**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**

**JUNE, 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, and 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 English News websites only.

**TABLE OF CONTENTS**

[CHAPTER 1 10](#_Toc42725879)

[INTRODUCTION 10](#_Toc42725880)

[1.0 Introduction 11](#_Toc42725881)

[1.1 Motivation 11](#_Toc42725882)

[1.2 Problem statement 11](#_Toc42725883)

[1.3 Goals and Objectives 12](#_Toc42725884)

[1.4 Scope of the study 12](#_Toc42725885)

[1.5 Process model 13](#_Toc42725886)

[1.5.1 Requirement gathering 13](#_Toc42725887)

[1.5.2 Work breakdown and decomposition 13](#_Toc42725888)

[1.5.3 Designing 13](#_Toc42725889)

[1.5.4 Coding 13](#_Toc42725890)

[1.5.5 Integration 14](#_Toc42725891)

[1.5.6 Testing 14](#_Toc42725892)

[1.5.7 Why incremental model 14](#_Toc42725893)

[1.6 Nature of the project 14](#_Toc42725894)

[1.7 Overview of the Report 14](#_Toc42725895)

[CHAPTER 2 15](#_Toc42725896)

[BACKGROUND AND EXISITING WORK 15](#_Toc42725897)

[2.0 Introduction 16](#_Toc42725898)

[2.1 Explanation of Important Constructs of the Application Domain 16](#_Toc42725899)

[2.2 Existing Systems 16](#_Toc42725900)

[2.2.1 Google News 16](#_Toc42725901)

[2.2.2 ABC News 16](#_Toc42725902)

[2.3 Comparison of Existing Systems 16](#_Toc42725903)

[2.4 Summary 17](#_Toc42725904)

[CHAPTER 3 18](#_Toc42725905)

[REQUIREMENTS SPECIFICATION 18](#_Toc42725906)

[3.0 Introduction 19](#_Toc42725907)

[3.1 Interface Requirements 19](#_Toc42725908)

[3.1.1 Software Interface Requirements 19](#_Toc42725909)

[3.1.2 Hardware Interface Requirements 19](#_Toc42725910)

[3.2 Functional Requirements 20](#_Toc42725911)

[3.2.1 User Registration 20](#_Toc42725912)

[3.2.2 News Search 20](#_Toc42725913)

[3.2.3 Display of Information 20](#_Toc42725914)

[3.2.4 Latest news 20](#_Toc42725915)

[3.2.5 Newsfeed 20](#_Toc42725916)

[3.2.6 Search History 21](#_Toc42725917)

[3.2.7 News Interests 21](#_Toc42725918)

[3.2.8 Save information on local machine 21](#_Toc42725919)

[3.3 Non-Functional Requirements 21](#_Toc42725920)

[3.3.1 Security 21](#_Toc42725921)

[3.3.2 Reliability 21](#_Toc42725922)

[3.3.3 Availability 21](#_Toc42725923)

[3.3.4 Maintainability 22](#_Toc42725924)

[3.3.5 Portability 22](#_Toc42725925)

[3.4 Use Case Model 22](#_Toc42725926)

[3.5 Use Cases 23](#_Toc42725927)

[3.5.1 Sign in 23](#_Toc42725928)

[3.5.2 Sign up 25](#_Toc42725929)

[3.5.3 Search news 26](#_Toc42725930)

[3.5.4 View search history 28](#_Toc42725931)

[3.5.5 Select news interests 29](#_Toc42725932)

[3.5.6 Save news 30](#_Toc42725933)

[3.5.7 Latest news 32](#_Toc42725934)

[3.5.8 Newsfeed 33](#_Toc42725935)

[3.6 Resource Requirements 34](#_Toc42725936)

[3.6.1 Web Scraper 34](#_Toc42725937)

[3.6.2 Tools 34](#_Toc42725938)

[3.7 Database Requirements 35](#_Toc42725939)

[3.8 Project Feasibility 35](#_Toc42725940)

[3.8.1 Technical Feasibility 35](#_Toc42725941)

[3.8.2 Operational Feasibility 36](#_Toc42725942)

[3.8.3 Legal and Ethical Feasibility 36](#_Toc42725943)

[3.9 Summary 36](#_Toc42725944)

[CHAPTER 4 37](#_Toc42725945)

[SYSTEM MODELLING 37](#_Toc42725946)

[4.0 Introduction 38](#_Toc42725947)

[4.1 System design 38](#_Toc42725948)

[4.2 Design Approach 38](#_Toc42725949)

[4.3 Interface design 39](#_Toc42725950)

[4.3.1 High fidelity Prototype 40](#_Toc42725951)

[4.4 4+1 view Model of Architecture 45](#_Toc42725952)

[4.4.1 Logical View 45](#_Toc42725953)

[4.4.2 Process View 47](#_Toc42725954)

[4.4.3 Development View 57](#_Toc42725955)

[4.4.4 Physical View 58](#_Toc42725956)

[4.5 Summary 59](#_Toc42725957)

[CHAPTER 5 60](#_Toc42725958)

[IMPLEMENTATION 60](#_Toc42725959)

[5.0 Introduction 61](#_Toc42725960)

[5.1 Modules of Project 61](#_Toc42725961)

[5.1.1 User Registration 61](#_Toc42725962)

[5.1.2 News Search 61](#_Toc42725963)

[5.1.3 Save News 61](#_Toc42725964)

[5.1.4 Search History 61](#_Toc42725965)

[5.2 Tools and Technology 62](#_Toc42725966)

[5.2.1 Flask 62](#_Toc42725967)

[5.2.2 Tesseract-OCR 62](#_Toc42725968)

[5.2.3 Pandas 62](#_Toc42725969)

[5.2.4 Beautiful Soup 62](#_Toc42725970)

[5.2.5 Requests 62](#_Toc42725971)

[5.2.6 NumPy 63](#_Toc42725972)

[5.2.7 NLTK 63](#_Toc42725973)

[5.2.8 SKLearn TFIDF Vectorizer 63](#_Toc42725974)

[5.2.9 SKLearn Cosine Similarity 63](#_Toc42725975)

