Summ-The-Search

By

Liaqat Ali

Shaharyar Shaikh

Umar Daraz Bhatti

Submitted to the Faculty of Computer Software Engineering, National University of Sciences and Technology, Islamabad in partial fulfillment for the requirements of a B.E. Degree in Computer Software Engineering

MAY 2016

CERTIFICATE OF CORRECTNESS AND APPROVAL

Certified that work contained in this thesis “Summ-The-Search” carried out by Liaqat Ali, Shaharyar Shaikh and Umar Daraz Bhatti under the supervision of Dr. Seemab Latif for partial fulfillment of Degree of Bachelor of Software Engineering is correct and approved.

Approved By

_____

(Dr. Seemab Latif)

Department of Computer Software Engineering

Military College of Signals, Rawalpindi

Dated: _____

ABSTRACT

Day by day with the extensive increase of internet data, its difficult to go through every website to find manual solution of any query. Single Document Summarization (SDS) has been used for summarization of internet web pages so far.

Summ-The-Search is an android application is using Multi Document Summary(MDS) for internet search results to summarize multiple web pages (English Language only), given as a result of a query search on Google. First five web pages will be taken for summarizing. These web pages will be first cleaned and pre-processed and then important information will be extracted using feature extraction technique. This will then be used for multi document summarization. Summary length will be decided by the user of the application in terms of percentage. Query string to meaningful summary(sentences) generation: Don't have to go through each and every page of suggested link , so no longer feel the dilemma of extracting information from multi dynamic pages.

The main fields is being used AI, Data Mining, Behavior analysis and different types of summary algorithms to produce meaningful results are being used.

It can be widely use by students as well as by office people. As it will be saving a lot of time so general public will also be benefited. In short any Google user can get help from this app an very easy way of gathering data from different websites.

COPYRIGHT NOTICE

We very solemnly declare that the work presented herewith is the result of sole effort of our group, comprising of Liaqat Ali, Umar Daraz Bhatti , Shaharyar Shaikh, and is free of any kind of plagiarism in part or whole. We also declare that the dissertation has never been submitted previously in part or whole in support of another award or qualification either at this institution or elsewhere.

DEDICATION

To our respected teachers whose kind guidance and unfailing support made this mammoth task easy for us and to our very dear parents whose unceasing prayers gave us strength and courage to complete the work of this magnitude.

ACKNOWLEDGEMENTS

We are very humbly grateful to Almighty Allah for bestowing us with the strength and resolve to undertake and complete the project**.**

We owe a special debt of gratitude to our supervisor, Dr. Seemab Latif for the continuous supervision, motivation and support provided to us and for their continuous and valuable suggestions, guidance, and instructions from time to time right through the project.

We would also like to acknowledge and thank the benign faculty members of the department of Computer Software Engineering for extending their kind and continuous support to help us in Mobile app development. It was their help and guidance which helped us complete the project in due time.

Table of Contents

List of Figures

List of Tables

Chapter 1

1. Introduction

Summ-The-Search is an android Application, providing text summarization to the world wide android users. It is an efficient and fast way of jumping to the result in search engine solutions.

The proposed project is an Android Application named as Summ-The-Search. The application will be used to generate the summary of the search pages. The scope of this project is limited to summarizing only first 5 web pages in English language, given by Google search. Extended scope includes this number to be variable and decided by the user of the application. Complete android user friendly application will be available at the end of this project.Day by day with the extensive increase of internet data, it's difficult to go through every website to find manual solution of any query. Single Document summarization (SDS) has been used for summarization of internet web pages so far.

Our Summ-The-Search an android application is using Multi Document Summary (MDS) for internet search results to summarize multiple web pages (English Language only), given as a result of a query search on Google. First five web pages will be taken for summarizing. These web pages will be first cleaned and pre-processed and then important information will be extracted using feature extraction technique. This will then be used for multi document summarization. Summary length will be decided by the user of the application in terms of percentage. Query string to meaningful summary (sentences) generation: Don't have to go through each and every page of suggested link .So No longer feel the dilemma of extracting information from multi dynamic pages.

The main fields are being used AI, Data Mining, Behavior analysis and different types of summary algorithms to produce meaningful results are being used.It can be widely used by students as well as by office people.As it will be saving a lot of time so general public will also be benefited. In short any Google user can get help from this app and very easy way of gathering data from different websites.

1.1. Motivation

Summ-The-Search is intended to narrow down the gap between the user and the easy way to access data by providing them with a common platform. The project has multidimensional effects on the way of addressing issues in the society. Industry can refer Summ-The-Search for carrying out data mining and NLP Projects. Society in general will also be able to help the user in their respective domains. Moreover the app will take care of the ever changing requirements of the user thus allowing the system to address there and then, whenever it is possible for them. Data Base updation will be performed on the basis of user's query. Similarly The user will be notified by the email and work will be saved on the server as long as user wants.

1.2. Project Scope

The proposed project is an Android Application named as Summ-The-Search. The application will be used to generate the summary of the search pages. The scope of this project is limited to summarizing only first 5 web pages in English language, given by Google search. Extended scope includes this number to be variable and decided by the user of the application. Complete android user friendly application will be available at the end of this project.

