CONTEXTUAL NEWS INFORMATION RETRIEVAL

Zain Ul Abedin

Annas Israr

Faizan Ahmad

Yaldram Shahzad

Supervised by:

Engr. Muhammad Umer Haroon

Submitted for the partial fulfilment of BS Software Engineering degree to the Faculty of Engineering and Computer Science

NATIONAL UNIVERSITY OF MODERN LANGUAGES
ISLAMABAD

FEBRUARY, 2020

ABSTRACT

Data generation and its growth rate is an abrupt process these days and will grow exponentially with each passing day. Users on the internet can enjoy abundant services and information in ecommerce websites, electronic newspapers, blog & social networks. Although this data is available for its consumption by users, quite an amount of time is spent retrieving this information and processing it. This has favoured the research in several fields such as web scrapping. Web scraping, a process of extracting useful information from HTML pages, which is the main formatting tool of information on the internet today. Web scraping is a hot topic in today's perspective, and it has multi faced applications. But two of the most important utilities of scraping are information retrieval for personal usage and for analytical purposes.

In this project, the aim is to do a survey of personalized information retrieval for statistical purposes, a specialized and crucial subsection of information retrieval and propose a system that will do this job on behalf of user. The proposed system will solve the above-mentioned problem by searching the web pages for the relevant information and extracting the information that is relevant to the user's context. Methods that are chosen for information retrieval as Web Scraping, a technique that is extremely popular and is proven to have multi-domain usage these days. The proposed system is currently limited to Pakistani news websites only.

TABLE OF CONTENTS

CHAPTER 1 10

INTRODUCTION 10

1.1 Introduction 11

1.2 Motivation 11

1.3 Problem statement 11

1.4 Scope 12

1.5 Process model 12

1.5.1 Requirement gathering 13

1.5.2 Work breakdown and decomposition 13

1.5.3 Designing 13

1.5.4 Coding 13

1.5.5 Integration 13

1.5.6 Testing 13

1.5.7 Why incremental model 13

1.6 Proposed system 14

1.7 Features 14

1.7.1 Text Scrapping 14

1.7.2 Scrapping of text in images 14

1.7.3 Sign up / Sign in 14

1.7.4 Guest user 14

1.7.5 Merging and referencing 15

1.8 Nature of the project 15

1.9 Summary 15

CHAPTER 2 16

BACKGROUND AND EXISITING WORK 16

2.1 Introduction 17

2.2 Existing Systems 17
2.2.1 Google News 17
2.2.2 ABC News 17
2.3 Limitations of Existing Systems 17
2.2.2 Providing Links 17
2.2.3 Unavailability of Information in Document Form 18
2.2.4 Not Analysing Pictorial Information 18
2.2.5 Time Consuming 18
2.2.6 Presenting the information 18
2.3 Solutions to Limitations of Existing Systems 18
2.3.1 Providing Links 18
2.3.2 Availability of Information in Document Form 18
2.3.3 Analysing Pictorial Information 18
2.3.4 Time Saving 18
2.3.5 Presenting the information 18
2.4 Comparison of Existing Systems and Proposed System 18
2.5 Summary 19
CHAPTER 3 20
REQUIREMENTS SPECIFICATION 20
3.1 Introduction 21
3.2 Interface Requirements 21
3.2.1 Software Requirements 21
3.2.2 Hardware Requirements 21
3.3 Functional Requirements 22
3.3.1 User Registration 22
3.3.2 News Search 22
3.3.3 Display of Information 22

3.3.4 Latest news....22
3.3.5 Newsfeed....22
3.3.6 History and Interest....22
3.3.7 Save information on local machine....22
3.4 Non-Functional Requirements....22
3.4.1 Security....23
3.4.2 Reliability....23
3.4.3 Availability....23
3.4.4 Maintainability....23
3.4.5 Portability....23
3.5 Use Case diagrams....23
3.5.1 Sign in....25
3.5.2 Sign up....26
3.5.3 Search news....28
3.5.4 Manage history....29
3.5.5 Manage interests....31
3.5.6 Manage news results history....32
3.5.7 Save news....34
3.6 Resource Requirements....35
3.6.1 Web Scraper....35
3.6.2 Tools....35
3.7 Database Requirements....36
3.8 Project Feasibility....36
3.8.1 Technical Feasibility....37
3.8.2 Operational Feasibility....37
3.8.3 Legal and Ethical Feasibility....37
3.9 Summary....37

CHAPTER 4 39
SYSTEM MODELLING 39
4.1 Introduction 40
4.2 System design 40
4.3 Design Approach 40
4.4 Interface design 41
4.5 High fidelity Prototype 42
4.5.1 Landing page 42
4.5.2 Signin 42
4.5.3 Signup 43
4.5.4 Search news 43
4.5.5 User homepage 43
4.5.6 Latest news 44
4.5.7 Newsfeed 44
4.5.8 Change account name 45
4.5.9 Change account password 45
4.5.10 Delete account 46
4.5.11 Clear search history 46
4.6 4+1 view model of architecture 47
4.7 Logical View 47
4.7.1 Class diagram 47
4.8 Process View 48
4.8.1 Activity Diagram 48
4.8.2 Sequence Diagram 50
4.9 Development View 55
4.9.1 Component diagram 56
4.10 Physical View 56

4.10.1 Deployment diagram....56
4.11 The Use Case view....57
4.12 Summary....57

LIST OF FIGURES

Figure 3.1: Use case diagram for Contextual News Information Retireval.....24
Figure 3.2: Signin.....25
Figure 3.3: Signup.....27
Figure 3.4: Search news.....28
Figure 3.5: Manage search history.....29
Figure 3.6: Manage news interests.....31
Figure 3.7: Manage news results 31
Figure 3.8: Save news.....34
Figure 4.1: Top down approach of system.....41
Figure 4.2: Class diagram of a system.....48
Figure 4.3: Activity diagram of a system.....49
Figure 4.4: Sequence diagram for Signin 50
Figure 4.5: Sequence diagram for Signup.....51
Figure 4.6: Sequence diagram for Search news.....52
Figure 4.7: Sequence diagram for Search history.....53
Figure 4.8: Sequence diagram for Neewsfeed 54
Figure 4.9: Sequence disgram for Search results.....55
Figure 4.10: Component diagram for a system.....56
Figure 4.11: Deployment diagram for a system.....57

LIST OF TABLES

Table 1.1: Scope of the project 12
Table 2.1: Comparison of existing and proposed system 19
Table 3.1: Sign in 25
Table 3.2: Sign up 27
Table 3.3: Search news 29
Table 3.4: Manage search history 30
Table 3.5: Manage news interests 31
Table 3.6: Manage result history 33
Table 3.7: Save news 35

CHAPTER 1

INTRODUCTION

1.1 Introduction

The world has entered in the modern era of technology and technology has captured everyone, everything and every process in its grip. There is an enormous and almost infinite amount of data and information that is now available on the internet, which is available for people to access.

This has made a major impact on the people's way of living and due to this almost every person is trying to get information online. The time of the users is also very precious and almost everyone wants to do their work in a more convenient way. There are about billions of articles and newspapers with numerous facts and important information. In order to look for some desired facts or news a person need to go through multiple sources online in order to get their required information.

So, in order to solve or minimize this problem we are proposing a web-based application for contextual extraction of relevant news information according to the required query of the user. The user will be able to search the desired news by using our application that will automate the process of searching and extract the required information for the user which would save a lot of time. This app will also merge gather the extracted information from different sources in a document along with the references that the user would be able to import in his local device.

1.2 Motivation

As there are tons of information over the internet, sometimes searching for your required news/information can be hectic and frustrating. The motivation behind the development of this application was to reduce wastage of time that people had to face in order to get the relevant news of their interest and secondly, to provide news from authentic sources.

1.3 Problem statement

In the modern world of technology internet is available to almost everyone. Internet is now the biggest means of mass media and by using this one can change the ideology and way of thinking of a huge number of people. Different type of fake news and propagandas keep circulating all around the internet. So, in order to reduce or eliminate such acts we have come towards the development of this system.

This system will extract your required information from trusted and authorized resources with links and references of these resources automatically and thus saving a lot of time and effort of the user.

1.4 Scope

Table 1.1: Scope of the project

Scope of the project	
Title	**Contextual News Information Retrieval**
Description	A web-based application that automatically searches and extracts the news from different sources on the internet based on the users query and provides the user with a merged document of the information along with the links and references.
Justification	The main objective of this system is to save the users time of browsing over the internet for desired information and to make sure that the news is coming from authorized and trusted sources.
Constraints	Limited number of Pakistani English newspapers.
Assumption	A fast & stable internet connection is available, a working web browser.
Stake holders	User, news researchers, news reporters, developers, testers, admins.
Risks	Lack of internet, slow internet connection, faulty internet browser.
Deliverables	A web-based application, final documentation report.

1.5 Process model

In the development of our system we have used the incremental process model. We broke down our work into different modules. The modules were then developed and integrated. Different phases are as follows.

1. Requirement Gathering
2. Work breakdown and decomposition
3. Designing
4. Coding
5. Integration
6. Testing
7. Deployment

1.5.1 Requirement gathering

The requirements and specifications that were needed for Contextual News Information Retrieval system were gathered from different sources.

1.5.2 Work breakdown and decomposition

The system was decomposed into the following modules in this phase.

- Scrapping Texts
- Image processing for conversion into text
- Scrapping data on images
- Extracting data from sources
- Merging data at one place
- Making a document
- Providing references of the extracted information

1.5.3 Designing

In this phase we transformed the different requirements of the system into properly defined high level specification of each of the modules of our system.

1.5.4 Coding

The actual development and coding according to the requirements and design specification has been done in this phase and has been repeated for each module of the system.

1.5.5 Integration

After the different modules were developed, they were now integrated at one place in this phase.

1.5.6 Testing

Different modules were being tested parallel to the development but the main system testing and high-level testing was done in this phase after the integration of the modules.

1.5.7 Why incremental model

- Tasks can be divided.
- Problems can be easily identified and managed.
- Less skilful team can manage it.
- Can focus on the module under development.

1.6 Proposed system

We have developed and proposed this system to successfully reduce the wastage of precious time of the user who want to look for relevant news of their choice and to minimize and eliminate the fake news by providing the user with the content from trusted and authorized sources.

This web-based application will allow the user to sign up to our system or use it as a guest user. The user will type a query of the wanted news and the system will automatically search our predefined sources and retrieve the desired information to the user and present it in a document that can be exported or downloaded into any device. We will use python 3.0 for backend development in which our scrapper and image processing algorithms will be developed. We will be using Flask SQLITE DB for maintaining our database and keeping the scrapping query history and news preferences of the user that will sign up.

1.7 Features

Following are the five main features of this system.

1.7.1 Text Scrapping

This feature will allow the user to write in a query of some sort and our application will start searching and scrapping the predefined sources for news that will somehow match the query of the user and be relevant and will extract it from the source.

1.7.2 Scrapping of text in images

This feature enables the system to look and search for the image with the desired data, convert the data to text and retrieve it back for the user that will be according to the query of the user.

1.7.3 Sign up / Sign in

This feature will enable the user to sign up to our web based application which will help him in setting up his preferences of news of his interests which would help us to provide him with news of his topics of interests by just one click.

1.7.4 Guest user

Our system does not limit our users to sign up. Any user that will come to our website will be able to avail its basic functionalities and will be provided with the top trending news by just a click.

1.7.5 Merging and referencing

Our system will merge and combine the data that is extracted from different sources in a text document and will provide the references of the different chunks of information parallel to them.

1.8 Nature of the project

Contextual news information retrieval is a web-based application that is made using python at the back end for data mining and machine learning algorithms like Image processing and Web Scrapping. We are using Flask SQLITE DB for maintaining our database.

1.9 Summary

In this chapter we have discussed the surface details of Contextual News Information Retrieval. We have highlighted what is the system that is being developed, why is it being developed, how will it be operated, how will it impact on the modern world, for whom it is being made, for which domain and field will benefit the most, where will it be organizationally located and where will it be in the future.

CHAPTER 2

BACKGROUND AND EXISITING WORK

2.1 Introduction

In this chapter, we are going to discuss about the existing systems that are related to CNIR (Contextual News Information Retrieval). News researchers and students of journalism have to spend a lot of their time to research about different topics that are under their consideration. CNIR provides a solution to save their time and effort that they have to spend. They can just simply search and leave rest of the work to the system, CNIR will search their related information and will retrieve the information related to their query. Google News is an existing system which does a pretty similar job, but it provides you links, and you have to individually go through those links to get your desired information and it contains only articles and blogs. ABC News is another system which provides the link to sources and does not have any mechanism to download the information on your system. CNIR will solve these problems.

2.2 Existing Systems

Following are some existing systems.

2.2.1 Google News

Google News is a news aggregator application to search different news made by google. It provides the facility of daily updates, provides search facility, it provides different categories for the users to search information. But its drawback is that it only provides you with links and you have to individually visit those links and there is no mechanism to download the information.

2.2.2 ABC News

ABC News is another news website by American Broadcast Company. It provides searching facility and allows you to see latest stories. But again, it does not have any mechanism to download information and also it does not analyse pictorial information and E-Papers.

2.3 Limitations of Existing Systems

2.2.2 Providing Links

All these existing systems provide links to resources but does not provide you with the extracted information that is relevant to your searched query.

2.2.3 Unavailability of Information in Document Form

Google News and ABC News does not have any mechanism to download the extracted information in the form of a document.