[5.3 Summary 63](#_Toc42725976)

[CHAPTER 6 64](#_Toc42725977)

[TESTING, ANALYSES AND VALIDATION 64](#_Toc42725978)

[6.0 Introduction 65](#_Toc42725979)

[6.1 Test Bed 65](#_Toc42725980)

[6.2 System Test Case 66](#_Toc42725981)

[6.3 Test Cases 66](#_Toc42725982)

[6.3.1 Sign in 66](#_Toc42725983)

[6.3.2 Sign up 68](#_Toc42725984)

[6.3.3 Search news 69](#_Toc42725985)

[6.3.4 Search history 70](#_Toc42725986)

[6.3.5 Select interests 71](#_Toc42725987)

[6.3.6 Save news 72](#_Toc42725988)

[6.3.7 Latest news 73](#_Toc42725989)

[6.3.8 Newsfeed 75](#_Toc42725990)

[6.4 Summary 76](#_Toc42725991)

[CHAPTER 7 77](#_Toc42725992)

[CONCLUSION AND FUTURE WORK 77](#_Toc42725993)

[7.0 Overview 78](#_Toc42725994)

[7.1 System Overview 78](#_Toc42725995)

[7.2 Milestones Achieved 78](#_Toc42725996)

[7.2.1 User Registration 78](#_Toc42725997)

[7.2.2 News Search 78](#_Toc42725998)

[7.2.3 Display of Information 79](#_Toc42725999)

[7.2.4 Latest news 79](#_Toc42726000)

[7.2.5 Newsfeed 79](#_Toc42726001)

[7.2.6 Search History 79](#_Toc42726002)

[7.2.7 News Interests 79](#_Toc42726003)

[7.2.8 Save Information on local machine 79](#_Toc42726004)

[7.3 Limitations 79](#_Toc42726005)

[7.4 Future Work 80](#_Toc42726006)

[7.4 Summary 80](#_Toc42726007)

**LIST OF FIGURES**

[Figure 3.1: Use case diagram for Contextual News Information Retireval 24](#_Toc32870342)

[Figure 3.2: Signin 25](#_Toc32870343)

[Figure 3.3: Signup 27](#_Toc32870344)

[Figure 3.4: Search news 28](#_Toc32870345)

[Figure 3.5: Manage search history 29](#_Toc32870346)

[Figure 3.6: Manage news interests 31](#_Toc32870347)

[Figure 3.7: Manage news results 33](#_Toc32870348)

[Figure 3.8: Save news 34](#_Toc32870349)

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](#_Toc32870342)

[Table 2.1: Comparison of existing and proposed system 19](#_Toc32870343)

[Table 3.1: Sign in 25](#_Toc32870344)

[Table 3.2: Sign up 27](#_Toc32870345)

[Table 3.3: Search news 29](#_Toc32870346)

[Table 3.4: Manage search history 30](#_Toc32870347)

[Table 3.5: Manage news interests 31](#_Toc32870348)

[Table 3.6: Manage result history 33](#_Toc32870349)

[Table 3.7: Save news 35](#_Toc32870350)

# CHAPTER 1

# INTRODUCTION

# 1.0 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. To look for some desired facts or news a person needs to go through multiple sources online in order to get their required information.

So, 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.

# 1.1 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 to get the relevant news of their interest and secondly, to provide news from authentic sources.

# 1.2 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, 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.

# 1.3 Goals and Objectives

The main goals that serve this project for users are to register user, search news, show newsfeed, show latest news, save news information, select user interests from predefined interest (Sports, Entertainment, Business, Pakistan), and show news search history of the users.

This project will assist those that are linked with news research including most of analysts, anchors, reporters. This project will retrieve valid and authentic news information from sites, articles and blogs available on internet according to users context. This project will then manage the gathered information by displaying information in an appropriate way to support the user in an efficient way.

# 1.4 Scope of the study

Table 1.1: Scope of the study

|  |  |
| --- | --- |
| **Scope of the study** | |
| **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 for Latest news and Newsfeed. |
| **Assumption** | A fast & stable internet connection is available, a working web browser. |
| **Stake holders** | Users, 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

### Requirement gathering

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

### 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 data downloadable
* Providing references of the extracted news information

### 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.

### 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.

### Integration

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

### 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.

### 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 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.

# 1.7 Overview of the Report

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

# 2.0 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 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. 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.

# Explanation of Important Constructs of the Application Domain

#### Query

When a user comes to an information retrieval system, he has some question in mind. Those questions are then transformed into a query which he writes and expects the system to answer that.

#### Information Provision

As the name suggests, the system is expected to provide the information that the user has asked in his query. Web Scraping refers to getting data from webpages by writing scripts. These scripts will automatically get the data from the page’s html when executed. Web scraping is used for providing the users with their desired information.

#### Information Representation

Information provision is another important construct. Information when not presented in a suitable format, has no impact on its reader.

#### Document Similarity

When providing information according to the context of user, it is particularly important to make sure the provided information relates to the entered query. This is made sure by finding the document similarity, which means by finding out the similarity between the entered query and information. The content which has the maximum similarity with reference to the query is then provided to the users.

# 2.2 Existing Systems

Following are some existing systems.

### 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.

### 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 it does not analyse pictorial information and E-Papers

# 2.3 Comparison of Existing Systems

A comparison the two existing systems is given below.

Table 2.1: Comparison of existing

|  |  |  |
| --- | --- | --- |
| **Features** | **Google News** | **ABC News** |
| Text Extraction | No | No |
| Text Extraction from Pictures | No | No |
| Analyzation of user’s script | No | No |
| Download Option | No | Yes |
| Searching Option | Yes | Yes |

# 2.4 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

# 3.0 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.

# 3.1 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 make the system communicating with all its components. As the system is a web application so the interface requirements are subdivided into further categories.

### Software Interface Requirements

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

### Hardware Interface 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.

# 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.

### User Registration

1. System should be able to register a new user (sign up) by prompting his name, email and a password and authenticate registered user (sign in) by prompting the email and a password.
2. System should be able to give control to users for updating their account details.

