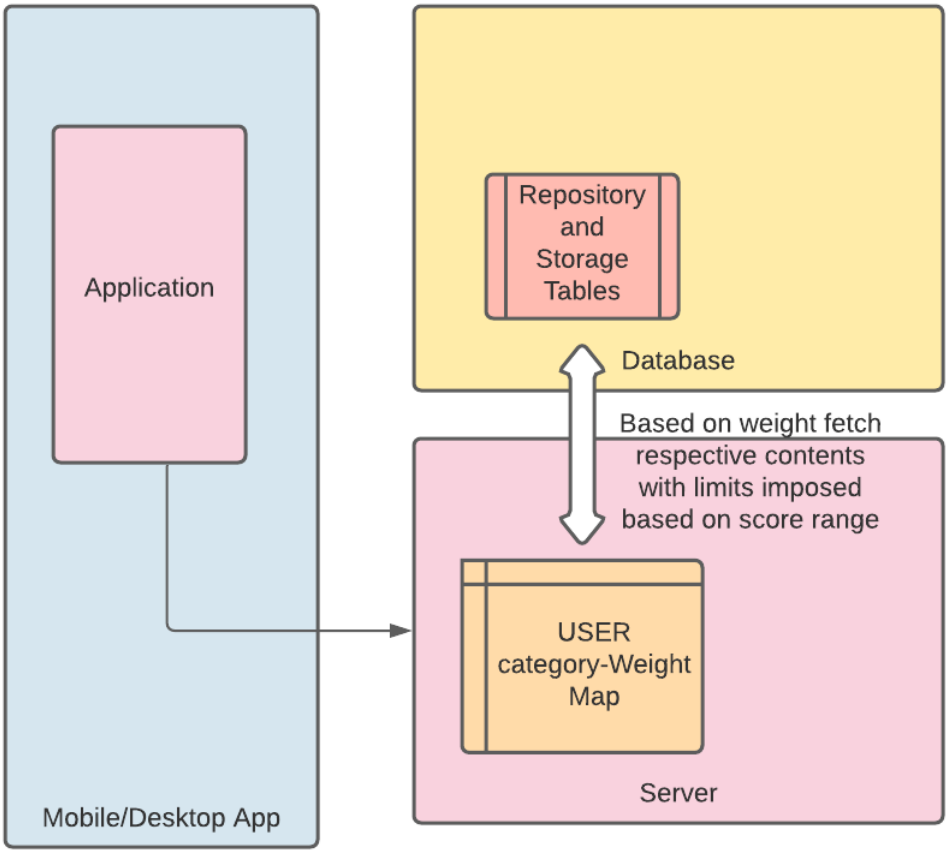

Figure 5.12: Weight-Map Algorithm Flow

```
        if preferredcategory in categoryList:
                category-map.weight = 2.0
    else
                category-map.weight = 0.0
    Add user with cmap
Monitor Actions:
  if(logged in):
        if click(Like):
           if(liked):
                fetch category for this feed
                    Add category-map.weight - 1.0
        else:
```

```
                fetch category for this feed
                Add category-map.weight + 1.0

        if click(Favorites):
            if(favorite):
                fetch category for this feed
                Add category-map.weight - 3.0
        else:
                fetch category for this feed
                Add category-map.weight + 3.0

        if click(Save):
            if(saved):
                fetch category for this feed
                Add category-map.weight - 2.0
        else:
                fetch category for this feed
                Add category-map.weight + 2.0
  else:
    print(Login to personalize!)
```

And to prevent the userfeed from overload with all news of a particular preferred category, the application imposed limits on the results to load the news collection from specified category based on its weight. More the range of weight, more are the number of results from that particular category and vice-versa. This enables to prioritize the top category preference and load them more instead of loading everything equally.

```
User-feed:
        fetch sorted category-map based on weight
        for sorted category in category-map
                if(weight > 50)
```

```
            fetch top 20 newsfeed on this category
if(weight > 10)
            fetch top 12 newsfeed on this category
if(weight > 5)
            fetch top 8 newsfeed on this category
else
            fetch top 5 newsfeed on this category
```

5.5.5 Visualization

Visualization is accomplished by a special library called amcharts which gives sensational maps and charts which are effective for locating the news around the globe and understanding the user chart data.
Visualization is applied in two areas in our application.

1. **User Charts:** User charts are visible in the user profile view where every user will be able to see their category chart which shows them their most favorite category and the overall percentage of interests he expressed while reading through the news articles in the application for all the available categories.
   This is represented in terms of pie-chart which shows the category-name and rate of interests in percentage along with the number of articles which made the figure.

   Pie-charts: Pie-charts represents a part-to-whole relationship. Every piece or slice of a pie-chart represents one component which together contributes the whole[2].
   Why Pie-charts: Pie-charts help clearly help us to understand how much each category contribute in a user's interests because pie-charts are best used in case of explaining and understanding part-whole relationship. Also, it is bit more attractive than simple column or bar charts in a way hence the choice of pie-charts.
   Pre-requisites: We installed the external library - **@amcharts/amcharts4** to have the basic functionalities from amcharts 4. Once integrated the amcharts, we can extend to any additional pluggins which we require in future.