2.2.4 Not Analysing Pictorial Information

ABC News does not analyse pictorial information. It only analyses the information that is available in the form of text.

2.2.5 Time Consuming

These existing systems provides links and users have to open each link individually and have to search for the data that is relevant to their context, it is really time consuming.

2.2.6 Presenting the information

These existing systems does not present the information, rather they just show the links.

2.3 Solutions to Limitations of Existing Systems

2.3.1 Providing Links

CNIR extracts the information and shows it to the user rather than just giving the links.

2.3.2 Availability of Information in Document Form

CNIR provides a feature to download the information that you searched in the form of a document or you can just read the information.

2.3.3 Analysing Pictorial Information

Analysing the information present in the form of image snippets etc. Is also an important task. CNIR analyses those image snippets and other pictorial means of information.

2.3.4 Time Saving

CNIR saves the time of user because user does not have to go through all the individual links by himself, instead he can search for the information and leave the work to the system which searches for the relevant information and extracts the information for the user.

2.3.5 Presenting the information

CNIR presents the information in a readable form, also the user can download the information in the form of CSV file. So, the system provides the information in readable form online and also in CSV.

2.4 Comparison of Existing Systems and Proposed System

A comparison between CNIR and the two existing systems is given below.

Table 2.1: Comparison of existing and proposed system

Features	**Google News**	**ABC News**	**Proposed System**
Text Extraction	No	No	Yes
Text Extraction from Pictures	No	No	Yes
Analyzation of user's script	No	No	Yes
Download Option	No	Yes	Yes
Searching Option	Yes	Yes	Yes

2.5 Summary

In this section, we will summarize that some existing systems for news search are Google News, ABC News etc but they all lack at some point i.e. data representation or download option etc. The proposed system solves these problems and tackles these features. In this way, the proposed system will facilitate its users and will provide them ease and convenience.

Most Importantly, the proposed system saves the time of its users and searches the information on their behalf rather than letting them go through the individual links themselves, and then the user can download and keep the results with himself which further provides convenience.

CHAPTER 3

REQUIREMENTS SPECIFICATION

3.1 Introduction

In this chapter, we are going to discuss system requirements of the proposed system. System Requirements are those software or hardware requirements that cause the system to do what it is supposed to. The absence of these requirements can cause serious system failure, system failure can be in the form of performance issues or compatibility issues. Performance issues are those which cause system to perform below the bar i.e. the system can hang, or crash and compatibility issues can be when the deployed system is not compatible with the working environment or the hardware installed at the workplace. In this chapter, system requirements of the system are discussed. System requirements of our system include interface requirements, database requirements, software requirements, hardware requirements, database requirements, functional requirements and non-functional requirements.

3.2 Interface Requirements

Interface Requirements are those necessary measures which need to be met in order to interface different components of the system, as the system has many different modules and components so interfacing all these separate components is very important in order to compile the whole system and get the required functionality. In this section, we need to discuss that what we need to have in order to make the system communicating with all its components. As the system is a web application so the interface requirements are subdivided into further categories.

3.2.1 Software Requirements

As the proposed system is a web application so in order to use the system the user must have an active internet connection and a web browser.

3.2.2 Hardware Requirements

These requirements include minimum processor speed, memory and disk space required to install windows. In almost all the cases, user wants to make sure that his hardware is enough to support his tasks and perform all the required functions and services like running applications on the server with ease. As our system is a web-based application so it can be easily accessed on a system with a 1.8 GHz dual core processor, 4GB of ram and 500 GB of disk space.

3.3 Functional Requirements

Functional requirements are those which are explicitly stated during the phase of requirements gathering. Functional requirements are must for any system to be successful, these are the critical requirements that must be met by the system.

3.3.1 User Registration

1. System should be able to register a new user by prompting his name, email and a password.
2. System should ask for user's interest areas and save them for future searches to provide a better user experience.
3. System should be able to give control to users for updating their account details.

3.3.2 News Search

1. System should enable the user to search news of his choice.
2. System must allow the user to search for some specific categories by clicking on the buttons without having the need to type each time.

3.3.3 Display of Information

1. System should be able to display the extracted information in a readable form.
2. System should display the information based on user context.

3.3.4 Latest news

System should be able show latest information based on trending topics.

3.3.5 Newsfeed

System should be able to show newsfeed based on the history and interest of registered users.

3.3.6 History and Interest

System should be able to maintain history and interest of its registered users.

3.3.7 Save information on local machine

1. System should be able to give control to user by providing options to save and export the information in user's device.

3.4 Non-Functional Requirements

Non-Functional requirements are those which are not explicitly stated during the phase of requirements elicitation. Non-Functional requirements are some general/common requirements are considered to improve the quality of the system. The more often they are

implemented the more quality the system can have. Almost all systems have these kinds of requirements.

3.4.1 Security

Server end data shall be accessible to developers/administrators.

3.4.2 Reliability

System shall provide the database and backup of database to ensure the reliability of overall system. Since the database is the backbone of system so it shall be maintained, updated, and secured along with all reliable components of system.

3.4.3 Availability

System shall be available anytime, anywhere to user whenever the users need it. However, the time when the server or database is down or there is some technical difficulty, the performance and availability of the system shall be compromised. Any possible backup shall be available and shall be provided to users in this case until the problem has been identified and corrected.

3.4.4 Maintainability

System shall be maintainable and shall be maintained regularly to ensure that new changes can be implemented. Flask SQLITE DB is used to maintain the database and server is used to maintain the database and to look after the overall application.

3.4.5 Portability

The system is a web-based application, and it shall run on all web browsers and platform including the most popular ones to ensure the portability and compatibility related to every browser available.

3.5 Use Case diagrams

Use case diagram is a diagram that is drawn based on user goals. It serves as communication interface that show the interaction of user with the system and response of the system towards user. This system also has a main use case diagram and written use cases with diagrams that has many functions for the user to do. The system shows a homepage or landing page to the use that has many sub features related to website. This home page has a search bar from where a random user can the news of its own choice and context. Then in that same home page there is also a feature of latest news from where user can see the trending news. Apart from these random users features there are some essential features that are meant to be used and performed by registered users only. These features include

newsfeed, saving of news, interest-based news, managing of search history and news result history. These interesting features makeup the system named as Contextual News Information Retrieval (CNIR). Below are the most essential and primary use cases with their diagrams.

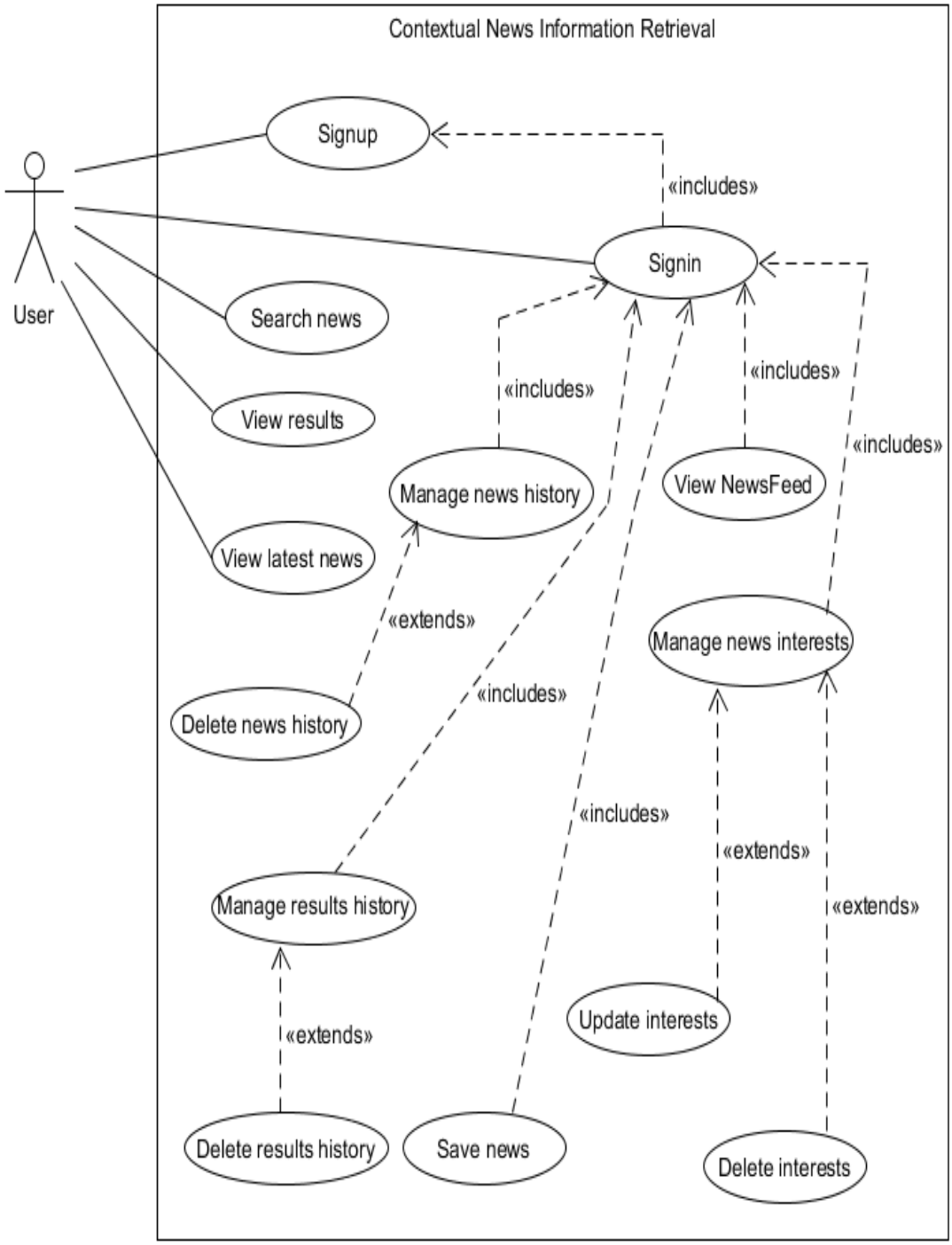


Figure 3.1: Use Case Diagram for Contextual News Information Retrieval

3.5.1 Sign in

In this use case the system prompts the user to enter the email and password and click sign in. In response to that the system verifies the email and password and passes the user to more feature rich home page to unlock and use special features. The system also gives response if the sign in gets unsuccessful and shows an appropriate massage.

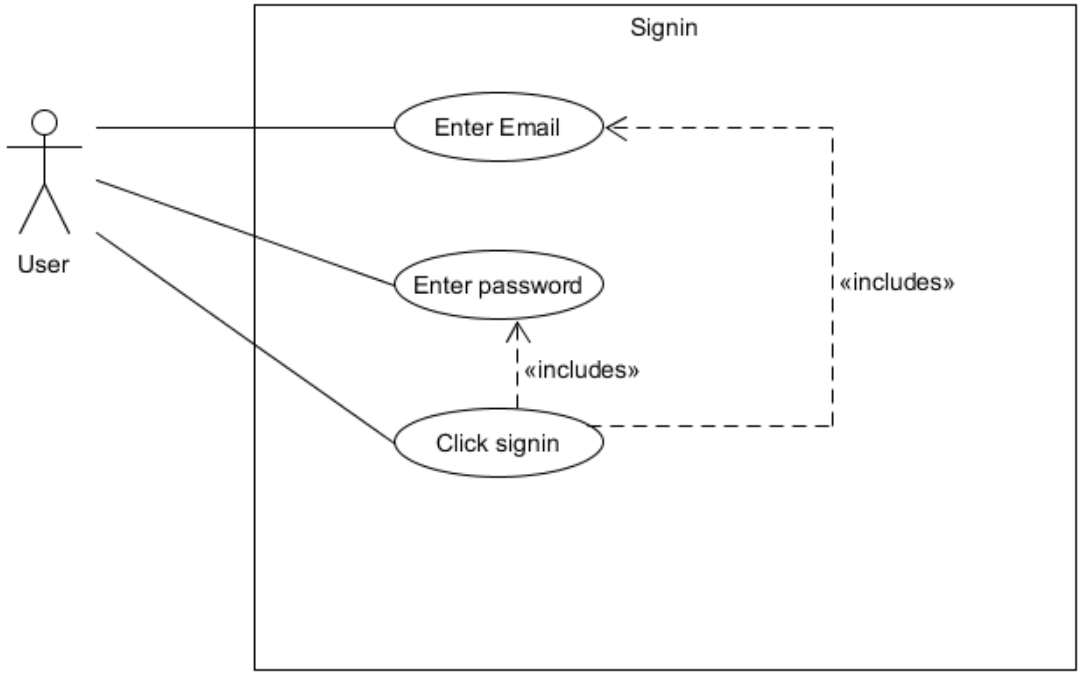


Figure 3.2: Use case diagram for Sign in

Table 3.1: Sign in

Use case ID	UC-1
Use case name	Sign in
References requirement	Requirement no. 1
Actors	User
Purpose	To sign in the user
Overview	The user enters the email and password and then clicks on sign in
Type	Primary and essential
Pre-Condition	User must be signed up

Post-Condition	User signed in successfully
Normal flow	
Actors Actions	System Response
1. The user enters the email and password that was entered during signup.	2. The system verifies the user email and password and shows next phase.
Alternative flow	
1a. The user enters the wrong password.	2a. The system does not pass the user to next phase and shows a message of incorrect password.