### News Search

1. System should enable the user to search news of his choice.
2. System must be able to analyse the users input to determine the users context.
3. System should be able extract news information available online and from online news images.

### 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.

### Latest news

System should be able extract latest news information from following news websites:

* The Nation
* Dawn
* Pakistan Today
* Daily Pakistan
* Express Tribune

### Newsfeed

System should be able to show newsfeed based on the search history and interest of registered users from these news websites:

* The Nation
* Dawn
* Pakistan Today
* Daily Pakistan
* Express Tribune

### Search History

System should be able to maintain search history of the registered users.

### News Interests

System should enable the users to select news interests from these categories:

* Sports
* Entertainment
* Business
* Pakistan

### Save information on local machine

System should be able to give control to user by providing option to save and export the news information in users device.

# 3.3 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.

### Security

Server end data shall be accessible to developers/administrators. The tools and technologies used to develop this project will also provide built-in security by itself.

### 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.

### 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.

### 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.

### 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.

# 3.4 Use Case Model

Use case is a model that is designed 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 guest user that has many sub features related to website. The home page has a search bar from where users can search 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 guest user features there are some essential features that are meant to be used and performed by registered users only. These features include user registration, newsfeed, search news, saving of news, news interest selection, viewing search history. These interesting features makeup the system named as Contextual News Information Retrieval (CNIR).

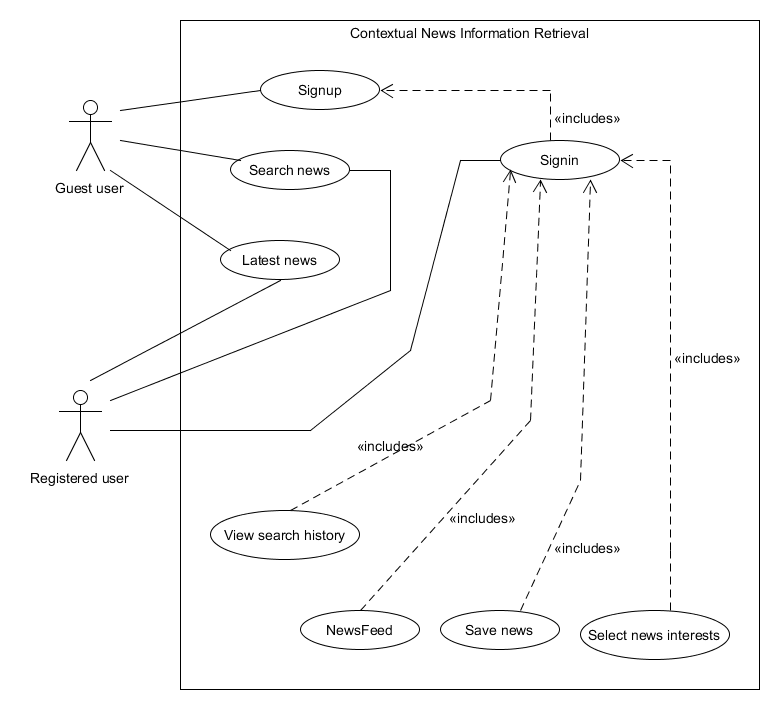


Figure 3.1: Use Case Diagram for Contextual News Information Retrieval

# 3.5 Use Cases

Below are the main and primary use cases that serve the purpose of use goals:

### 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.

A close up of a map

Description automatically generated

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 | Registered user |
| Purpose | To sign in registered 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. | 1. 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. |

### Sign up

In this use case the system prompts the guest 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.

A close up of a map

Description automatically generated

Figure 3.3: Use case for Sign up

Table 3.2: Sign up

|  |  |
| --- | --- |
| Use case ID | UC-2 |
| Use case name | Sign up |
| References requirement | Requirement no.2 |
| Actors | Guest user |
| Purpose | To sign up the guest 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 |
| 1. The user enters the first name, last name, email, password and confirms the password. | 1. The system creates an account against the user inputs and pass it to next phase. |
| Alternative flow | |
| 1a. The user enters the wrong password in confirm password field. | 2a. The system does not pass the user to next phase and shows a message of mismatch password. |

### Search news

In this use case the system prompts the user to enter the keywords and click search. In response to that the system analyses the users 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.

A close up of a logo

Description automatically generated

Figure 3.4: Use case for search news

Table 3.3: Search news

|  |  |
| --- | --- |
| Use case ID | UC-3 |
| Use case name | Search news |
| References requirement | Requirement no. 3 |
| Actors | Guest User, registered 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 |
| 1. The user enters the keywords. | 1. The system passes the keywords to find results. |
| Alternative flow | |
| 1a. The user does not enter anything. | 2a. The system shows a message to enter some keywords. |

### View search history

In this use case the system shows search history of the registered user and allows the registered user to manage it by giving an option to clear 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.

A close up of a logo

Description automatically generated

Figure 3.5: Use case for Managing search history

Table 3.4: Manage search history

|  |  |
| --- | --- |
| Use case ID | UC-4 |
| Use case name | View search history |
| References requirement | Requirement no. 4 |
| Actors | Registered User |
| Purpose | To show search history to registered user |
| 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 |
| 1. The user clicks on delete history. | 1. The system shows message that search history deleted. |
| Alternative flow | |
| 1a. There is no history to delete and user clicks delete history. | 2a. The system shows message no history found to be deleted. |

### Select news interests

In this use case the system shows predefined news interest to registered user and allows the user to update the news interest by choosing any one of them and save them. The system gives response updating of news interest and shows an appropriate massage. The system also gives response on updating of news interests and shows an appropriate massage.

A close up of text on a white background

Description automatically generated

Figure 3.6: Use case for managing news interests

Table 3.5: Manage news interests

|  |  |
| --- | --- |
| Use case ID | UC-5 |
| Use case name | Select news interests |
| References requirement | Requirement no. 5 |
| Actors | Registered user |
| Purpose | To choose and save news interest of registered user |
| Overview | The user sees the news interests and choose the predefined interests and clicks save |
| Type | Primary and essential |
| Pre-Condition | User must be signed in |
| Post-Condition | News interest successfully selected |
| Normal flow | |
| Actors Actions | System Response |
| 1. The user sees the news interests to chooses interests and then the user clicks save to save news interests. | 1. The system shows news interests saves news interests selected by user. |
| Alternative flow | |
| 1.a The user does not select any news interests and clicks update interests. | 2a. The system does not save news interests |

### Save news

In this use case the system gives option to save news in users local machine to registered users. The system shows response by creating a file containing news result and make it downloadable.