To achieve amcharts pie-charts we required two modules - **am4core and am4charts** which were installed.

2. **Globe View:** Another area where we have the visualization is the map or the globe view in the application where it projects a map in orthogonal projection.

   Pre-requisites: For acquiring the maps, we need **am4maps** module which gives the globe view and we can make this to appear in any projection like a globe or a flat map.

   Why Orthographic projection: We have chosen **Orthographic projection** in this application which is a globe structure because it is more interactive and looks appealing to the users as it keeps rotating around which makes it lively.

Here is how it looks when projected: **How it is designed to help us:**

Figure 5.13: Amcharts Map

Visual Country Filter: This view of globe allows users to chose which country's news they would like to read through anytime. This helps in filtering the news for

each users preferred country.

Category Filter: At the same time, when the users prefer to know more about each country specific category news, the application gives them another option to select any category they need from a list of categories which again filters selected category data for selected country.

Combined Filters: Thus two filters are applied on user's preference. So, users can read any country news and any category news and compare the category news data against any other country if required.

So, technically the combination of filters is achieved using **Behavioral subject** in angular. This helps in retaining the category selected for change in country and change the category selection only on subscription.

For example: When the user selects 'Australia' the application loads the Australia country results with 'All' categories selected. When the user changes the category to 'Sports' the user will be able to look specifically Australia's Sports data and change to any other category or country as preferred.

Personalized Visualization:

Initially to make it more personalized, we load the **user location** by default with **'All'** categories loaded in the results. This allows the user to open the view and have a look at his/her country data and option to see any category data for the location.

Whom It helps Mostly: It is mostly helpful for journalists or news organizations who compare news against countries and want to understand the worldwide news in compare and contract approach. Also, it obviously helps the general users who are interested to know more information anytime.

5.5.6 Recommendation

Recommendation is provided for each user based on user's interests rate compared against other users who have similar interest pattern.

Mahout Recommendation System: This is achieved using **Mahout Recommenda-**

tion system which handles real-time users interests data and recommends articles based on Collaborative Filtering algorithm. It is chosen because it supports different approaches for Collaborative filtering both item-based and user-based and also highly scalable for large datasets[19]. For Example, Amazon and Netflix recommendation systems.

Recommendation System Architecture: Recommendation works in several stages as follows:

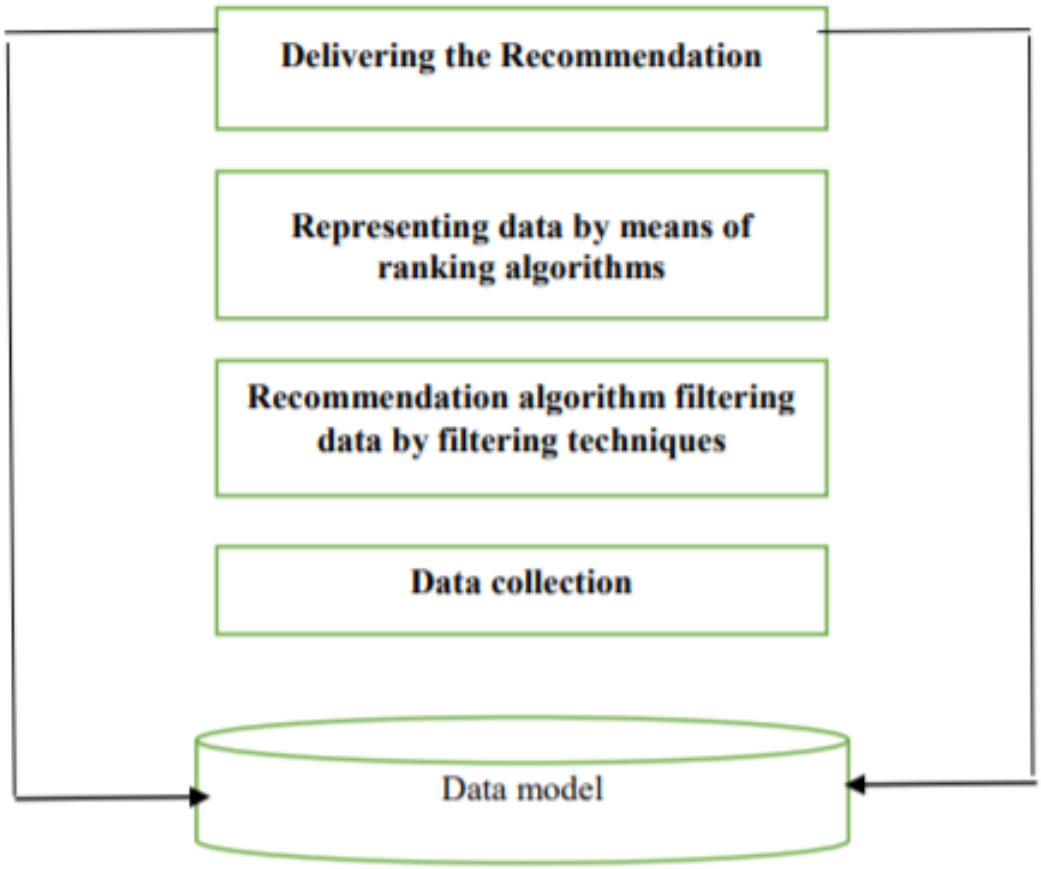

Figure 5.14: Recommendation Architecture

Proposed Approach - User Based Recommendation: User-based approach is chosen for this application and it works on the concept that users who have similar interests will be interested towards similar categories[14].