3.5.2 Sign up

In this use case the system prompts the user to enter first name last name, email and password and click sign up. In response to that the system validates the details and shows an appropriate message on successful sign up. The system also gives response if the sign up gets unsuccessful and shows an appropriate massage.

Figure 3.3: Use case for Sign up

Table 4.2: Sign up

Use case ID	UC-2
Use case name	Sign up
References requirement	Requirement no.2
Actors	User
Purpose	To sign up the user
Overview	The user enters the first name last name, email, password, password and then click on sign in
Type	Primary and essential

Pre-Condition	None
Post-Condition	User successfully signed up
Normal flow	
Actors Actions	System Response
3. The user enters the first name, last name, email, password and confirms the password.	4. The system creates an account against the user inputs and pass it to next phase.
Alternative flow	
3a. The user enters the wrong password in confirm password field.	4a. The system does not pass the user to next phase and shows a message of mismatch password.

3.5.3 Search news

In this use case the system prompts the user to enter the keywords and click search. In response to that the system analysis the user's context and shows the news results in an appropriate way. The system also gives response if keywords are too short to be use as context and shows an appropriate massage.

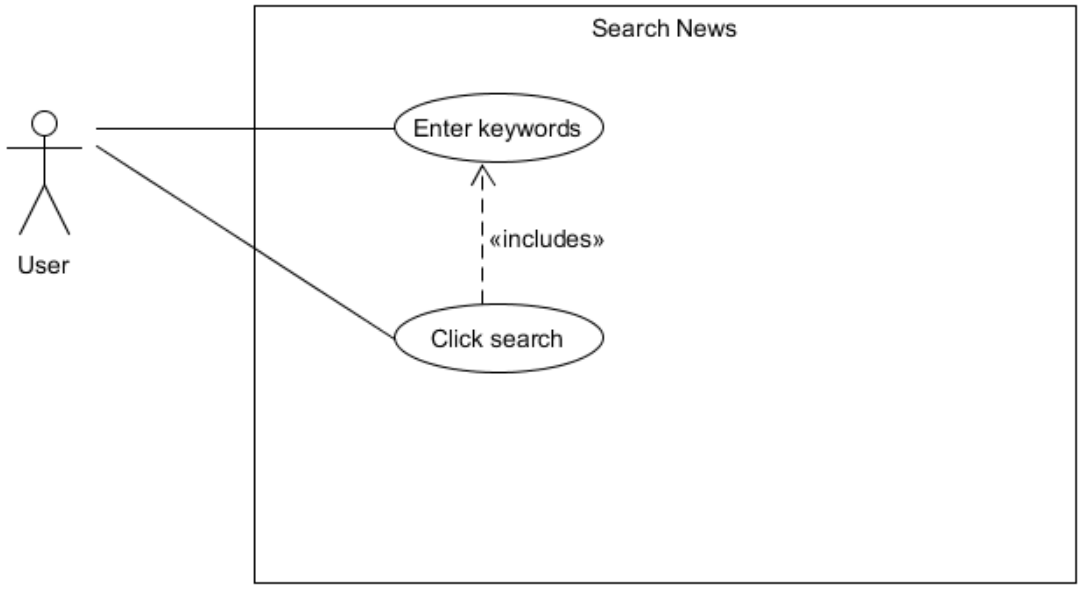


Figure 3.4: Use case for search news

Table 5.3: Search news

Use case ID	UC-3
Use case name	Search news
References requirement	Requirement no. 3
Actors	User
Purpose	To search news
Overview	The user enters the keywords and then clicks on search
Type	Primary and essential
Pre-Condition	None
Post-Condition	Search successfully performed
Normal flow	
Actors Actions	System Response
5. The user enters the keywords.	6. The system passes the keywords to find results.
Alternative flow	
5a. The user does not enter anything.	6a. The system shows a message to enter some keywords.

3.5.4 Manage history

In this use case the system shows search history of the user and allows the user to manage it by giving an option to delete the search history. The system gives response on successful deletion of search history and shows an appropriate massage. The system also gives response on unsuccessful deletion of search history and shows an appropriate massage.

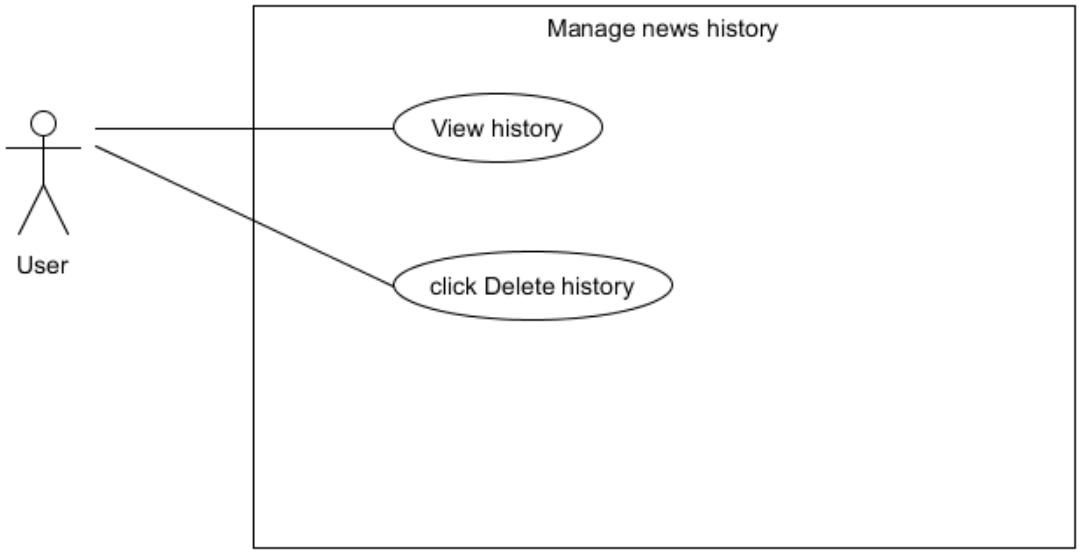


Figure 3.5: Use case for Managing search history

Table 6.4: Manage search history

Use case ID	UC-4
Use case name	Manage news history
References requirement	Requirement no. 4
Actors	User
Purpose	To manage search history of news
Overview	The user sees search history and later clicks on delete history to delete history
Type	Primary and essential
Pre-Condition	User must be signed in
Post-Condition	Search history successfully managed
Normal flow	
Actors Actions	System Response
7. The user clicks on delete history.	8. The system shows message that history deleted.
Alternative flow	

7a. There is no history to delete and user clicks delete history.	8a. The system shows message no history found to be deleted.

3.5.5 Manage interests

In this use case the system shows news interest of user and allows the user to manage it by giving an option to delete and update the news interest. The system gives response on successful deletion or updating of news interest and shows an appropriate massage. The system also gives response on unsuccessful deletion or updating of news interests and shows an appropriate massage.

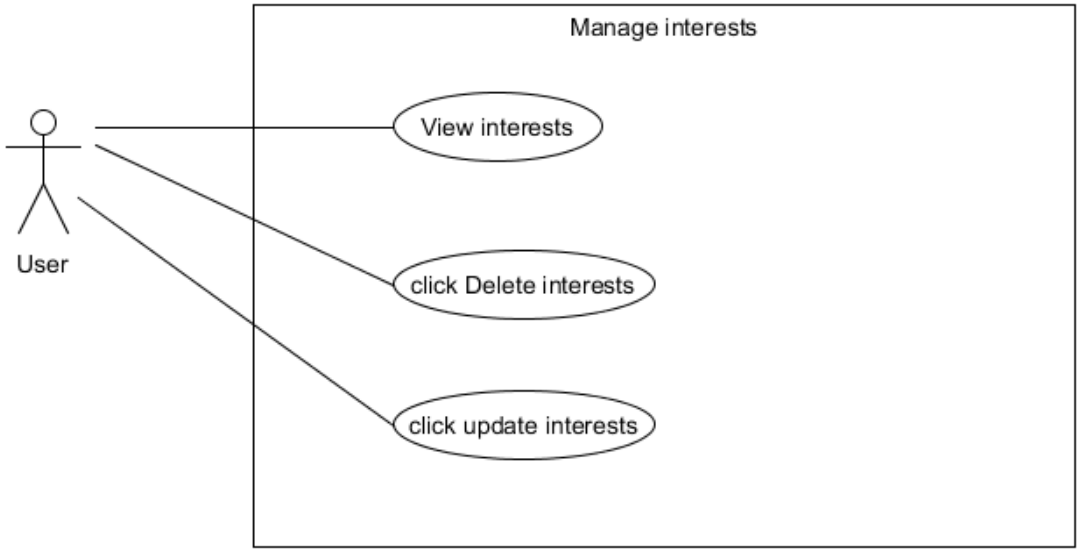


Figure 3.6: Use case for managing news interests

Table 7.5: Manage news interests

Use case ID	UC-5
Use case name	Manage interests
References requirement	Requirement no. 5
Actors	User
Purpose	To manage user new interests

Overview	The user sees the news interests and later clicks on delete or update interests to delete or update news interests
Type	Primary and essential
Pre-Condition	User must be signed in
Post-Condition	News interest successfully managed
Normal flow	
Actors Actions	System Response
9. The user sees the news interests and clicks delete interests. 11. The user sees the news interests and clicks update interests.	10. The system deletes the news interest and shows news interests again to demonstrate changes. 12. The system deletes the news interest and shows news interests again to demonstrate changes.
Alternative flow	
9a. The user does not have any news interests and clicks delete interests. 11a. The user does not have any news interests and clicks update interests.	10a. The system shows the message about no interest found. 12a.The system shows the message about no interest found.

3.5.6 Manage news results history

In this use case the system shows news result history of user and allows the user to manage it by giving an option to delete the news result history. The system gives response on successful deletion of news result history and shows an appropriate massage. The system also gives response on unsuccessful deletion of news result history and shows an appropriate massage.

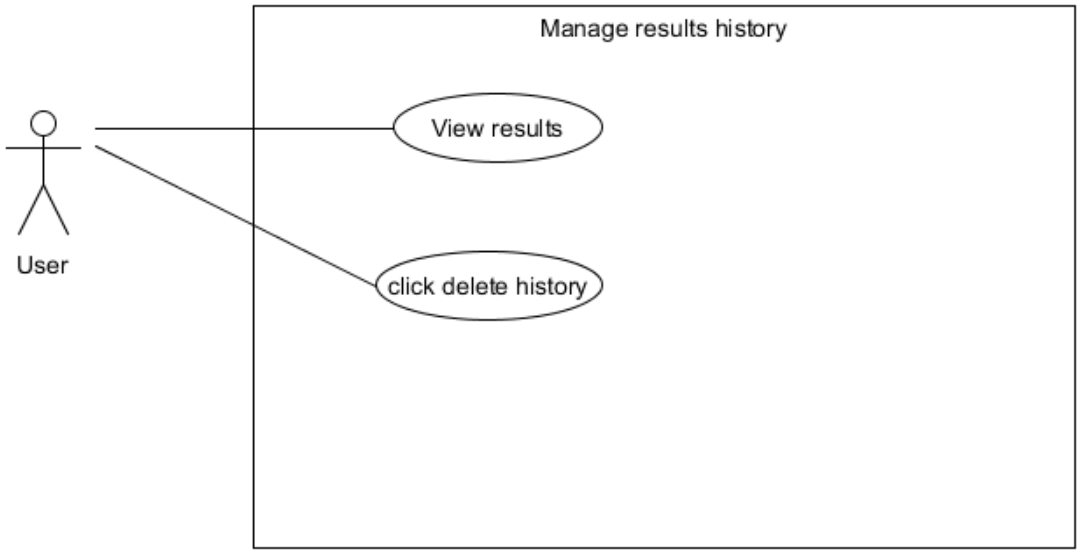


Figure 3.7: Use case for Managing news result history

Table 8.6: Manage result history

Use case ID	UC-6
Use case name	Manage result history
References requirement	Requirement no. 6
Actors	User
Purpose	To manage news result history
Overview	The users see news information results and later clicks on delete to delete news information results.
Type	Primary and essential
Pre-Condition	User must be signed in
Post-Condition	News result history successfully managed
Normal flow	
Actors Actions	System Response
13. The user clicks on delete news results.	14. The system shows message that news results are deleted.
Alternative flow	

13a. There are no news results to delete and user clicks delete history.	14a. The system shows message no news results found to be deleted.

3.5.7 Save news

In this use case the system shows news result of user and allows the user to delete the news result. The system gives response on successful saving of news result and shows an appropriate massage. The system also gives response on unsuccessful saving of news result and shows an appropriate massage.