1.3. Deliverables

1. Android Application
2. Documentation (User Manual)

1.4. Document Organization

This document provides basic knowledge about Search-The-Search. Initially the description of project has been given, the next chapter discusses related software products and their functionalities and how Summ-The-Search is different from them. Requirement specifications have been covered in chapter 3, while the Design specifications are discussed in chapter 4. Implementation of the system is described in chapter 5 and further chapters ponders upon the analysis of the software, testing techniques employed and suggestion about the future work for enhancing the functionalities and capabilities of the system.

1.5. Summary

This chapter is an introduction to Summ-The-Search and its scope, objectives, and motivation for developing this software. A brief introduction to document contents is also included.

Chapter 2

2. Literature Review

2.1. Introduction

Summ-The-Search will provide the user with the context oriented and compact summary generated from multiple links. The main purpose of the App is to set up a platform for the student, office worker and other online user to generate the summary of his desire topic which would be in result of saving a lot of their time. This App is using Extracting techniques. Abstracting approach could be used for the future work and is the research area.

2.2. Problem Domain

Currently there are many online text summarizer to generate the summary for the user, but those summarizers are generate summary of only single document i.e Single Document Summarization (web pages) and we are making multi-document summary without wasting of user's precious time. The main aim of the project is to develop an android application to summarize multiple web pages (English Language only), given as a result of a query search on Google. First five (extendable) web pages will be taken for summarizing. Online users require a lot of time to search for his related topic or to get the information of his domain. This App will provide the user with the compact solution of his user and would save much of his time.

2.3. Related Work

The related work helps in understanding the existing solution of the problem domain with a better insight. Currently there are many online text summarizers to generate the summary for the user, but those summarizers are generating summary of only single document i.e. Single Document Summarization (web pages). Except that a MDS website is also there but it's not for free and is used by private organizations to get the summary from multiple links. Our App would generate the summary of multiple links and would be available for the common man.

Iresearch-reporter (a MDS website but it’s not free).

Gist(a single document summarization app)

Ultimate Research Assistant

JistWeb is a query specific multiple document summarizer.

2.4. Limitations

1. The contents of the application will be in English language only.
2. The server will be unavailable in case of maintenance and testing issues.
   No backup server configuration is provided.

2.5. Technological & Software Requirements

Summ-The-Search will be developed for Android operating system using Android Studio. To save the history of the query searched by the user will be as separate Database. For that purpose we will use Shared Preferences. For sending the summary as the email we are using Email Intent built in Android.

2.6. Hardware Requirements

The Hardware required for the implementation of our project includes:

1. Smart Phone and tablets having Android as its operating system (as a testing Device) .
2. Android Studio IDE.
3. A working internet connection.

2.7. Summary

This chapter introduces us to existing systems for community services and their features and limitations. The section further highlights some of the shortcomings in Summ-The-Search, while mentioning about the technological and hardware requirements of the system.

Chapter 3

3. Software Requirement Specifications

3.1. Product Perspective

Extracting most relevant information from first five pages suggested by Google search and providing the summary data to user at the end. Extended version may have the ability to work on extra details.

3.2. Product Functions

1. Search and summarize
2. Displaying the summary
3. Emailing a summarized file
4. Displaying the history
5. Updating and saving the history
6. Clearing the history

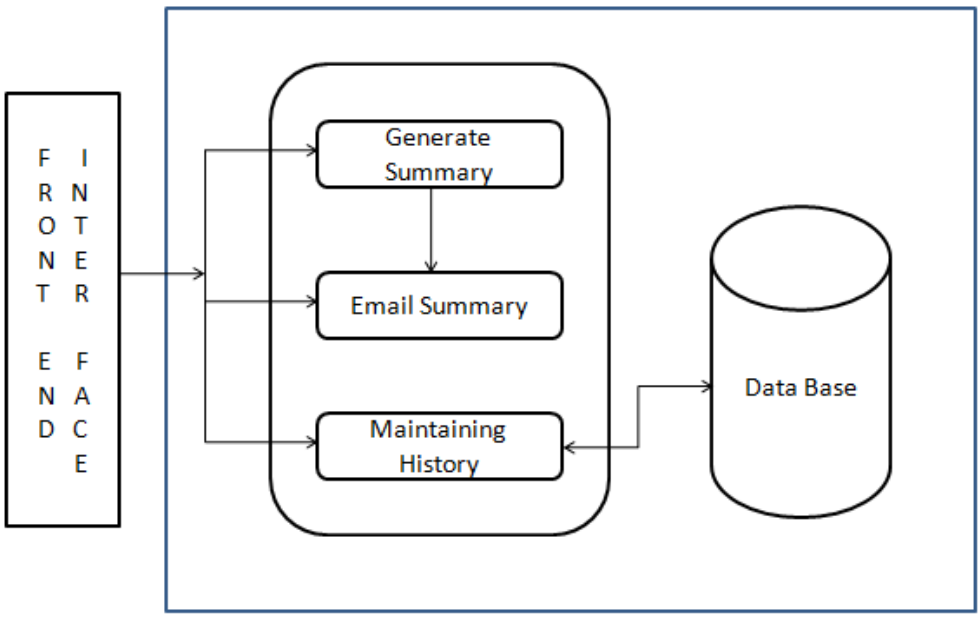


Figure 1: Abstract Design of Summ-The-Search

Following are the functions that Summ-The-Search should provide:

3.3. Search and Summarize

3.3.1. Description and Priority

The purpose is to allow the customer to generate the summary.

