**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 users context. Methods that are choosed 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 6](#_Toc32867853)

[INTRODUCTION 6](#_Toc32867854)

[1.1 Introduction 7](#_Toc32867855)

[1.2 Motivation 7](#_Toc32867856)

[1.3 Problem statement 7](#_Toc32867857)

[1.4 Scope 8](#_Toc32867858)

[1.5 Process model 8](#_Toc32867859)

[1.5.1 Requirement gathering 9](#_Toc32867860)

[1.5.2 Work breakdown and decomposition 9](#_Toc32867861)

[1.5.3 Designing 9](#_Toc32867862)

[1.5.4 Coding 9](#_Toc32867863)

[1.5.5 Integration 9](#_Toc32867864)

[1.5.6 Testing 9](#_Toc32867865)

[1.5.7 Why incremental model 9](#_Toc32867866)

[1.6 Proposed system 10](#_Toc32867867)

[1.7 Features 10](#_Toc32867868)

[1.7.1 Text Scrapping 10](#_Toc32867869)

[1.7.2 Scrapping of text in images 10](#_Toc32867870)

[1.7.3 Sign up / Sign in 10](#_Toc32867871)

[1.7.4 Guest user 10](#_Toc32867872)

[1.7.5 Merging and referencing 11](#_Toc32867873)

[1.8 Nature of the project 11](#_Toc32867874)

[1.9 Summary 11](#_Toc32867875)

[CHAPTER 2 12](#_Toc32867876)

[BACKGROUND AND EXISITING WORK 12](#_Toc32867877)

[2.1 Introduction 13](#_Toc32867878)

[2.2 Existing Systems 13](#_Toc32867879)

[2.2.1 Google News 13](#_Toc32867880)

[2.2.2 ABC News 13](#_Toc32867881)

[2.3 Limitations of Existing Systems 13](#_Toc32867882)

[2.3.1 Providing Links 13](#_Toc32867883)

[2.3.2 Unavailability of Information in Document Form 13](#_Toc32867884)

[2.3.3 Not Analysing Pictorial Information 14](#_Toc32867885)

[2.3.4 Time Consuming 14](#_Toc32867886)

[2.3.5 Presenting the information 14](#_Toc32867887)

[2.4 Solutions to Limitations of Existing Systems 14](#_Toc32867888)

[2.4.1 Providing Links 14](#_Toc32867889)

[2.4.2 Availability of Information in Document Form 14](#_Toc32867890)

[2.4.3 Analysing Pictorial Information 14](#_Toc32867891)

[2.4.4 Time Saving 14](#_Toc32867892)

[2.4.5 Presenting the information 14](#_Toc32867893)

[2.5 Comparison of Existing Systems and Proposed System 14](#_Toc32867894)

[2.6 Summary 15](#_Toc32867895)

[CHAPTER 3 16](#_Toc32867896)

[REQUIREMENTS SPECIFICATION 16](#_Toc32867897)

[3.1 Introduction 17](#_Toc32867898)

[3.1 Interface Requirements 17](#_Toc32867899)

[3.1.1 Software Requirements 17](#_Toc32867900)

[3.1.2 Hardware Requirements 17](#_Toc32867901)

[3.2 Functional Requirements 18](#_Toc32867902)

[3.2.1 User Registration 18](#_Toc32867903)

[3.2.2 News Search 18](#_Toc32867904)

[3.2.3 Display of Information 18](#_Toc32867905)

[3.2.4 Latest news 18](#_Toc32867906)

[3.2.5 Newsfeed 18](#_Toc32867907)

[3.2.6 History and Interest 18](#_Toc32867908)

[3.2.7 Save information on local machine 18](#_Toc32867909)

[3.3 Non-Functional Requirements 18](#_Toc32867910)

[3.3.1 Security 19](#_Toc32867911)

[3.3.2 Reliability 19](#_Toc32867912)