A close up of a logo

Description automatically generated

Figure 3.8: Use case for saving news results

Table 3.7: Save news

|  |  |
| --- | --- |
| Use case ID | UC-6 |
| Use case name | Save news |
| References requirement | Requirement no. 6 |
| Actors | Registered user |
| Purpose | To save news |
| Overview | The user sees the news information results and later clicks on save news to download locally |
| Type | Primary and essential |
| Pre-Condition | User must be signed in |
| Post-Condition | News successfully saved |
| Normal flow | |
| Actors Actions | System Response |
| 1. The user sees the news results and clicks save news. | 1. The system saves the news results in users device. |
| Alternative flow | |
| 1a. The user sees the news results and does not clicks save news results and leaves the page. | 2a. The system shows a message to ask if the user wants to save the search news results. |

### Latest news

This news case will show latest news to the users. The system will respond by showing the latest news and will provide appropriate feedback on successful result of latest news.

A close up of text on a white background

Description automatically generated

|  |  |
| --- | --- |
| Use case ID | UC-7 |
| Use case name | Latest news |
| References requirement | Requirement no. 7 |
| Actors | Registered user, guest user |
| Purpose | To show latest news |
| Overview | The user clicks on latest news and then system will show latest news to the respective user. |
| Type | Primary and essential |
| Pre-Condition | Users must be active |
| Post-Condition | Latest news successfully shown |
| Normal flow | |
| Actors Actions | System Response |
| 1. The user clicks on latest news and see the latest news | 1. The system shows the latest news to the users. |
| Alternative flow | |
| 1a. The user clicks on latest news button but does not see latest news | 2a. The system shows the message about unavailability of latest news. |

### Newsfeed

This news case will show newsfeed to the users. The system will respond by showing the newsfeed and will provide appropriate feedback on successful result of newsfeed.

A close up of text on a white background

Description automatically generated

|  |  |
| --- | --- |
| Use case ID | UC-8 |
| Use case name | Newsfeed |
| References requirement | Requirement no. 8 |
| Actors | Registered user |
| Purpose | To show newsfeed |
| Overview | The user clicks on newsfeed and then system will show newsfeed to the registered users. |
| Type | Primary and essential |
| Pre-Condition | User must be signed in |
| Post-Condition | Newsfeed successfully shown |
| Normal flow | |
| Actors Actions | System Response |
| 1. The user clicks on newsfeed and see the latest news | 1. The system shows the newsfeed to the users. |
| Alternative flow | |
| 1a. The user clicks on newsfeed button but does not see newsfeed news | 2a. The system shows the message about unavailability of newsfeed. |

# 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:

### 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.

### 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.

### 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.

### 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.

### 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

# 4.0 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.

# 4.1 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.

# 4.2 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 part of it.

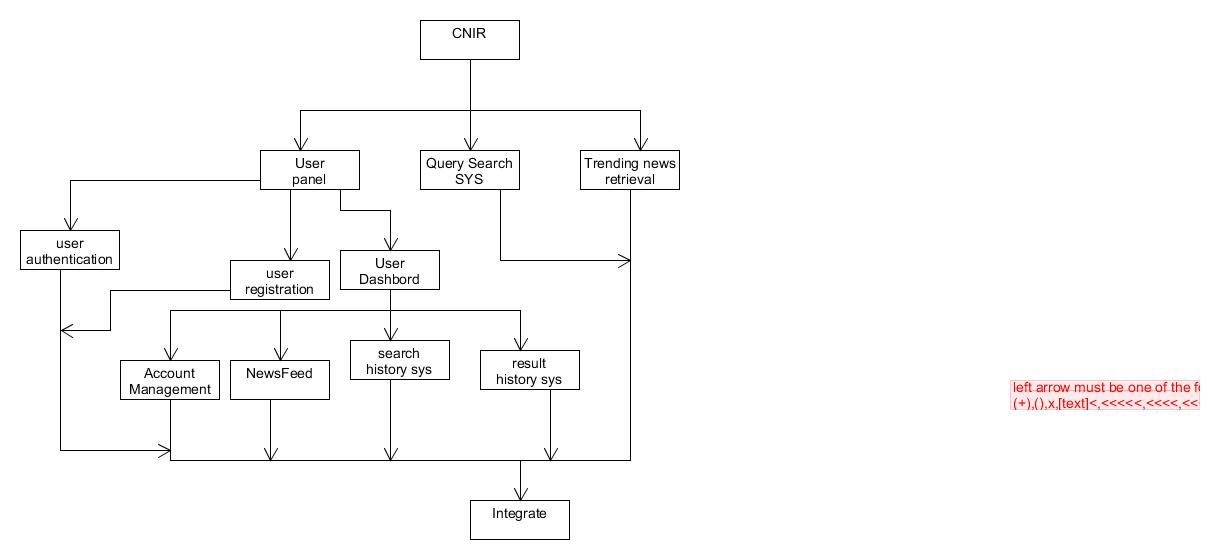
The top-down approach for our system is given below.

Figure 4.1: Top down approach for system

# 4.3 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.

### 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.

#### Landing page

#### Sign in

A screenshot of a computer

Description automatically generated

#### Sign up

A screenshot of a computer screen

Description automatically generated

#### Search news

A screenshot of a cell phone

Description automatically generated

#### Homepage for Registered user

A screenshot of a cell phone

Description automatically generated

#### Latest news

A screenshot of a social media post

Description automatically generated

#### Newsfeed

A screenshot of a social media post

Description automatically generated

#### Change account name

A screenshot of a computer

Description automatically generated

#### Change account password

A screenshot of a computer

Description automatically generated

#### Dashboard

A screenshot of a computer screen

Description automatically generated

#### Select news Interests

A screenshot of a computer screen

Description automatically generated

#### Dashboard

A screenshot of a computer screen

Description automatically generated

# 4.4 4+1 view Model of Architecture

All views which are application to this project are described below diagrams.