Priority Level: HIGH

3.3.2. Stimulus/ Response Sequences

First user will write the query. Second it will select no of search links means from how many links he want to make summary. Finally it will give the length of the summary (in term of Sentence length and word length).

3.3.3. Functional Requirements

If user has not written anything in the query or has left anything then it will not proceed and will generate dialog box.

3.4. Displaying the Summary

3.4.1. Description and Priority

The application allows the user to read, cancel and email the summary.

Priority Level: HIGH

3.4.2. Stimulus/ Response Sequences

Once the user press the “Search & Summarize” button then new window will appear. New window will show the summary and now user can cancel it or email it.

3.4.3. Functional requirements

If the user cancels the summary then it will go back to “Main page”.User can also mail it on desired email address.

3.5. Emailing a Summarized File

3.5.1. Description and Priority

This feature of application will allow the user to email that summary

Priority Level: HIGH

3.5.2. Stimulus/ Response Sequences

Once the user gave command for email after summary generation then new window will appear. User has to give desired email address and in pdf format file will be emailed.

3.5.3. Functional Requirements

New email window requires the email address, if anyone missed or wrong email is given then it will not proceed and if user does cancels summary then in history “TO” parameter would be “NA”.

3.6. Displaying the History

3.6.1. Description and Priority

The purpose is to allow the customer to check, delete or clear the history.

Priority Level: HIGH

3.6.2. Stimulus/ Response Sequences

Once the user press the button “History” from Home page then it will directs the user to history page from where can come back to home page. In history Query Search, Date, Time, and email addresses are saved.

3.6.3. Functional requirements

Once user reaches on history page, user can check the history. Further user can clear the whole history it would be saving last 10 searches.

3.7. Updating and Saving the history

3.7.1. Description and Priority

This functionality of the application will save and maintain the history after every deletion or summarization.

Priority Level: Medium

3.7.2. Stimulus/ Response Sequences

After pressing the button “Email” after summarization, application will directs the user to email page where user will give email address.

After giving all the information in email intent, text will appear giving successfully sent message and given parameter will be saved in history.

3.7.3. Functional requirements

Once user summarized anything, simultaneously application will also add that thing into the history. User deletes/generates anything from history, simultaneously application will update itself.

3.8. Other Nonfunctional Requirements

3.8.1. Performance Requirements

3.8.1.1. Response Time

The response time of the application would be less than 30 seconds as soon as it extracts the relevant data from document.

3.8.2. Capacity

The application can summarize one document at a time and can save last 10 summaries parameters.

3.8.3. Accuracy

The application would be able to take out relevant data and generate the precise summary and grammatically correct as much as possible.

3.8.4. Availability

Application will be available on Google store and can be used for summarization (web pages only).

3.8.5. Safety Requirements

The application shall not accidently lose/delete the files associated with it, such as summary generated or history which is maintaining.

3.9. Security Requirements

- Only end-user would be able to generate summary of the document.
- Summarized file will be sent only to email address given by that end-user.

3.10. Software Quality Attributes

3.10.1. Availability

The application can be used 24/7 as long as the cell phone works properly.

3.10.2. Correctness

If the user gives right/defined commands for a particular action the options should be selected correctly.

3.10.3. Extensibility and Maintainability

- Changes required by law will applied at least three months.
- The application can be further extended.
- The application can be improved to generate more than one summary at a time.

3.10.4. Portability

The application can be installed on any compatible hardware(with android OS) meeting the requirements

3.10.5. Reliability

The application will be available to users 98% of normal working hours.
The factors needed to establish the software expected reliability are:

- The user inputs should be valid and within the given range.
- Normal termination of the program

3.10.6. Usability

The application will be very easy to use by the users after being trained. However, the output of the application can be understood by any web Surfer i.e. End -Users

3.10.7. Business Rules

As per defined Google store policies this application can be used.

3.11. User Classes and Characteristics

There would be one user class that would be searching a query on web and application will generate summary of first 5 links and present it to the user. Users will also have options of giving query summary length, selecting number of pages for summary generation and they will also be having preference after summarization to cancel that summary or to email it. While emailing, user will be able to get that file in pdf.

3.12. Operating Environment

The Summ-The-Search operates, either directly or indirectly, with the following external hardware:

3.12.1. Hardware

Workstations:

Any android supported smart phone can run this application.

Customer Peripherals:

The smart phones (android) used by end-users over the Internet using a web interface.

Internet:

The global network (Wifi,3G,4G) will use for communication.

3.12.2. Software

Summ-The-Search operates, either directly or indirectly, with the following software:

Operating System:

Android based, only with networking allowed, through which this application shall communicate with internet. Android 2.2 or superior is required.

3.13. Design and Implementation Constraints

3.13.1. Data and Content Constraints

Each and every summarized file will be stored in database (History) and will be present till deleted and also if it comes in last 10 generated summaries.