[3.3.3 Availability 19](#_Toc32867913)

[3.3.4 Maintainability 19](#_Toc32867914)

[3.3.5 Portability 19](#_Toc32867915)

[3.4 Use Case diagrams 19](#_Toc32867916)

[3.4.1 Sign in 20](#_Toc32867917)

[3.4.2 Sign up 22](#_Toc32867918)

[3.4.3 Search news 23](#_Toc32867919)

[3.4.4 Manage history 25](#_Toc32867920)

[3.4.5 Manage interests 26](#_Toc32867921)

[3.4.6 Manage news results history 27](#_Toc32867922)

[3.4.7 Save news 28](#_Toc32867923)

[3.5 Resource Requirements 30](#_Toc32867924)

[3.5.1 Web Scraper 30](#_Toc32867925)

[3.5.2 Tools 30](#_Toc32867926)

[3.6 Database Requirements 31](#_Toc32867927)

[3.7 Project Feasibility 31](#_Toc32867928)

[3.7.1 Technical Feasibility 31](#_Toc32867929)

[3.7.2 Operational Feasibility 31](#_Toc32867930)

[3.7.3 Legal and Ethical Feasibility 31](#_Toc32867931)

[3.8 Summary 32](#_Toc32867932)

**LIST OF FIGURES**

[Figure 3.1: Use case diagram for Contextual News Information Retireval 23](#_Toc32870342)

[Figure 3.2: Signin 24](#_Toc32870343)

[Figure 3.3: Signup 26](#_Toc32870344)

[Figure 3.4: Search news 27](#_Toc32870345)

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

[Figure 3.6: Manage news interests 30](#_Toc32870347)

[Figure 3.7: Manage news results 32](#_Toc32870348)

[Figure 3.8: Save news 33](#_Toc32870349)

**LIST OF TABLES**

[Table 1.1: Scope of the project 11](#_Toc32870342)

[Table 2.1: Comparison of existing and proposed system 18](#_Toc32870343)

[Table 3.1: Sign in 24](#_Toc32870344)

[Table 3.2: Sign up 26](#_Toc32870345)

[Table 3.3: Search news 28](#_Toc32870346)

[Table 3.4: Manage search history 29](#_Toc32870347)

[Table 3.5: Manage news interests 30](#_Toc32870348)

[Table 3.6: Manage result history 32](#_Toc32870349)

[Table 3.7: Save news 34](#_Toc32870350)

# CHAPTER 1

# INTRODUCTION

## 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 and which is for the access of people.

This has made a major impact on the people’s way of living and due to this almost every business is trying to provide every facility to their customers and users online and one click away. The businesses aim to do this in order to make the processes effective and save time in processes at the business end. Just like the business community wants to save their time. 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.

## Motivation

As there are trillions of data bytes of information over the internet therefore searching for the required news and information can be very much time consuming 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 interests. Secondly to eliminate or minimize the fake news and Hoax that circulate on the internet.

## 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.

## 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 used 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 internet connection is available, proper functionality of the database. |
| **Stake holders** | User, news researchers, news reporters, developers, testers, admins. |
| **Risks** | Lack of internet, slow internet connection, complex database management. |
| **Deliverables** | A web-based application, final documentation report. |

## 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 using the waterfall model. 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 a document
* Providing references of the extracted 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.

## 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.

## Features

Following are the five main features of this system.

### 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.

### 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.

### 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.

### Guest user

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

### 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.

## 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.

## 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

# 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 physically 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.

## 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 physically 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 also it does not analyses pictorial information and E-Papers.