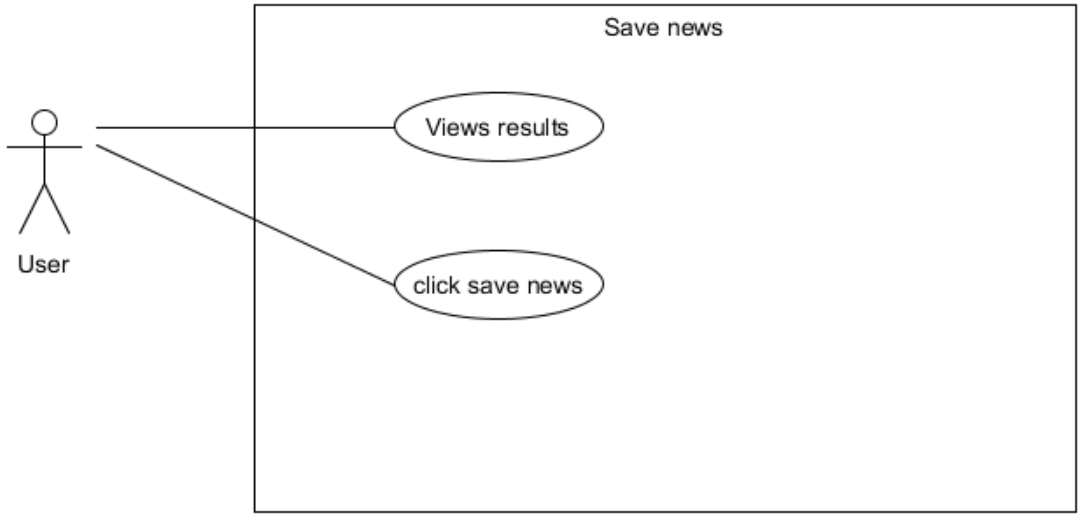


Figure 3.8: Use case for saving news results

Table 9.7: Save news

Use case ID	UC-7
Use case name	Save news
References requirement	Requirement no. 7
Actors	User
Purpose	To save news
Overview	The user sees the news information results and later clicks on save news information to save news locally
Type	Primary and essential
Pre-Condition	User must be signed in
Post-Condition	News successfully saved
Normal flow	
Actors Actions	System Response
15. The user sees the news results and clicks save news results.	16. The system saves the news results in users device.
Alternative flow	
15a. The user sees the news results and does not clicks save news results and leaves the page.	16a. The system shows a message to ask if the user wants to save the search news results.

3.6 Resource Requirements

These are the items that are used as resource in software and hardware to achieve a task. These tools are used in the project.

3.6.1 Web Scraper

Web scraper is data extraction tool to extract data from websites. It helps in extracting data from web pages, tables, html pages etc. It uses algorithms of different python frameworks, packages and libraries to scrap data. Logics can be made to scrap data of specific choice and context.

3.6.2 Tools

Here are some languages, frameworks, architectures that are used as tools to develop our system.

- **HTML:** It is a mark-up language that is used to create webpages for web.
- **CSS**: It is a Style Sheet that is used with HTML to design and decorate the webpages to look better.
- **Bootstrap:** It is framework of HTML/CSS that is used to handle the responsiveness of webpages on different screen.
- **JavaScript:** It is a scripting language that is used to handle and control the behaviour of web pages on web.
- **Python:** It is a high-level language that is used at frontend and backend as well. It is used because of its powerful frameworks.
- **Flask:** It is a micro framework of python that is used at back end to run webpages and integrate them to be ready to work with front end pages with all the integrated backend services.
- **SQLite DB:** It is built-in database in Flask that is used to manage the user data and other system data.
- **Visual Studio Code:** It is an open source and cross platform software that provides a workplace to work with above languages and frameworks.
- **Pandas:** It is a Python data analysis library that is used to manipulate data in tables.
- **Scrappy:** It is an open source web crawling framework of python that is used to extract data from web pages and web contents.
- **Beautiful Soup:** It is a parsing package for python that is used to parse web pages to extract data from HTML and XML documents.
- **PyCharm:** It is a Python IDE that provides a workplace to work with python, its frameworks and libraries.

3.7 Database Requirements

Flask SQLITE DB is used for this system because of its built-in compatibility with python and its libraries. SQLITE is a simple, relational database that is best for this system, because this does not have any complex and large amount of user and system data to store.

3.8 Project Feasibility

Moving to a web scrapping technique from general roaming and searching for collecting data would affect the efficiency of gathering information in all manners. Many people would save their time because web scrapping would allow them to gather information at one defined place rather than searching randomly and gathering it and then making it at one

place to make it suitable. Now the time has come to introduce this method of scrapping to achieve high level of efficiency. This project is the step-in right direction to lesser the frustration of common user who want to get relevant information at one place.

3.8.1 Technical Feasibility

Technically this project is more feasible because it provides more efficient way of gathering information from different sources. Contextual News Information Retrieval (CNIR) system is secure and less time consuming hence it provide more technical overheads than any other random system. The main module of our system is based on web scraping and python libraries, packages, frameworks and architecture.

3.8.2 Operational Feasibility

Operational feasibility is a measure of how well a proposed system solves the problems and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. Our system is performing all its operations accurately. All functions include web scrapping, data gathering, displaying gathered data document.

3.8.3 Legal and Ethical Feasibility

Legal feasibility is basically the analyzation of at which extent your system is legally meeting the requirements that exist for implementation. Our system is legally and ethically feasible as:

1. It does not violate any country law.
2. It is designed to fulfill people requirements of gathering data and minimize their effort on it to aid them in every possible manner.
3. Data of the user is completely secure, and it cannot be accessed without consent.

3.9 Summary

This section was about requirement specification, in which system requirements were discussed. System requirements include Interface requirements, Functional requirements, non-functional requirements, database requirements, resource requirements. Interface requirements include Software requirements and Hardware requirements. Functional requirements include must fill requirements whereas non-functional requirements include common requirements of every system. At the end Project feasibility was discussed which include technical feasibility, operational feasibility and Legal & Ethical feasibility which

means that the project is feasible within the time, budget and does not break or against the rules and regulations of the country.

CHAPTER 4

SYSTEM MODELLING

4.1 Introduction

In this section we will get the aid of some type of models for graphic representation of our system from different type of perspectives of different stakeholders. Our representation will be done by using unified modelling language (UML).

UML is a general-purpose developmental modelling language that used as an industry standard in the field of software engineering. In this chapter we will use 4+1 view to represent the architecture of our system.

4.2 System design

It is the process of theoretically defining the development and usage of the product through architecture, models, and interactions, interfaces suitable data that is or will be processed by the system.

4.3 Design Approach

Since our system is subdivided into many subsystems therefore, we would be using top-Down design approach. By using this approach, the sub programs or modules of the system would also be identified and elaborated easily. The whole system is taken as a single entity and divided into multiple sub systems and then each of these sub systems are taken as a single entity and divided further. First defining the basic and generic model then going is depth and defining each and every part of it.

The top-down approach for our system is given below.

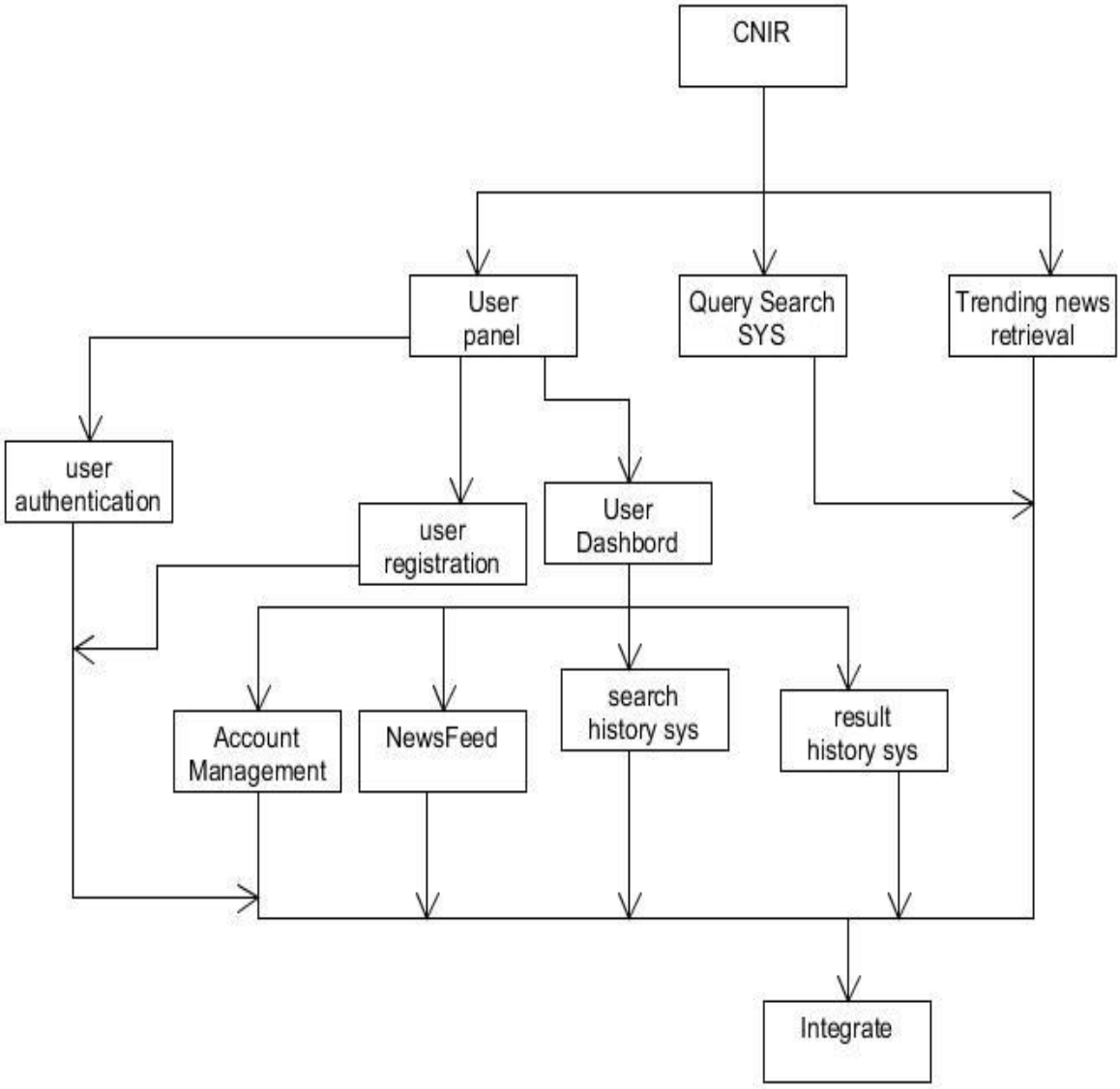


Figure 4.1: Top down approach for system

4.4 Interface design

Interface design means the visual layout and placement of the system. An interactive or responsive manner layout used by the user of the system by some elements. These elements may be menus, buttons, radio buttons, check lists etc.

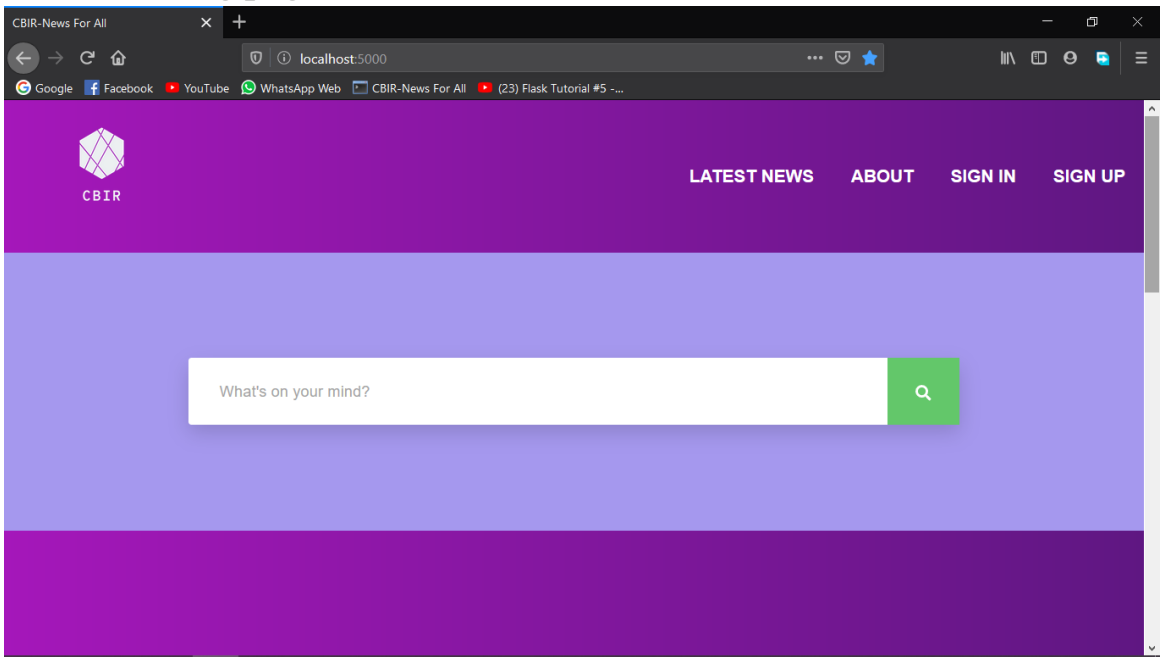
The designs must not only be attractive but also display and explain the functionalities of the system.