3.13.2. Hardware Constraints

Data is stored in memory of respective cell phone with minimum of 2 GB of storage is required.

3.13.3. High-Level Languages

Android (java + xml) will be used to design and implement this application.

3.14. User Documentation

A few documents shall be delivered along with the software:

1. Installation guides
2. Usage manuals with pictures and text for using the software.
3. Tutorials
4. Online help links

3.15. Assumptions and Dependencies

3.15.1. Technical knowledge limitation

The group is only well versed with one language of the 4th generation languages today (e.g. C#, JAVA, .NET) that enable creation of GUIs and modern APIs and libraries. The group has received formal training in C and has studied Object Oriented Programming, networking in java and also has studied networking or web applications that would enable them to develop networked applications that would run on the internet and telephone lines. The group shall, therefore, in accordance with the training they have and the one they are undergoing in their 8th semester of their undergraduate degree, try to develop as good a product as they can according to their expertise level. The product shall not meet market requirements, but this is already understood by both the developers and the customer.

3.16. Summary

This chapter describes the requirements of the system as described by Project Supervisor. It includes interface, functional and nonfunctional requirements along with the main features, system would provide to the end user. These requirements have been set after checking the feasibility of the system. These requirements have been considered as the fundamental principles for testing and standardization of the product.

Chapter 4

4. System Design Specifications

4.1. Overview of the modules

The system will be architected mainly in 3 fundamental modules “Front End Interface”, “Core System”, and “Database” having other sub modules too as shown in the following abstract diagram:

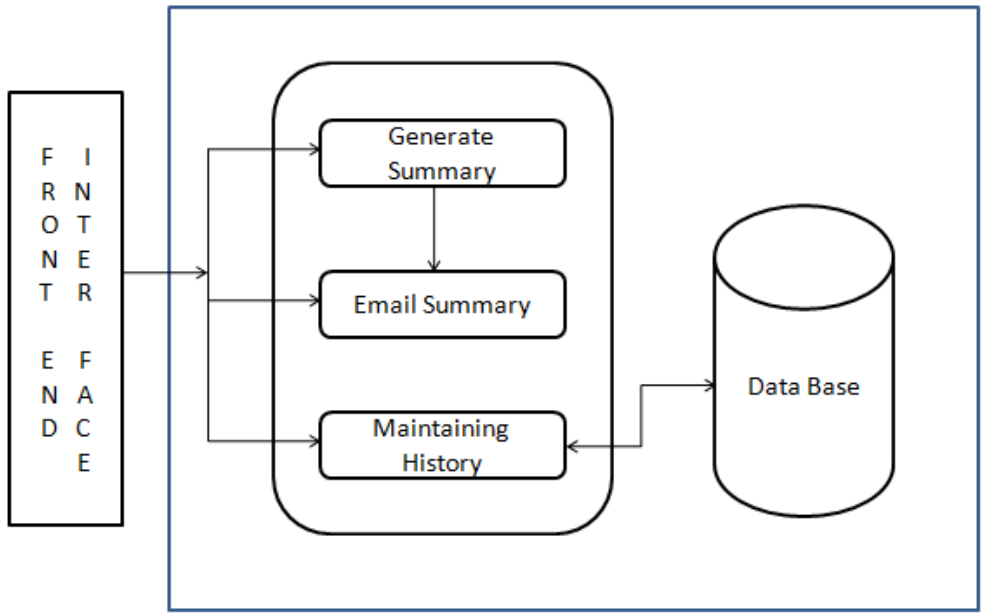


Figure 2: Overview of the Modules

4.2. Description of the modules

The “Front End Interface” will be used to pass on the information such like query, no of links and length of summary to the “Generate Summary Module” which will determine the input and removes the ads and images and references links and will generate the summary by applying different filters (Keyword searching, query word, position of sentence) out of it. The generated summary will be passed to “Email Module” which will use for emailing it, after emailing, history will be recorded in the “History Module” through a database afterwards user can clear it completely or last 10 summaries would be available and keep on updating by “History Module”.

4.3. System Architecture

Layered Architecture (3-Tiers) will be used to implement Summ-The-Search. From a high level perspective, a service-based solution can be seen as being composed of multiple services, each communicating with the others by passing messages. Conceptually, the services can be seen as components of the overall solution. However, internally, each service is made up of software components, just like any other application, and these components can be logically grouped into presentation, business, and data layers. Other applications can make use of the services without being aware of the way they are implemented.