### 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.

#### 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.

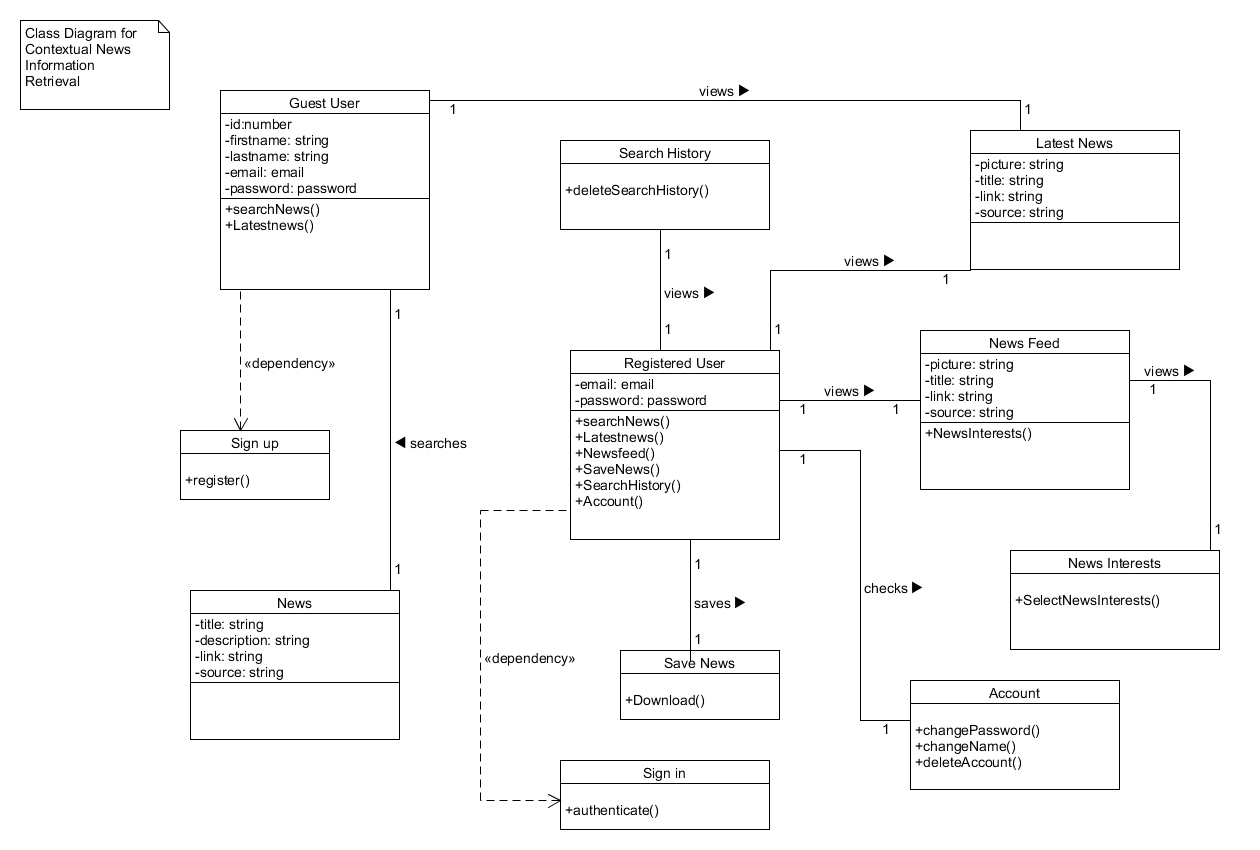


Figure 4.2: Class diagram of a system

#### ERD

### 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

#### 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

A picture containing text, map

Description automatically generated

Figure 4.3: Activity diagram of a system

#### State Machine Diagram

State machine diagram is diagram that shows the discrete behaviour of the system. It contain entities, states and connect flow between them.