4.5 High fidelity Prototype

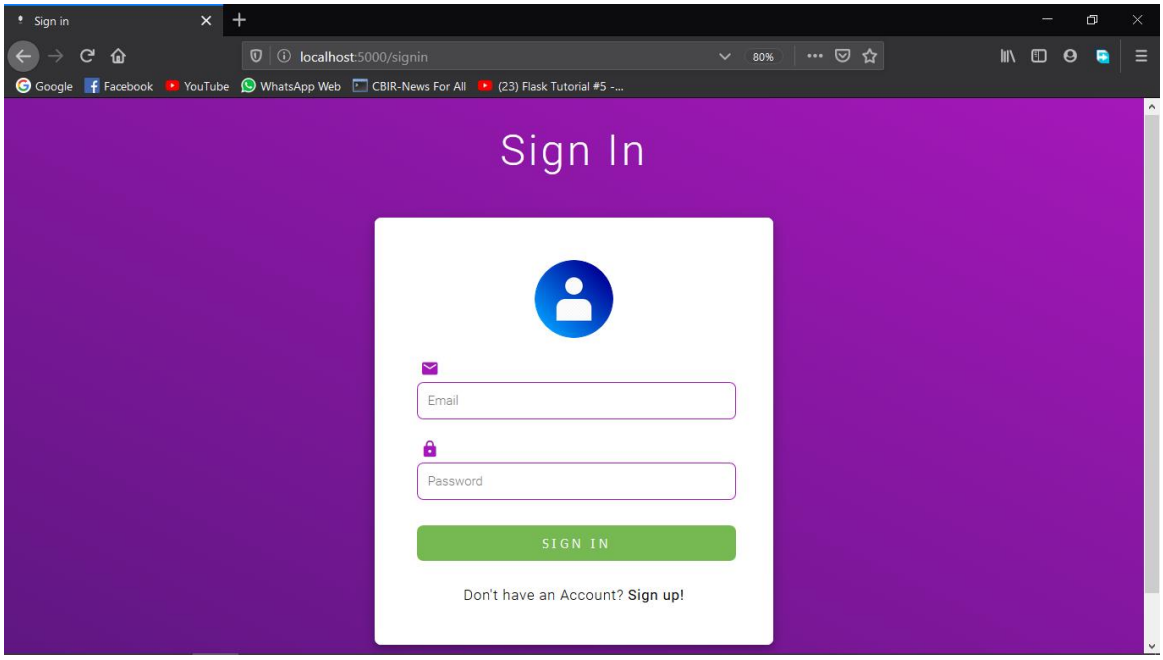
This is a type of prototype which is very much attractive and shows the functionality, operating and workflow of the system. These types of prototypes are very close to the final design of the fully functional system.

The level of detail and comprehensiveness in this type of prototype allows us to test, Analyse and question the usability of the system workflow.