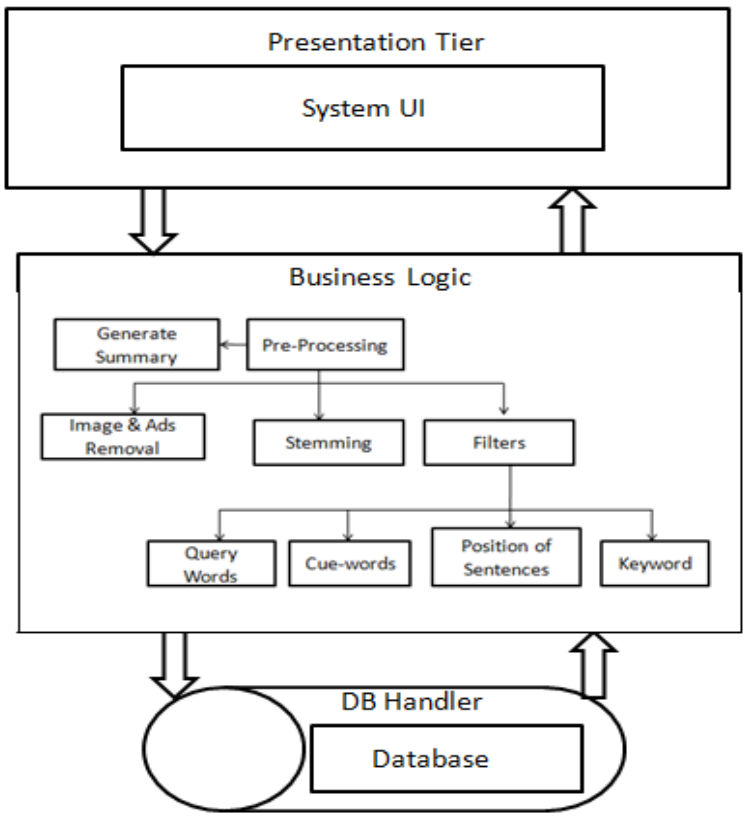


Figure 3: 3-Tier Architecture

4.3.1. Layers Details

The details of the layers have been discussed below:

4.3.1.1. Presentation Layer

The presentation layer provides the platform for interaction of the users with the system. It displays homepage to the user and accepts input from the user. The Presentation layer can only receive requests from, and return responses to, an outside agent. This is usually a person, but may be another piece of software. Here, it can only send requests to, and receive responses from, the Business Logic layer. It cannot have direct access to the Database.

4.3.1.2. Business Logic

Business Logic has been used as a service layer to expose the business functionality of the application. Comprising of elements of “Generate Summary”, “Pre-Processing”, History”, Stemming, AdsRemoval, Extraction (Keyword Search, Position of Sentence, Cue-Word, Query word) the layer caters for the core functionality of the system. The Business Logic layer effectively allows the clients to get access to the Database through it, and get the desired search queried in a profound manner.

4.3.1.3. Data Access Layer

This layer receives request from the Business Logic Layer and sends back the data after querying it from the Database server.

4.4. Structure and Relationships

Focusing upon the internal structure of the system, this section ponders upon the interrelationships and dependencies among various components.

4.4.1. Overall Structure of the system

The diagram shows the main components of the system along with their interactions with each other. It mainly describes the system structure which is further augmented by the explanatory text as follows:

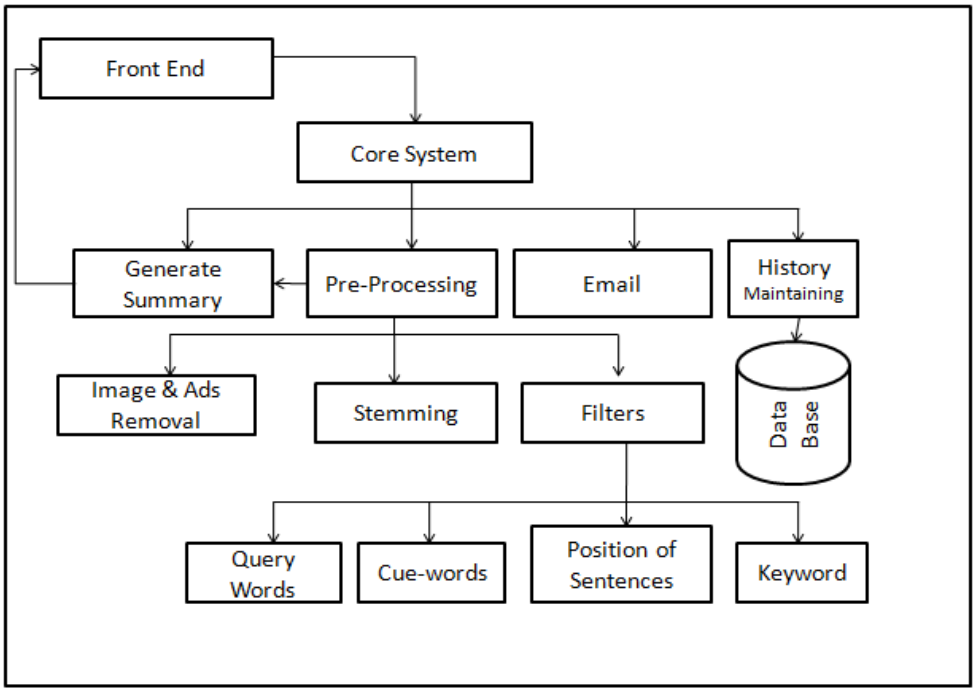


Figure 4: Over Structure of System

4.4.1.1. Front End Interface

Front End Interface caters for the visual needs of the application, wherein the Human-Computer-Interaction aspects are considered to enable the user to communicate with the system profoundly. It is connected with Core System and Generate Summary modules to pass on the user inputs to the core system as well as display the output feedback to the user respectively.