Algorithm works in three steps[14].

```
For every item I that u has no preference for yet
For every other user v that has a preference for i
    Compute a similarity s between u and v
```

```
Incorporate v s preference for I weighted by s to avg
Return the top items, ranked by weighted average
```

Explanation of Algorithm:

1. 1. Find user u with no preference for item say. i
2. 2. Find other users v who has preference towards i and have similarity model between u and v
3. 3. Compute average for the preference for items and recommend items with top weight.

How does it work for Buzzfeed: Buzzfeed also works out this algorithm in three stages except that incorporating weight is not necessary here as we have them pre-calculated for every user in user-category map.

Recommendation Input Data: The input data is obtained from user-category map stored in database table where we have the user category and respective weights for every user which we convert to csv files for Recommendation process.

How it Works: Let's discuss the working process with an example. When user 'u' has more interest weights for categories 'Sports' and 'Science' and there are other similar users with more which is around similar weight as of 'u' for 'Sports', 'Science' and additionally has more weight for other category 'Club', so the user 'u' is recommended with some items in 'Club' which might sound interesting for the user.

Observed Limitation: This is more related when the there are more number of categories and users are not sure which one they find to read sometimes. For our application, we have less number of categories due to some dataset limitations but it helps in understanding how it works clearly.

Chapter 6

Implementation

Providing explanations and if possible piece of code to understand the implementation of the desired designs highlighted in the previous chapter. This chapter should clearly explain how the designs are implemented at a detailed level.

Chapter 7

Evaluation

Explanation for the testing to be achieved and some evidence of testing screenshots and results after resolving the bugs if possible. Preferably use-cases for testing observations conducted during the process could be helpful.It should have the testing framework, sample evidences of user testing and observed results and bug-fixes to make the application complete.

Chapter 8

Challenges and Contributions

This chapter should explain the challenges and results of accomplished work and future considerations for enhancements if available.

8.1 Challenges Faced

Provide enough examples or evidence of where the application was challenging to achieve and what all the solutions developed to overcome the challenges. If possible provide a detailed explanation of the root cause, source and solutions for the challenges faced and if unresolved provide the reason and workarounds if applicable.

8.2 Contributions and reflections

Explain the process of achieving the requirements through various stages and plans. Provide the results achieved with the application and the test results if applicable. Establish the overall goals achieved through the project and the beneficial aspects.

Bibliography

[1] 21, S. S. o. A. Use the facebook algorithm to create meaningful interactions, Jun 2021.

[2] Javascript charts and maps, Nov 2018.

[3] ANNE. New german edition gets deep personalization powered by machine learning, May 2018.

[4] DANIELS, C. Personalization deep dive: Contextual vs. behavioral, Apr 2017.

[5] Erin, Y. K. How to set up java spring boot jwt authorization and authentication, Dec 2020.

[6] Facebook news feed algorithm, Dec 2018.

[7] Feng, C., Khan, M., Rahman, A. U., and Ahmad, A. News recommendation systems - accomplishments, challenges amp; future directions. *IEEE Access 8* (2020), 16702–16725.

[8] Garrigs, I., Gomez, J., and Houben, G.-J. Specification of personalization in web application design. *Information and Software Technology 52*, 9 (2010), 991–1010.

[9] Gitsham, O. The new bbc news app good or bad user experience?, Jun 2015.

[10] Jonnalagedda, N., Gauch, S., Labille, K., and Alfarhood, S. Incorporating popularity in a personalized news recommender system. *PeerJ Computer Science 2* (06 2016), e63.

[11] Lee, Y. Recommendation system using collaborative filtering. Master's Projects. 439.

[12] Liu, J., Pedersen, E., and Dolan, P. Personalized news recommendation based on click behavior. In *2010 International Conference on Intelligent User Interfaces* (2010).

[13] Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. Recommender system application developments: A survey. *Decision Support Systems 74* (2015), 12–32.

[14] Madhushree, B. A novel research paper recommendation.

[15] Mishra, U. What is a content-based recommendation system in machine learning?

[16] News api search news and blog articles on the web.

[17] Ngo, N. Android software development: Case:logo quiz world mobile application.

[18] Rao, R., and Swamy, S. Review on spring boot and spring webflux for reactive web development.

[19] Seminario, C., and Wilson, D. Case study evaluation of mahout as a recommender platform. vol. 910.

[20] Spring boot architecture - javatpoint.

Appendix A

User Manuals

Provide user stories to help with the workflow process for a new user to understand the application and go through the features.