A picture containing text

Description automatically generated

#### 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.

##### Sign in

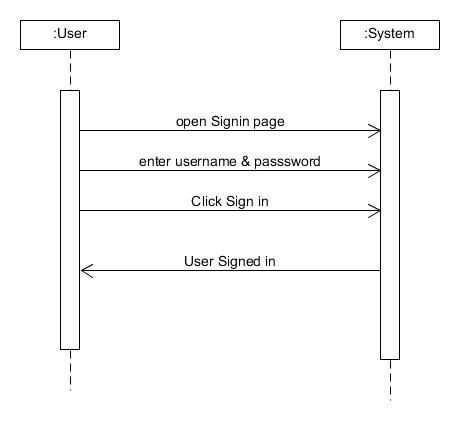


Figure 4.4: Sequence diagram for Signin

##### Sign up

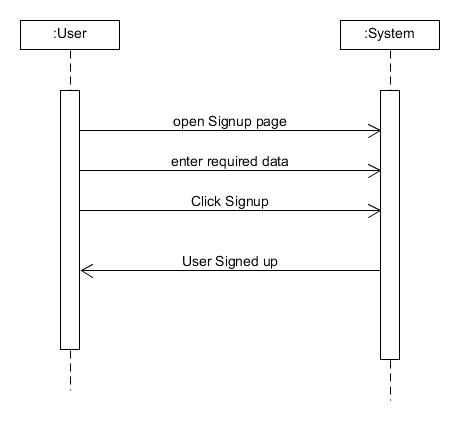


Figure 4.5: Sequence diagram for Signup

##### Search News

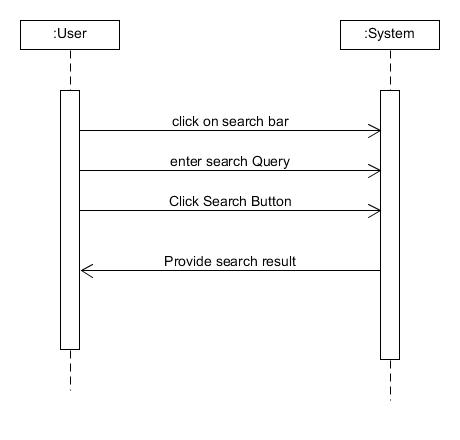


Figure 4.6: Sequence diagram for Search news

##### Search History

A picture containing screenshot

Description automatically generated

Figure 4.7: Sequence diagram for Search history

##### News Feed

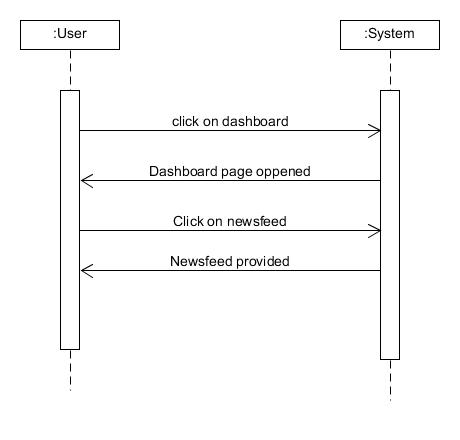


Figure 4.8: Sequence diagram for Newsfeed

##### Latest News

A screenshot of a social media post

Description automatically generated

Figure 4.8: Sequence diagram for latest news

##### Select news interests

A screenshot of a cell phone

Description automatically generated

Figure 4.8: Sequence diagram for Select news interests

##### Save news

A screenshot of a cell phone

Description automatically generated

Figure 4.8: Sequence diagram for save news

### Development View

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

#### 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.

A close up of text on a white background

Description automatically generated

Figure 4.10: Component diagram of a system

### 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.

#### 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.

A close up of a logo

Description automatically generated

Figure 4.11: Deployment diagram of a system

# 4.5 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.

# CHAPTER 5

# IMPLEMENTATION

# 5.0 Introduction

This chapter discusses the implementation of the system. Module by module implementation of the system is explained in this chapter. How each module of the system was developed, and which library, framework, or service was used. Detail of implementation of every module of the system is discussed. The working of the library, framework, or service that was used in the module is also explained in detail.

# **5.1 Modules of Project**

Modules of the system are given below with their implementation details.

### User Registration

The module user registration will allow the guest users of the system to register (sign up) themselves on the system, the user will enter first name, last name, email, password to register on the system.

This module will allow registered users of the system to authenticate (sign in) themselves on the system, the user will enter email and password to authenticate on the system.

### News Search

The module News search of the system will allow users to search for the news. The user can enter the keyword related to the story, and our website will scrap the required data, which is the best match for these keywords. The user can only read the information on-site and visit the site via the link which is provided with the news.

### Save News

One of the most important modules of our system is save news. This module provides a platform to record data extracted data from relevant sources. Users can use this platform to have factual data in the form of document and download it without wasting time and effort.

### Search History

This Module of the system will allow users to check their previous searches. The user will log in to the system and click on the dashboard where the history option is provided. Our system maintains the complete details of the user history in that section, and the user can clear the history.

# 5.2 Tools and Technology

The following are tools, libraries, framework, and services that were used in developing the modules of the system.

### Flask

Flask is a framework or micro-framework which is supported by python programming language, and it is mostly written in python. It is maintained by community of individual python developers and companies. Flask can be used as a base in the development of different functions or applications, as it is optimal for fetching changing data that needs to be recorded. We used to Flask for developing the modules of our system.

### Tesseract-OCR

Tesseract OCR is an optical character recognition drive for a different type of operating systems. It is optimized for complex character recognition and simple character recognition, layout analysis in the back end of systems for desktop and web-based applications. Tesseract is licensed by Apache and maintain by Google from 2006.

### Pandas

Pandas is a software library written in different programming languages includes Python, and C. Pandas developers develop it for different mathematics and time series analysis, including different tables data. Consequently, Panda is free software which is used for different useful functionalities.

### Beautiful Soup

Beautiful Soup is a python-based package which is used for parsing. Parsing is done for HTML and XML documents, which is later used for scrapping with modular architecture and applications using plug-and-play components that are aimed for use within private enterprises. It offers a unique approach to the consensus that enables performance at scale while preserving privacy.

### Requests

Request is a library developed for the better understanding of extracting data from different HTML and XML, and it is human friendly and easy to use.

### NumPy

NumPy is an important library of python programming language mainly designed for dealing with matrices; matrices involved two dimensional, three dimensional, and so on and other mathematical programming calculations. It is also popular in dealing with different arrays.

### NLTK

It is a tool kit where you can develop your Python programming language to deal with different human languages and make programs to understand the natural language more closely.

### SKLearn TFIDF Vectorizer

SKLearn TFIDF vectorizer is mainly used to tokenize the document, analyse the text for reverse document frequencies and encode the material, and it is also used for weighting the document and learning the vocabulary of it. In this project it has been used to find term frequency and inverse document frequency of the entered query by user so the system can get the context of what the user is searching.

### SKLearn Cosine Similarity

SKLearn Cosine Similarity is used to find the similarity between the documents, it mainly calculates its frequency of similarity with the help of x and y in cosine angle.

## Summary

In this chapter, we have discussed the module implementation of the system. Tools that were used in developing were discussed briefly. We have used Flask, Pandas, NumPy, Beautiful Soup and Requests libraries in the back end for the developing of the searching and scrapping modules of our system. Flask was used in the backend development as well as in the front-end development of the scrapping module of the system. NLTK was used to judge the human language more closely and is used for more efficient solution to the natural language extraction. HTML, CSS, JavaScript, and acts as a foundation for developing front end design, validation with modular architecture and applications using plug-and-play components. Detailed discussion about the implementation of the system is done in this chapter.