4.4.1.2. Core System Module

This module is comprised of Pre-Processing, Email, History, Generate Summary. It looks after the intra-communication of all these modules.

4.4.1.3. Pre-Processing

This module process the user inputs (query, no of links and summary length) and removes the ads and images and references links and will generate the summary by applying different filters (Keyword searching, query word, position of sentence) out of it by applying different filters.

4.4.1.4. Images and Ads Removal

This is one of the sub modules of pre-processing in which images and ads would be removed from the extracted links.

4.4.1.5. Stemming

This is one of the sub modules of pre-processing in which words having different forms used in documents would be converted into simplest or first form such eating, ate and eaten into eat, eat and eat.

4.4.1.6. Filters

This is one of the sub modules of pre-processing in which data would be extracted by applying different filters (Query Words, Cue-words, Position of Sentences, Keyword) after removing the ads and image removal.

4.4.1.7. Query Words

This is one of the sub modules of filters in which sentence would be selected on the basis of query word.

4.4.1.8. Cue-Words

This is one of the sub modules of filters in which we would be able to find out the query related sentences.

4.4.1.9. Position Of Sentence

This is one of the sub modules of filters in which we filter out the sentences on the basis of position of sentences normally in the documents important sentences are in the beginning or conclusion paragraph.

4.4.1.10. Generate Summary

This module will generate summarized file after pre-processing and applying different filters and that would be emailed.

4.4.1.11. Email

This is the one of the modules of core system in which generated summary would be sent to given email address and user will also be able to select document type of that summarized file.

4.4.1.12. History Maintaining

This is the one of the modules of core system in which users entered query, email address, date and time of emailed file would be recorded and maintained herein and database would be used Shared Preferences.

4.5. UML Diagrams

4.5.1. Use case Diagram

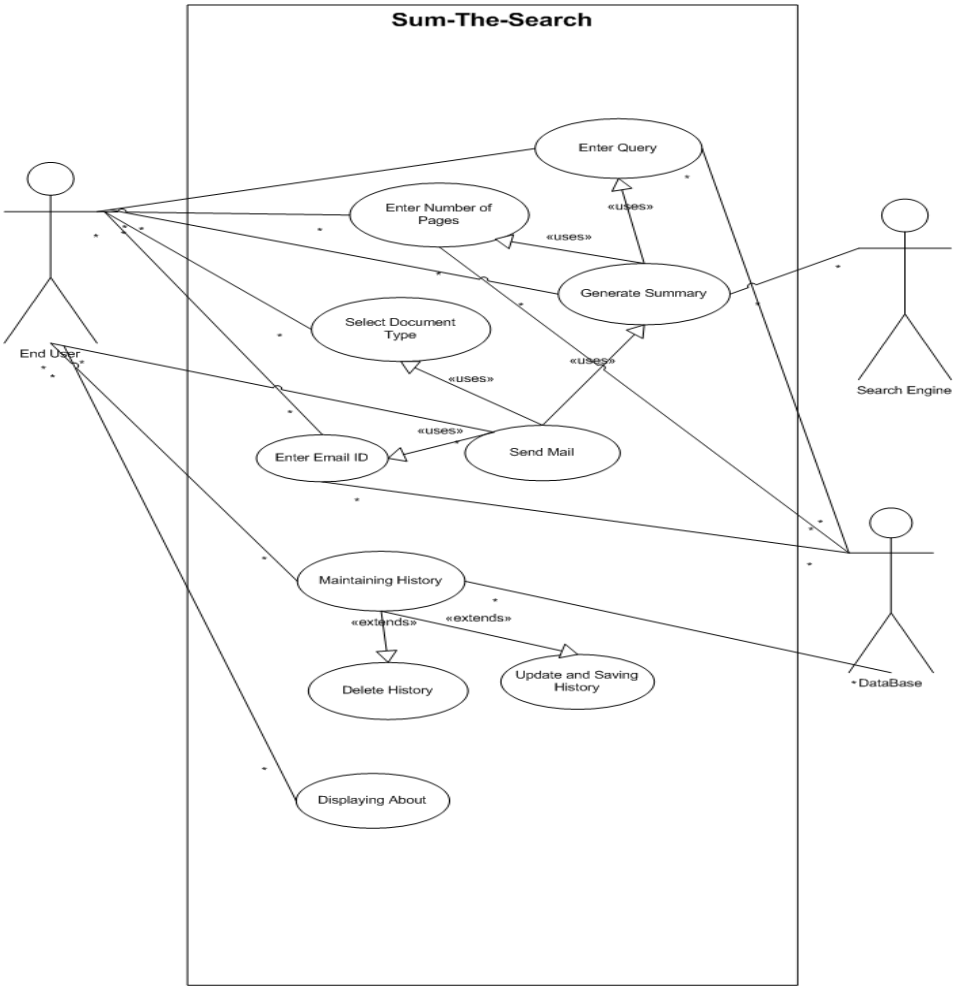


Figure 5: Use-Case Diagram

4.5.2. Use case Description

Description is attached as appendix ‘A’.