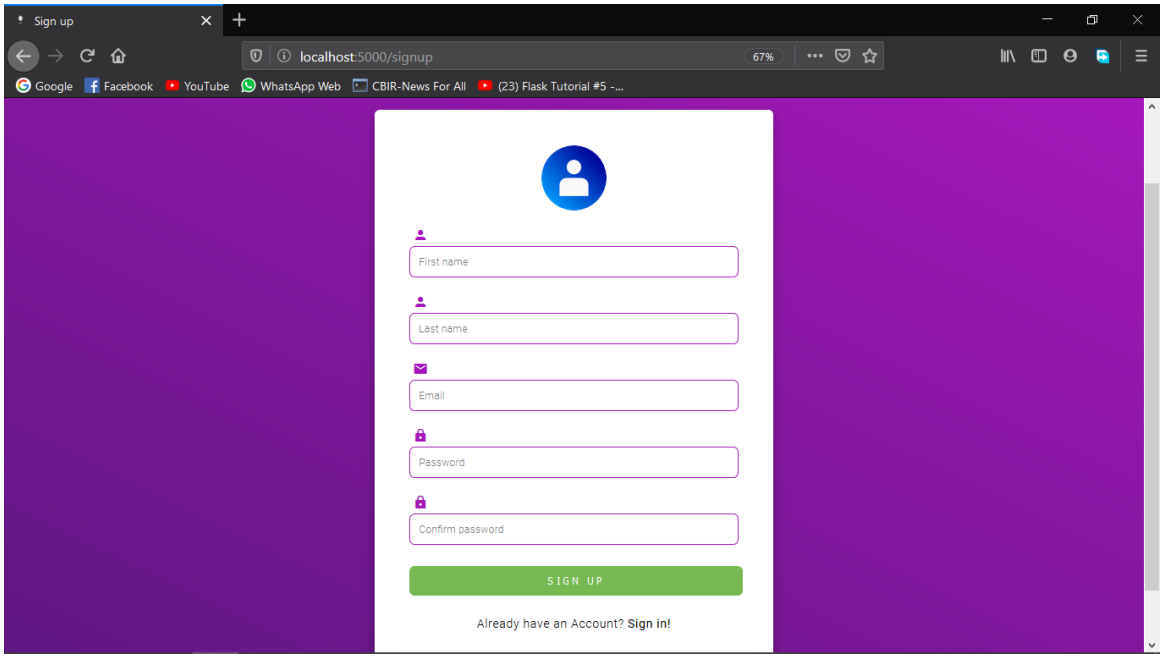
4.5.1 Landing page



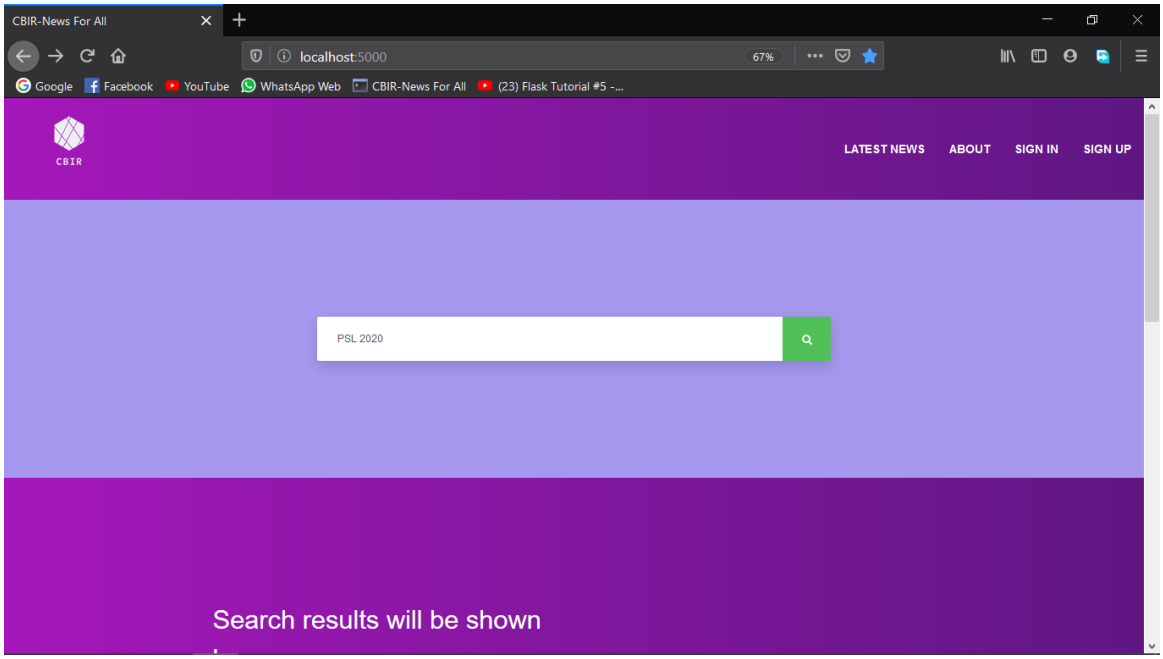
4.5.2 Signin



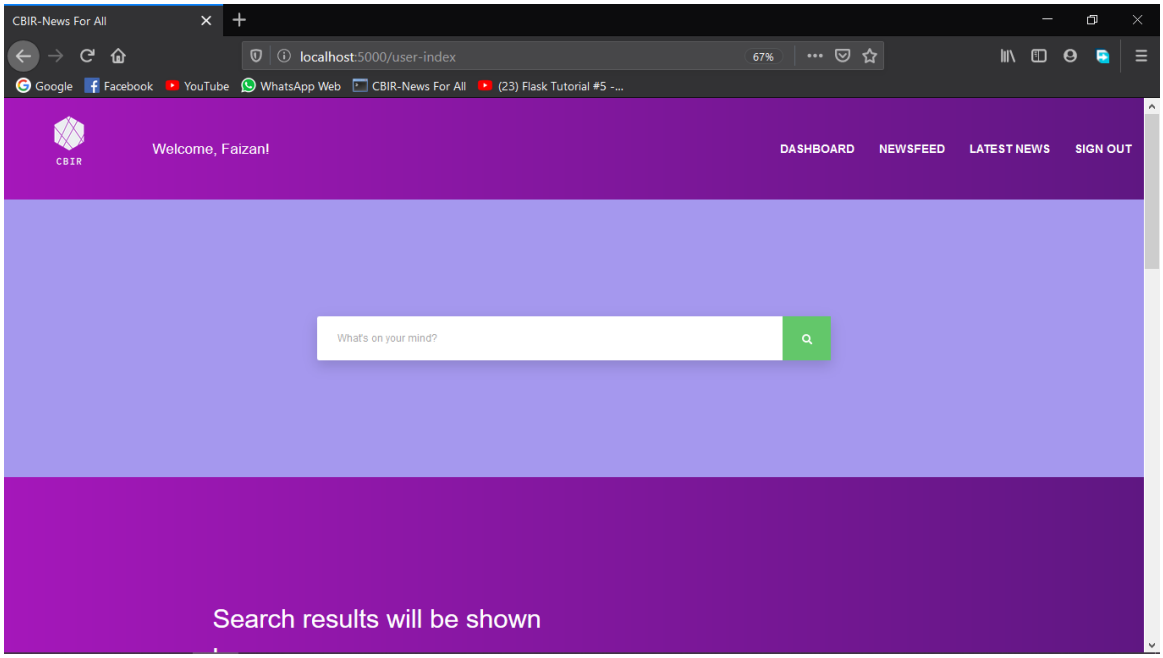
4.5.3 Signup



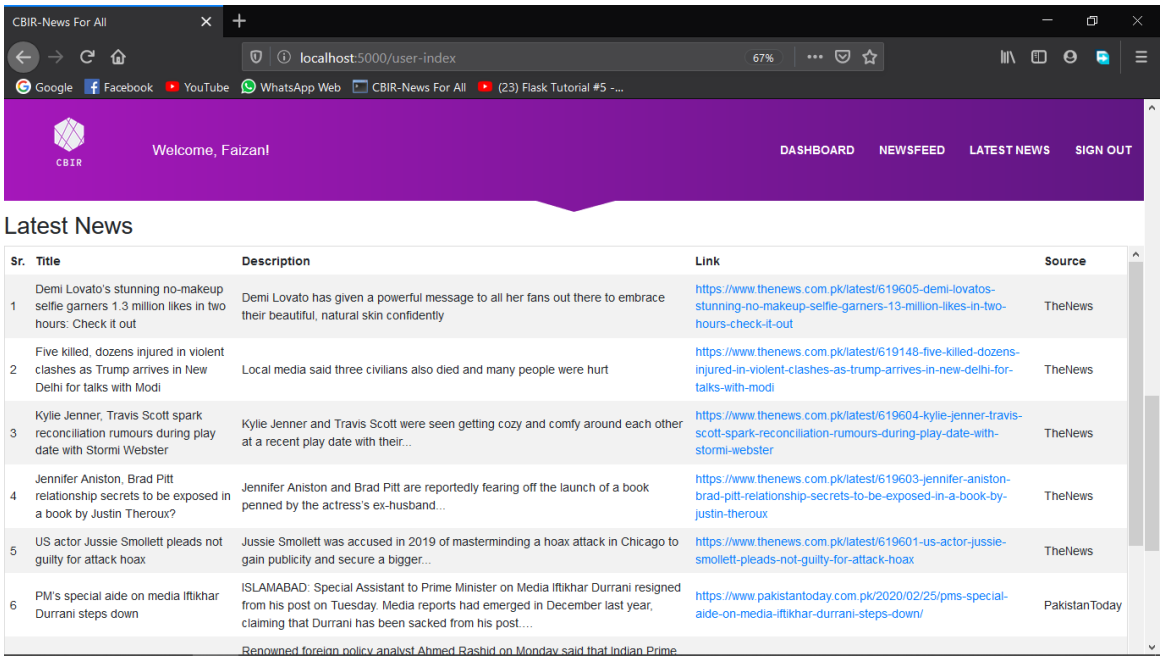
4.5.4 Search news



4.5.5 User homepage

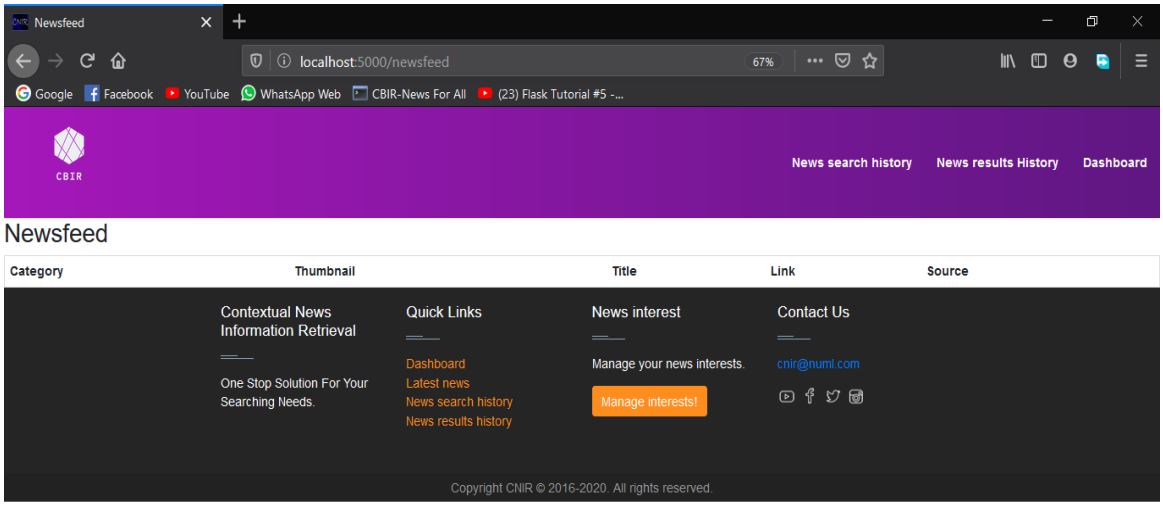


4.5.6 Latest news

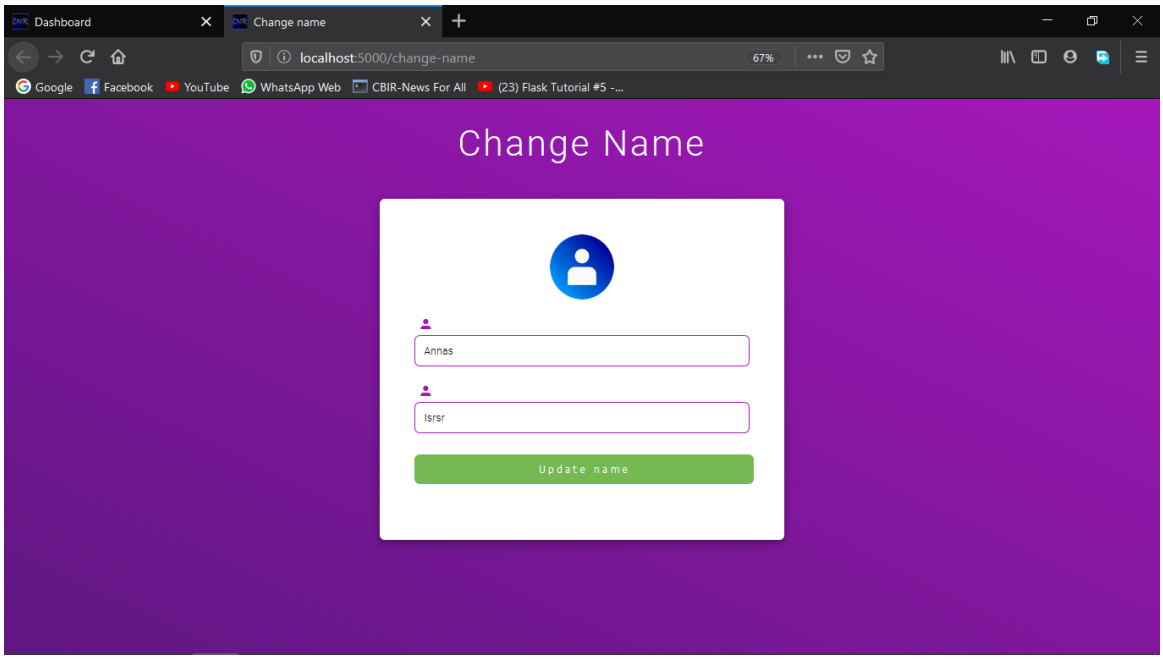


Sr.	Title	Description	Link	Source
1	Demi Lovato's stunning no-makeup selfie garners 1.3 million likes in two hours: Check it out	Demi Lovato has given a powerful message to all her fans out there to embrace their beautiful, natural skin confidently	https://www.thenews.com.pk/latest/619605-demi-lovatos-stunning-no-makeup-selfie-garners-13-million-likes-in-two-hours-check-it-out	TheNews
2	Five killed, dozens injured in violent clashes as Trump arrives in New Delhi for talks with Modi	Local media said three civilians also died and many people were hurt	https://www.thenews.com.pk/latest/619148-five-killed-dozens-injured-in-violent-clashes-as-trump-arrives-in-new-delhi-for-talks-with-modi	TheNews
3	Kylie Jenner, Travis Scott spark reconciliation rumours during play date with Stormi Webster	Kylie Jenner and Travis Scott were seen getting cozy and comfy around each other at a recent play date with their...	https://www.thenews.com.pk/latest/619604-kylie-jenner-travis-scott-spark-reconciliation-rumours-during-play-date-with-stormi-webster	TheNews
4	Jennifer Aniston, Brad Pitt relationship secrets to be exposed in a book by Justin Theroux?	Jennifer Aniston and Brad Pitt are reportedly fearing off the launch of a book penned by the actress's ex-husband...	https://www.thenews.com.pk/latest/619603-jennifer-aniston-brad-pitt-relationship-secrets-to-be-exposed-in-a-book-by-justin-theroux	TheNews
5	US actor Jussie Smollett pleads not guilty for attack hoax	Jussie Smollett was accused in 2019 of masterminding a hoax attack in Chicago to gain publicity and secure a bigger...	https://www.thenews.com.pk/latest/619601-us-actor-jussie-smollett-pleads-not-guilty-for-attack-hoax	TheNews
6	PM's special aide on media Iftikhar Durrani steps down	ISLAMABAD: Special Assistant to Prime Minister on Media Iftikhar Durrani resigned from his post on Tuesday. Media reports had emerged in December last year, claiming that Durrani has been sacked from his post....	https://www.pakistantoday.com.pk/2020/02/25/pms-special-aide-on-media-iftikhar-durrani-steps-down/	PakistanToday
		Renowned foreign policy analyst Ahmed Rashid on Monday said that Indian Prime		

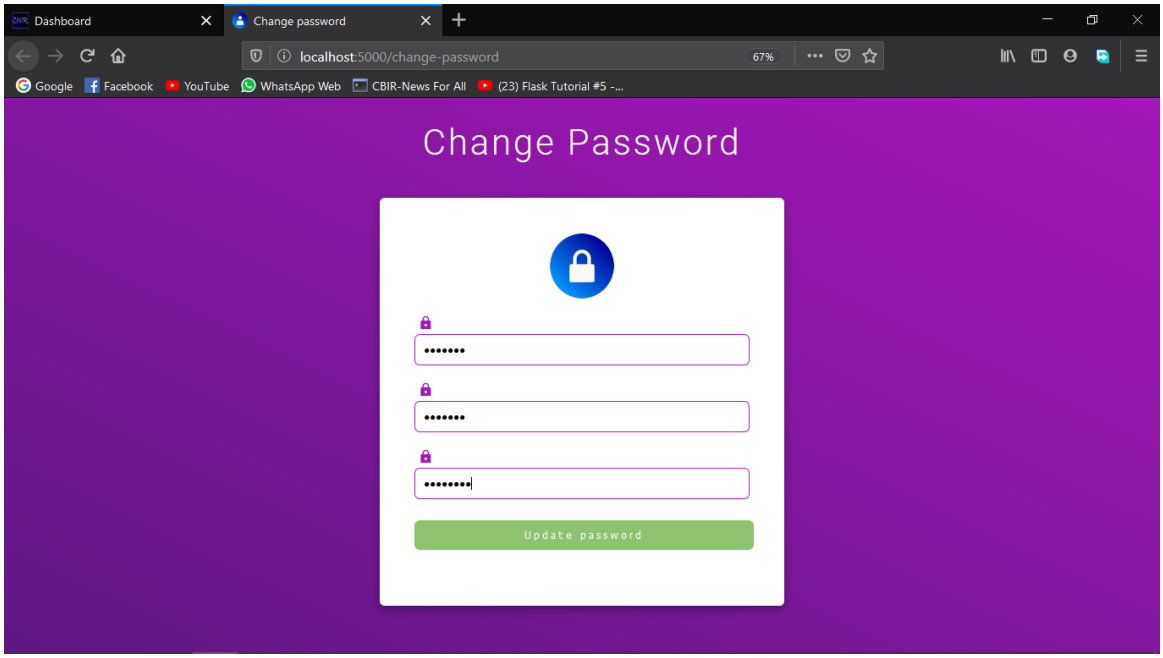
4.5.7 Newsfeed



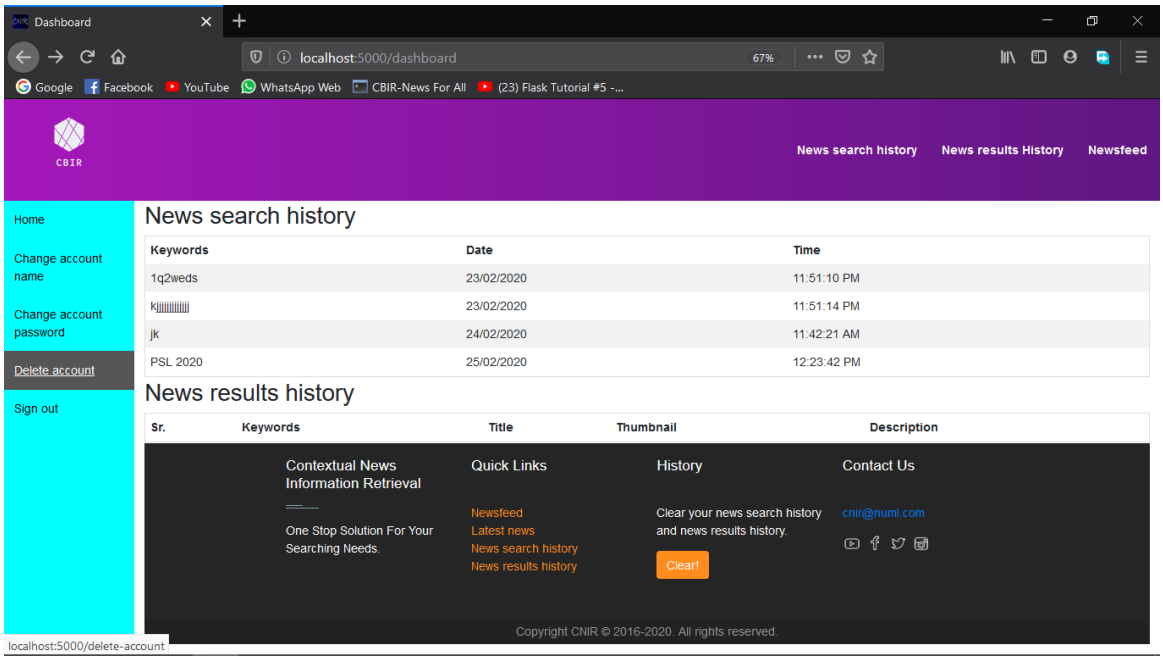
4.5.8 Change account name



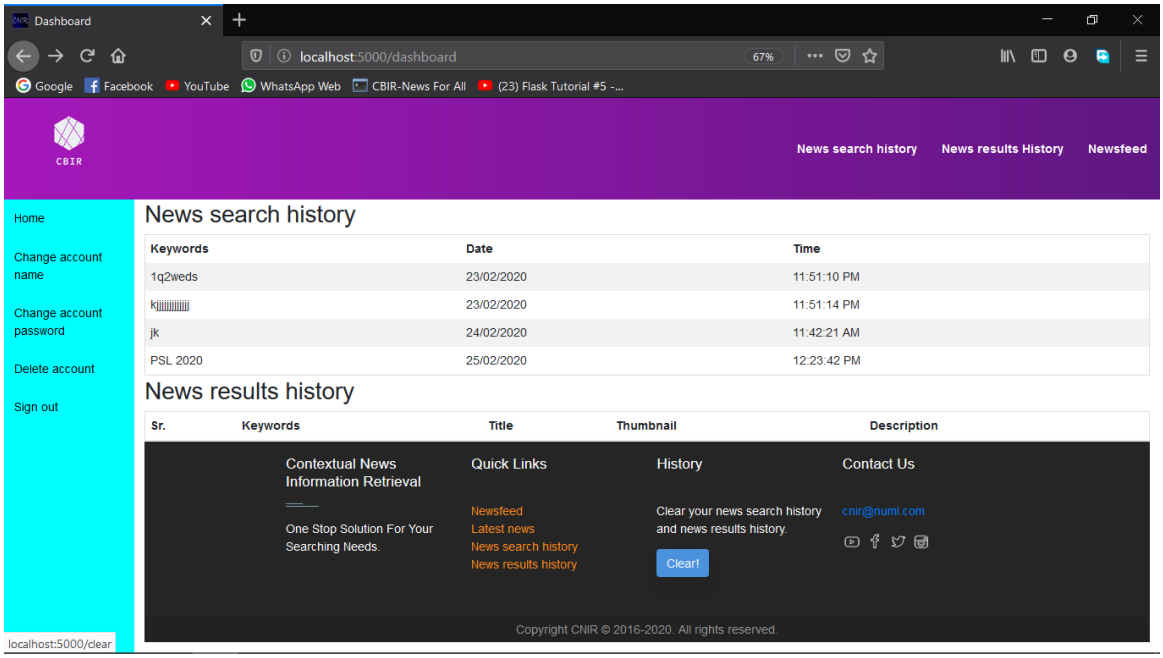
4.5.9 Change account password



4.5.10 Delete account



4.5.11 Clear search history



4.6 4+1 view model of architecture

In November 1995, Philippe Kruchten published a paper entitled: "Architectural Blueprints the “4+1” View Model of Software Architecture". The intent was to come up with a mechanism to separate the different aspects of a software system into different views of the system. Why? Because different stakeholders always have different interests in a software system. Some aspects of a system are relevant to the Developers; others are relevant to System administrators. Developers want to know about things like classes; System administrators want to know about deployment, hardware, and network configurations and don't care about classes.