# CHAPTER 6

# TESTING, ANALYSES AND VALIDATION

# 6.0 Introduction

All the necessary reporting and documentation required for testing and validation in this chapter. As we know that in testing, we run our project with the intention to find the errors and remove them to improve the functionalities of our system. We will make the use of test cases to test the functionalities of the system by analysing and validating the results gathered from those test cases. As we know that testing is a process of executing the system in a real life and then analysing and evaluating the results. Our goal is to find and correct as much errors and defects we can before launching the system in the market. Some of the testing technique that we have used in the testing process are unit testing in which a specific capability or a functionality of the system is tested, integration testing which is performed after some parts of the whole system are integrated in order to test how one functionality of the system works with the other when they are kept and run in parallel, in order for us to check the overall functionality of the system by observing the results according to some one given inputs black box testing comes for the aid. But these type of testing technique have nothing to do when it comes to finding and removing the specific integral parts of the code therefor we use white box testing in which in code and the modules are tested in very much detail.

When we consider all of this it is very clear that testing is a very important step of the development process and by using the testing methods by a tester or a testing team we can make sure that weather the requirements or the demands of the are being flailed by the system or not. The test cases and the scenarios are analysed by giving input opposing to every module and the component of the system. The test cases of the python logic and code have been executed on Visual Studio Code and the test cases for the web have been executed using Google Chrome, Visual Studio Code etc. Test cases are very important in discovering the defects errors and bugs of the system and they point out the components and the modules of the code that are not working properly therefor we make changes in the code and correct our mistakes.

# 6.1 Test Bed

A test is an execution environment used for testing. A Test bed consists of specific collection hardware, software, Operating system, network, the product being tested, other system software and application software. All the functionalities of the system and the modules are tested on this test bed. The test bed used for the testing of “Contextual News Information Retrieval” is a Laptop with Windows 10 running Operating System, 8GB RAM and 1 Terabyte Hard Disk, and the applications used in testing are visual studio code and google chrome.

# 6.2 System Test Case

The system test cases are written and executed to find all the possible errors and bugs in the system. System testing helps us to improve the functionality of the system and it also helps us to ensure and validate that our system is being made according to the requirements of the user. Throughout the testing of the system we have found a large number of errors and bugs and that is exactly what our goals was many of the bugs were corrected but many were very hard to settle there for they have been contained and minimized.

# 6.3 Test Cases

Test cases are the scenarios or the conditions that are used by a tester to ensure that the system that is being made is according to the requirement of the user and to detect the defects and the errors in the system before launching it in the market. Test cases help us to insure that our system is error free and performing its functionality without any risks involved the test cases have aided the testers to detect errors and have aided the development team to remove these errors. Test cases have also highlighted some of those defects which were very hard to deal with and to be removed therefor the development team has done some providers to minimize their impact and their containment.

All the test cases consist of the following components which are test case ID, test case reference, QA test engineer, name of personnel, test date, revision history, objective, testing environment, assumptions, preconditions, steps, and status of a test case. Test cases of all features of the system are given below in detail.

### Sign in

The sign in of the user is a necessary functionality of the system. Without this feature the user is not able to perform some important functionalities that are related with the system.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-1 (sign in) | Tester | | Annas Israr |
| **Test Date** | 15-01-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To sign in or log in the user | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is present at sign in page and has registered before. | | | |
| **Pre-Requisite** | System is in functional state. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Display sign in page  Input the data to sign in (Email and Password) | | Page displayed  The user has input the data. | |
| **Comments:** sign in without any delay | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Sign up

The registration of the user is a very important feature. The user of the system registers himself on the system by using this feature. This use case starts when the user enters the specific data of himself

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-2(signup) | Tester | | Faizan Ahmad |
| **Test Date** | 15-01-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To register new users | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is present at signup page | | | |
| **Pre-Requisite** | System is in a functional state. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Display signup page  Input the data of the user on system | | Page displayed  The user has input the data. | |
| **Comments:** Signup functionality performed without any delay | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Search news

Search news is the main functionality of the system. This feature allows us to search the news according to the query of the user. This functionality allows the user to extract the data from different sources of the interests and bring the result to the user.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-3(Search news) | Tester | | Zain Ul Abedin |
| **Test Date** | 20-05-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To search the desired news | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in | | | |
| **Pre-Requisite** | System is in a functional state. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2.  3 | Display home page  Input the search query  Search event triggered by click | | Page displayed  The user has input the data.  User has clicked the search button | |
| **Comments:** news retrieved based on the context with 15-20 seconds delay | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Search history

This is a very important feature and can be used only by the registered users the guests users cannot use this feature this feature help to maintain and manage the history of the search quires that the user made and provide the user with a list of search queries he had made.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC4(search history) | Tester | | Yaldram Shahzad |
| **Test Date** | 21-05-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To view the queries of search of the user | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in | | | |
| **Pre-Requisite** | System is in a functional state and the user is registered. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Display dashboard page  See under the search History | | Page displayed  History displayed | |
| **Comments:** updated history list of the query is displayed at the dashboard page. | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Select interests

This is a very important feature and can be used only by the registered users, the guest users cannot use this feature. This feature allows the signed in user to select his or her interests according to which he wants his newsfeed to be updated.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-5(select interests) | Tester | | Faizan Ahmad |
| **Test Date** | 25-05-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To select the topics of interests from mentioned categories by the system | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in | | | |
| **Pre-Requisite** | System is in a functional state and the user is registered. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Display dashboard page  Select the interest of news you wish to see | | Page displayed  Interests selected and saved in the database | |
| **Comments:** dashboard is refreshed, and newsfeed is updated | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Save news

This is a very important feature and can be used only by the registered users, the guest users cannot use this feature. This feature allows the signed in user to save the results of the search that he has made locally to their device in a readable form.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-6(Save new) | Tester | | Annas Israr |
| **Test Date** | 25-05-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | To save the results of the search to a local drive | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in | | | |
| **Pre-Requisite** | System is in a functional state and the user is registered. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Search news  Press the save/download/export button. | | Search result displayed  A CSV file will be download to local drive. | |
| **Comments:** saved the contents of the search in a readable file | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Latest news

This is a very important feature and can be used by the registered users and the guest users. This feature allows the system to display the latest news events that are happening throughout the world to the users.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-7(latest news) | Tester | | Yaldram Shahzad |
| **Test Date** | 30-05-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | Display the latest news | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in | | | |
| **Pre-Requisite** | System is in a functional state and the user is registered. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2. | Open landing page  Click on the latest news button | | Page opened  Latest news of interests displayed | |
| **Comments:** The news is displayed of the latest topic. | | | | |
| **Status** | Pass  Fail Not Executed | | | |