4.6. Class Diagram

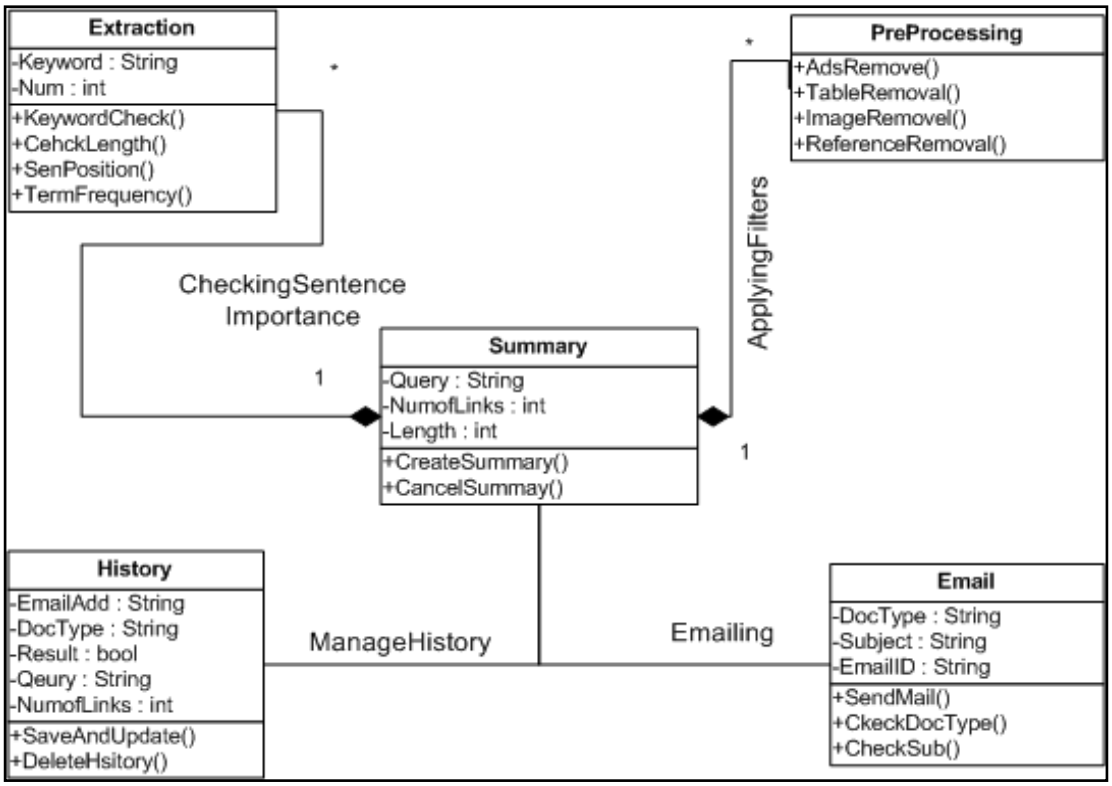


Figure 6: Class –Diagram

4.6.1. Class Diagram Description

Description is attached as appendix ‘B’.

4.7. UML Activity Diagram

This section shows the activities that a user need to perform to accomplish a task.