We will discuss each view of our system with its diagrams.

4.7 Logical View

This view shows the logical parts of the system and contains the information about these parts. This view helps in determining the logical connectivity of different parts of the systems.

4.7.1 Class diagram

Class diagram is a structural diagram in UML which shows about the static structure of the system. This diagram includes classes, attributes and methods and relationship among different objects. The class diagram of our system is shown below.

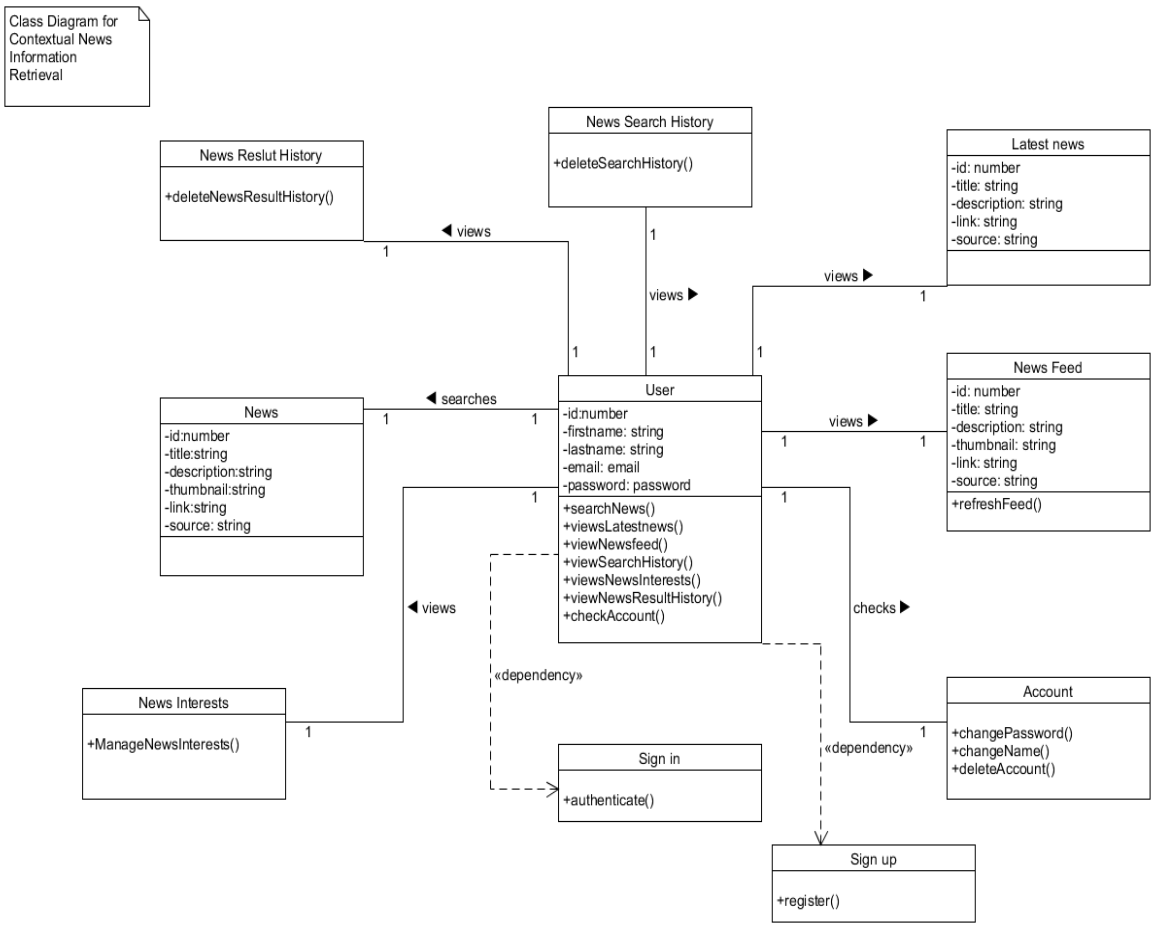


Figure 4.2: Class diagram of a system

4.8 Process View

This view shows the workflow and moving of functionality of a system and tells us the concurrent processes of the system. This view helps us in analysing the performance and availability of the system

4.8.1 Activity Diagram

The activity diagram is a behavioural diagram in the UML diagram to describe the dynamic aspects of the system.

The overall flow of system is displayed below

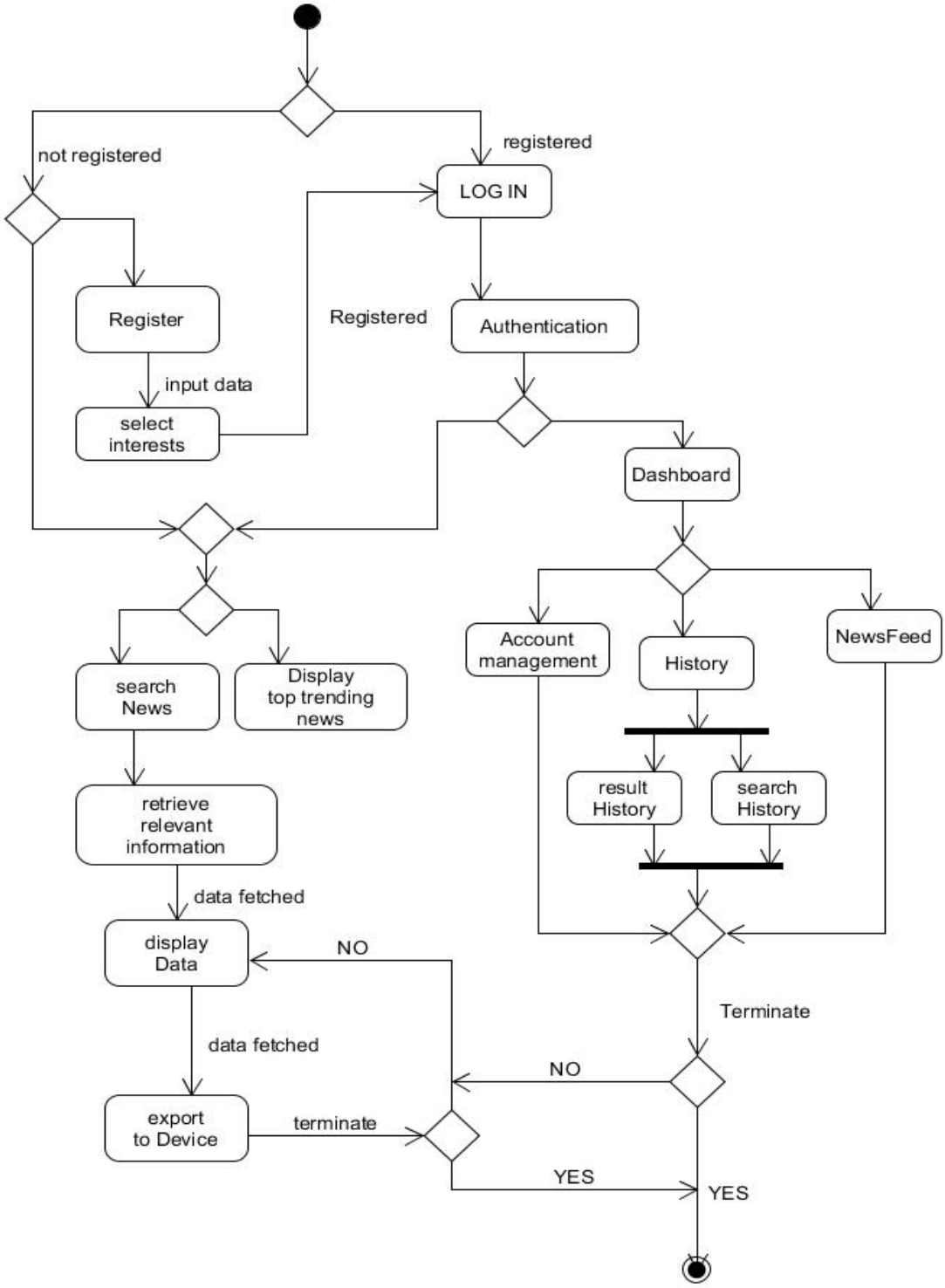


Figure 4.3: Activity diagram of a system

4.8.2 Sequence Diagram

Sequence diagrams helps us to represent the behaviour of the system in response to the interactions and input of the user. They display the functionality of each module by providing dynamic results based on the input of the user.

It is called sequence diagram because it shows the behaviour and functionality of the system and its response to the user's interactions in a sequential manner.