### Newsfeed

This is the most attractive feature of the system which cannot be used by a guest user the user has to be signed in to use this functionality. This feature displays a list of news to the user which are according to the interests that he had selected before and according to his past searches. This is a modified information retrieval to the user.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test Case ID** | **Use Case Reference** | **QA Test Engineer** | | **Name of Personnel** |
| TC-1 | UC-7(latest news) | Tester | | Yaldram Shahzad |
| **Test Date** | 05-06-2020 | | | |
| **Revision History** | None | | | |
| **Objective** | Display the news of the user’s interests | | | |
| **Environment** | User Mode | | | |
| **Assumptions** | User is signed in and has selected the interests | | | |
| **Pre-Requisite** | System is in a functional state and the user is registered. | | | |
| **Steps #** | **Execution Description** | | **Procedure Result** | |
| 1.  2.  3. | Open landing page  Click on the newsfeed  News according to selected interest displayed | | Page opened  Newsfeed page opened  See the news of yours interests | |
| **Comments:** the news is displayed of the interests selected before | | | | |
| **Status** | Pass  Fail Not Executed | | | |

# 6.4 Summary

In this chapter, we did the testing, analysis of the test results, and validated our system. First, we have done testing; it is an activity or process of finding errors and then removing those errors so that the system contains no defects and performs its functions correctly. To have this part done different kinds of testing have been done for achieving this goal that includes system testing, integration testing, black box testing and white box testing. We have evaluated and tested all the specific parts of the system therefor unit testing has been done. Integration testing and white box has been a continuous procedure and have been done in parallel to development and after compilation along with black box testing.

The errors or the testing results that we faced were then analyzed and corrected some of the errors were very difficult to deal with and to be corrected so that is why contained and minimized them.

After the correction of these errors we have then validated our system to ensure that the functionalities of the system are working fine, and our system is doing what it is supposed to do and fulfilling the requirements.

# CHAPTER 7

# CONCLUSION AND FUTURE WORK

# 7.0 Overview

This chapter is about the conclusion and future recommendations of the system, that how much functionalities according to the ones discussed earlier in proposed system, are successfully implemented, what milestones are achieved and what future work can be done to improve the system. Milestones achieved in developing the system are discussed briefly. Limitations of the project are also discussed that needs to be handled in future if the system is upgraded.

# 7.1 System Overview

“Contextual News Information Retrieval” is a web-based application for searching news as required with ease. Users can register themselves to access all the features of the system. A user can search his desired news over the internet through using the system, users can view their search history and can download the information to their system. The thing that differentiates this system from other existing systems of this type is that this system can search for the required news in any format, although it is a pdf file or an E-Newspaper.

# 7.2 Milestones Achieved

Milestones testify the completion of a project from its starting to ending date. They identify what functionalities or features have been implemented.

### User Registration

User can register himself successfully by providing a few details about him, that include his first name, last name, email, and password. After registration, the user can login by entering his email and password.

### News Search

After successful login, user can enter his desired query in the search box. The system will perform the context annotation on his entered words to get the context of what has been written in the search box, after this the system will first search for required data in the information already saved and will find the similarity, if the similarity of the news will be more than 50% that is the similarity threshold we have set, then it will display the information. If it fails to find the required data, then it will search over the web to provide the user with the results he is asking for. This functionality has been successfully implemented.

### Display of Information

This feature is about how the system shows the searched information. This feature has been successfully implemented and the system shows the searched results to user in an understandable format.

### Latest news

This feature is about the latest trends that are happening around. The system has this feature successfully implemented and it shows the latest trending topics to both (logged-in user and guest user).

### Newsfeed

This feature was intended to show the user news according to his selected interests. The system now successfully shows the news about topics that the user had selected, from the five websites that were mentioned earlier.

### Search History

This feature was intended to save the user’s search history. It has been implemented successfully and the system is maintaining the search history of the registered user.

### News Interests

This feature was about the user’s selection of interests from the following four categories that include: Sports, Entertainment, Business, Pakistan. The system now allows the user to select his interests from these four categories and provide newsfeed according to these interests.

### Save **Information** on local machine

This system feature was intended to allow the user to download the information on his system. This functionality has been successfully implemented and users can now download the information.

# 7.3 Limitations

Although the system has implemented the features that were proposed, but still there are some limitations that emerged from discussions. Following are some limitations of the system.

1. System is currently limited to five news websites only.
2. News Interests of the user are limited to only four categories.
3. Mainly the sources that are covered by the system are Pakistani news websites.

# 7.4 Future Work

In general, if we talk about a news retrieval system then it should be able to retrieve any kind of news from any source, but still this system will facilitate many researchers, journalists and student to get their required data with minimum effort. If we mainly talk about expanding the scope of the application to more sources and websites, it can be time-consuming as separate scrappers need to be designed for each website and each template.

The possibility of implementing a full-fledged news retrieval system with a broader scope is still an open debate that can be worked on in future. News retrieval systems have a large amount of data as something is happening around in the world with each passing second. So, a more dynamic system can be developed that can handle any website coming its way with ease.

# 7.4 Summary

In this section, we discussed about the findings of the project as now we aim to close the project. The limitations that came in the way of developing the system are also discussed. Milestones are something that trace the completion of a specific work and so all the milestones that were identified in the initial stages of development are successfully achieved. A user can now register/log-in to the system, can view his newsfeed according to the interests he selected. User can download the information successfully. A very important portion of the system was searching the news according to users query, which has also been implemented successfully. System now understands the context of the user’s entered text and can successfully find the data related to that in the saved information with 50% similarity threshold with 15-20 seconds delay, or over the web. Search history of the user is being maintained. Limitations that were on the way to developing the system were also discussed in this chapter, these limitations can be addressed in future versions of the project. There is some future work that was discussed and can be done with advancement in available libraries in future. Future recommendations were also discussed.