## Limitations of Existing Systems

### 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.

### 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.

### Not Analysing Pictorial Information

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

### 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.

### Presenting the information

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

## Solutions to Limitations of Existing Systems

### Providing Links

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

### 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.

### 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.

### 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.

### 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.

## 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 users script | No | No | Yes |
| Download Option | No | Yes | Yes |
| Searching Option | Yes | Yes | Yes |

## 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

# 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.

## 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.

### 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.

### 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.

## 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 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.

### 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.

### 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 show latest information based on trending topics.

### Newsfeed

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

### History and Interest

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

### 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 users device.

## 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.

### 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.

## 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.

A close up of a map

Description automatically generated

Figure 3.1: Use Case Diagram for Contextual News Information Retrieval

### 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 | Signin |
| 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. | 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 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.

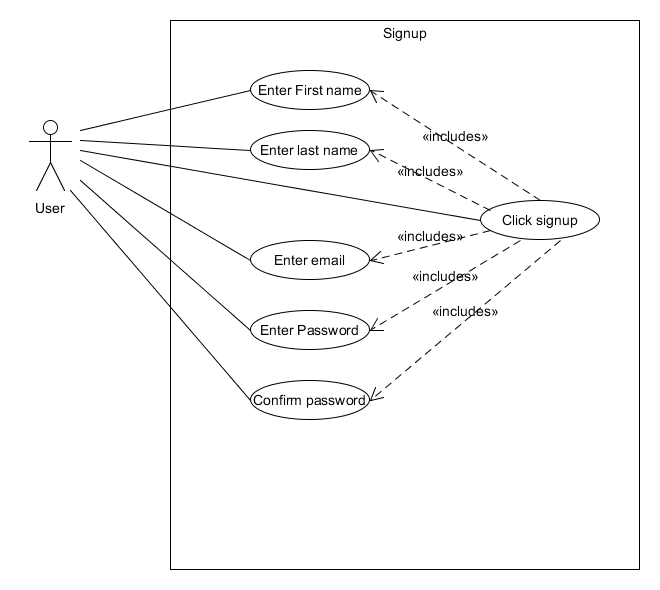


Figure 3.3: Use case for Sign up

Table 3.2: Sign up

|  |  |
| --- | --- |
| Use case ID | UC-2 |
| Use case name | Signup |
| References requirement | Requirement no.2 |
| Actors | User |
| Purpose | To signup 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 |
| 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 | |
| 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. |

### 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 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 map

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 | 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 | |
| 5a. The user does not enter anything. | 6a. The system shows a message to enter some keywords. |

### 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.

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 | 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 |
| 1. The user clicks on delete history. | 1. 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. |

### 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 updation of news interest and shows an appropriate massage. The system also gives response on unsuccessful deletion or updation of news interests and shows an appropriate massage.

A close up of a map

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 | 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 |
| 1. The user sees the news interests and clicks delete interests. 2. The user sees the news interests and clicks update interests. | 1. The system deletes the news interest and shows news interests again to demonstrate changes. 2. 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. |

### 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.

A screenshot of a cell phone

Description automatically generated

Figure 3.7: Use case for Managing news result history

Table 3.6: 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 |
| 1. The user clicks on delete news results. | 1. 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. |

### 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.

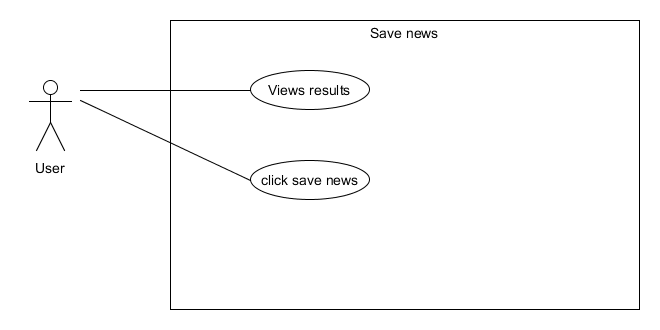


Figure 3.8: Use case for saving news results

Table 3.7: 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 |
| 1. The user sees the news results and clicks save news results. | 1. 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. |

## 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 framework of python that is used at front end to run webpages and integrate them to be ready to work with 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.

## 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.

## 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.

## 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.