4.8.2.1 Sign in

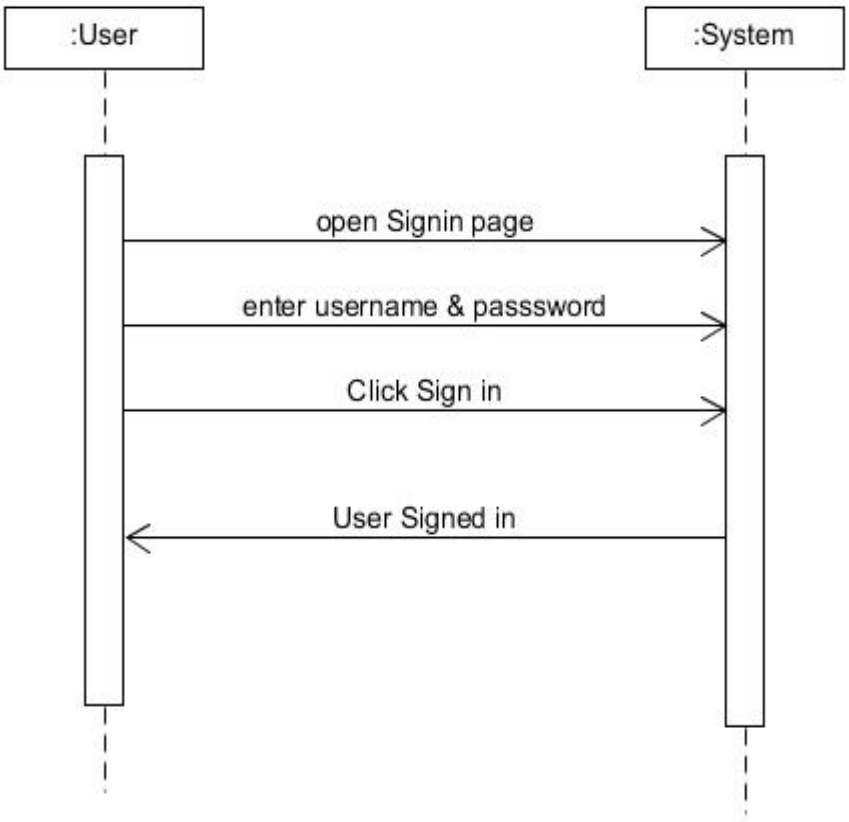


Figure 4.4: Sequence diagram for Signin

4.8.2.2 Sign up

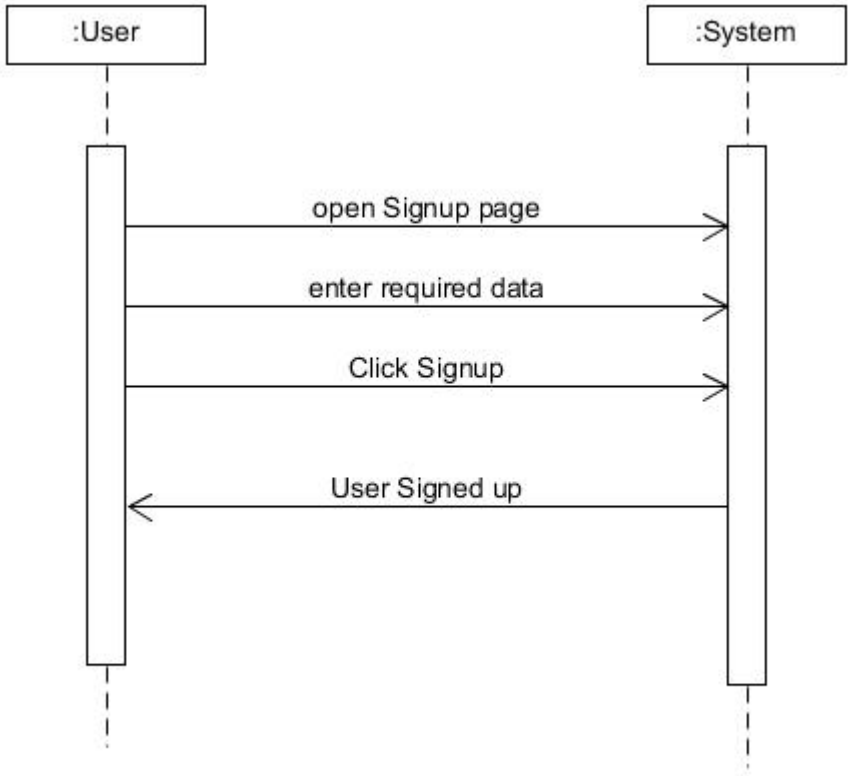


Figure 4.5: Sequence diagram for Signup

4.8.2.3 Search News

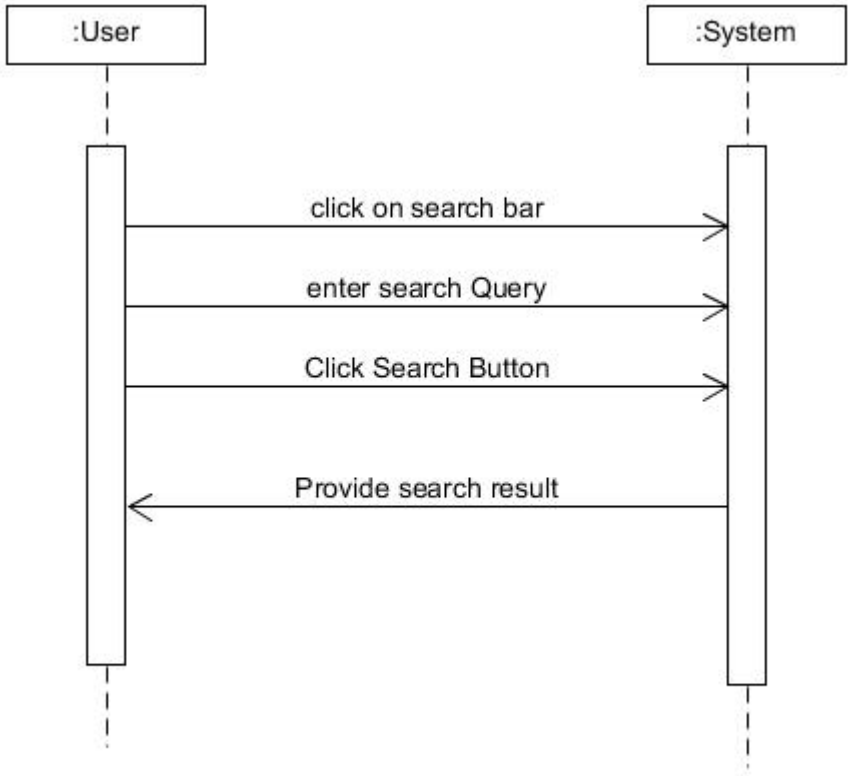


Figure 4.6: Sequence diagram for Search news

4.8.2.4 Search History

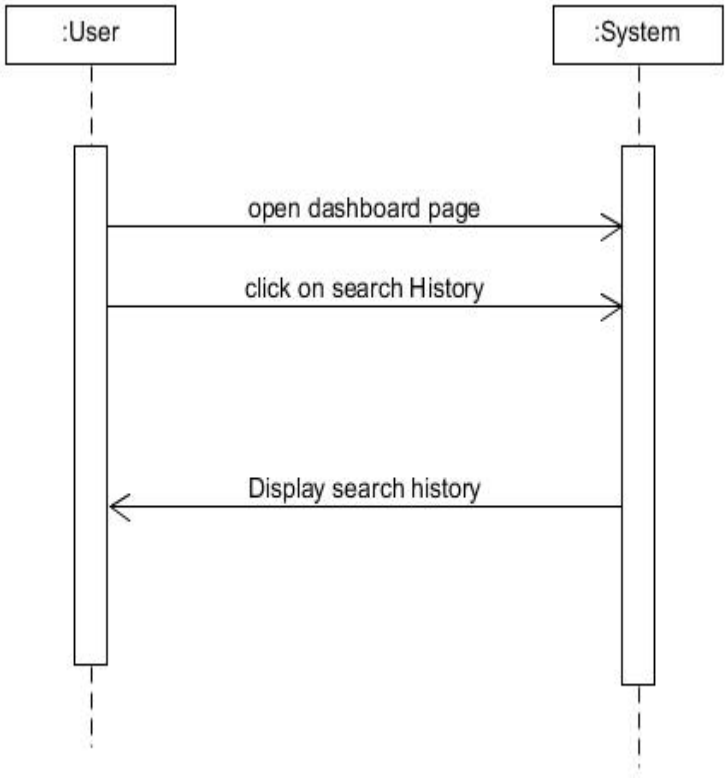


Figure 4.7: Sequence diagram for Search history

4.8.2.5 News Feed

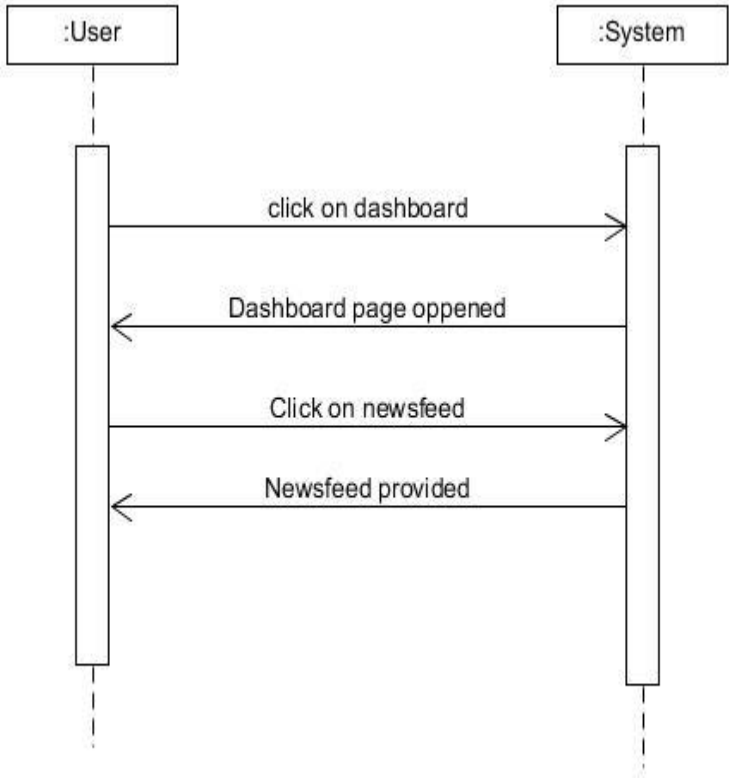


Figure 4.8: Sequence diagram for Newsfeed

4.8.2.6 Result history

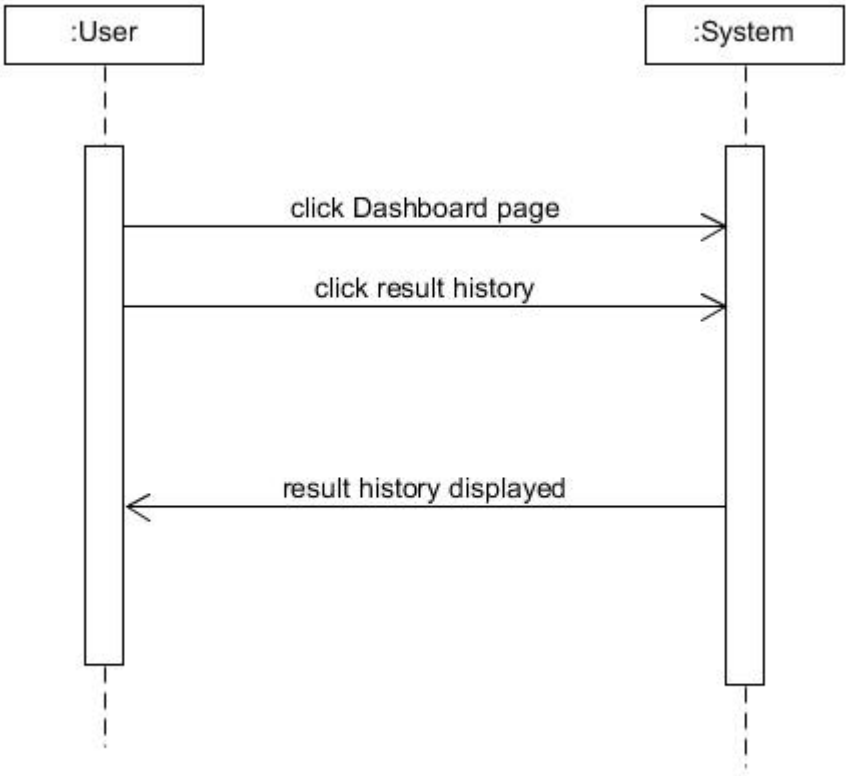


Figure 4.9: Sequence diagram for Result history

4.9 Development View

This view helps us to model and demonstrates the components, modules and subsystems of our main system.

4.9.1 Component diagram

Component diagrams are utilized in modelling the physical parts of Object-Oriented systems that are utilized for picturing, indicating, and documenting component based.

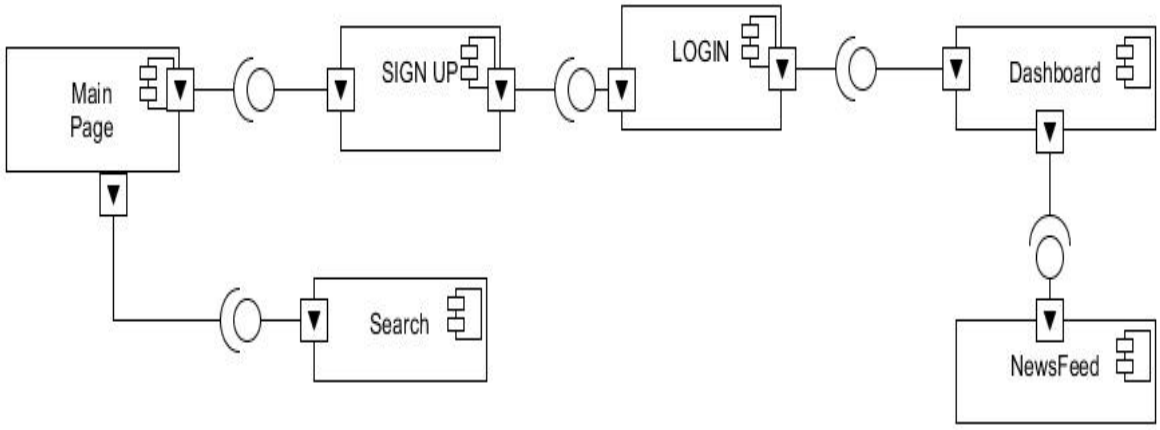


Figure 4.10: Component diagram of a system

4.10 Physical View

This type of view tells us the physical and structural deployment of the system and the type of network and physical layout this system is using and is designed.

4.10.1 Deployment diagram

This diagram shows us the physical deployment of the system. The arrangement of nodes, what type of functionality is performed by the nodes. It helps us to figure out the stability and availability of the system.

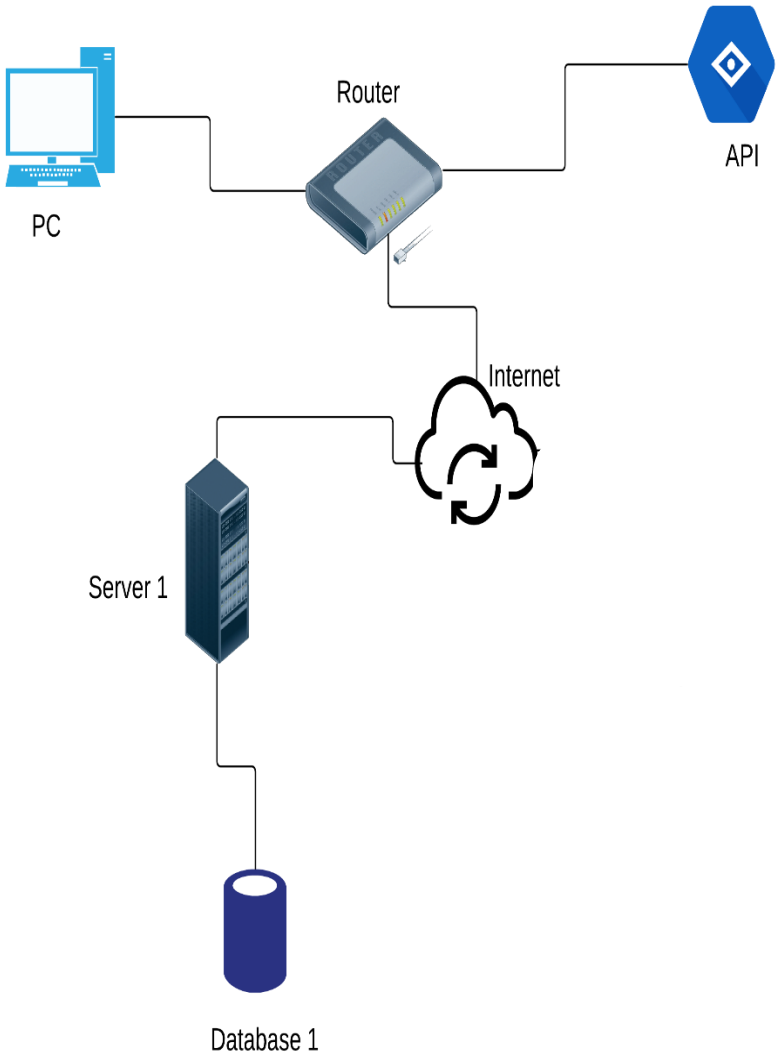


Figure 4.11: Deployment diagram of a system

4.11 The Use Case view

The use case views help us to understand the interactions with the system by the perspective of different stakeholders.

All use cases with their diagrams have been discussed in chapter 3.

4.12 Summary

In this chapter we have giving the complete modelling of our system from start to end. A high-level representation of the system has been given.

In the activity diagram we have represented the dynamic behaviour of our web-based application and all the alternates of different activities. In sequence diagrams we discussed

the different responses of the system opposing to different interactions of the user and the sequence of steps required to perform the functionality. Component diagram helps to represent the different subsystems and modules of the system to easily enhance the functionality.