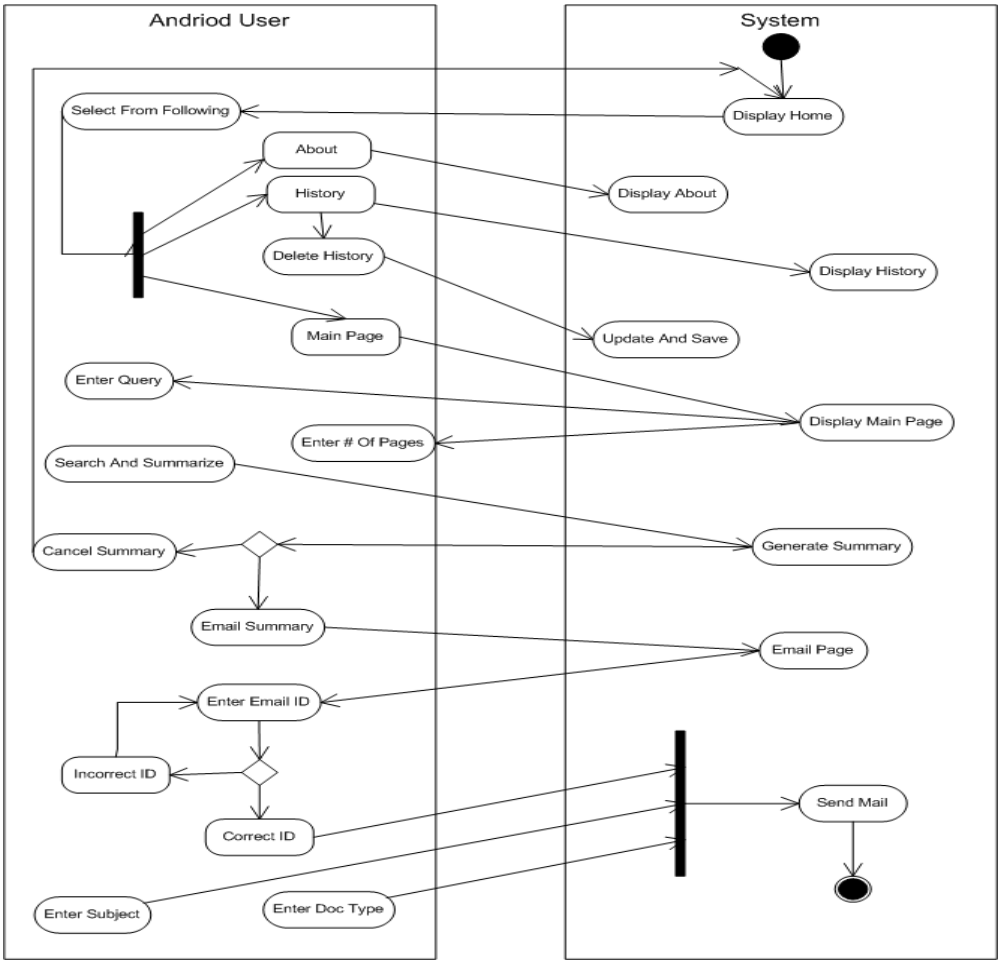


Figure 7: Activity Diagram

4.7.1. Activity Diagram Description

Description is attached as appendix ‘C’.

4.8. UML Sequence Diagram

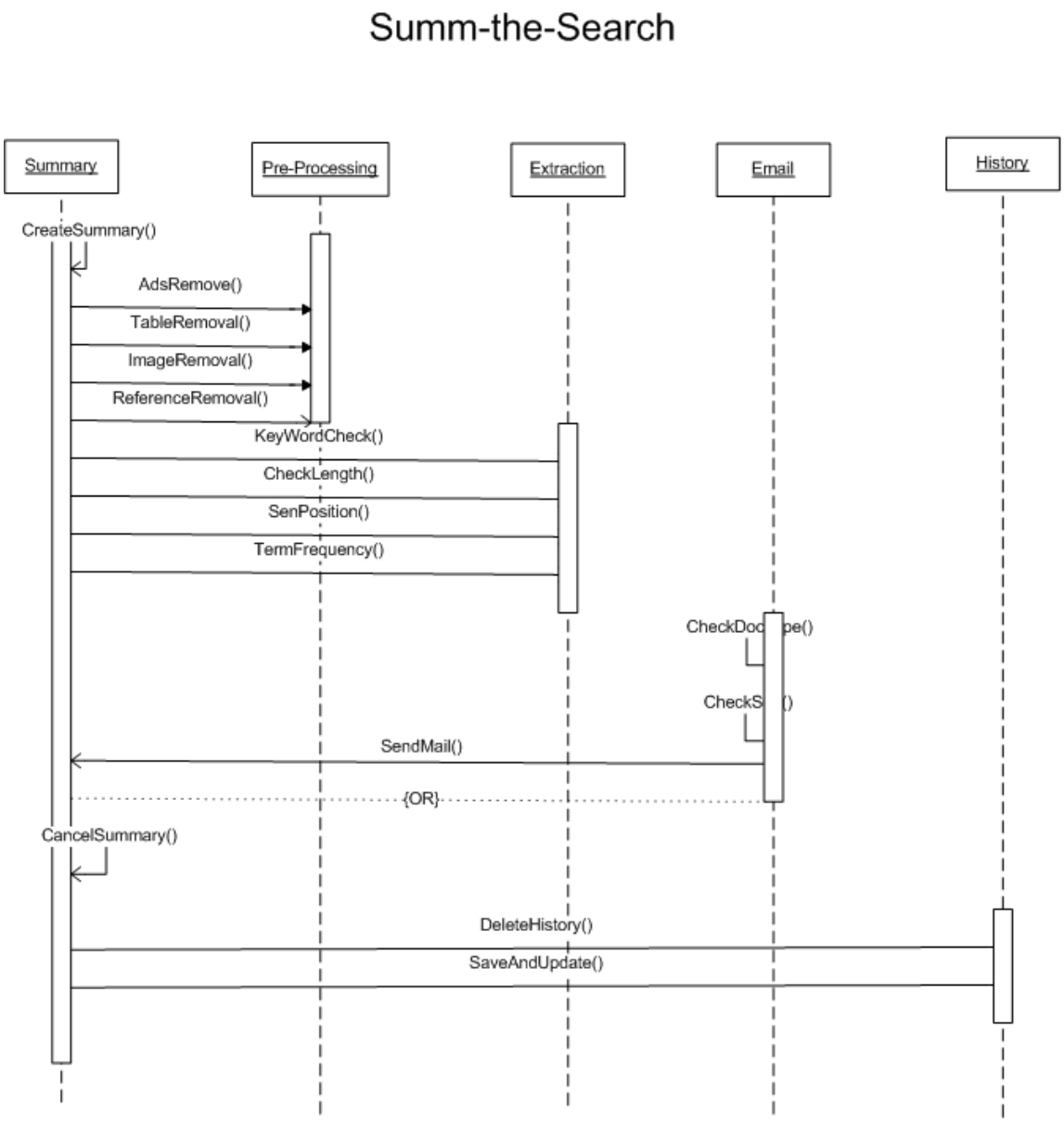


Figure 8: Sequence Diagram

4.8.1. Activity Diagram Description

Description is attached as appendix ‘D’.

4.9. User Interface Design

Summ-The-Search is an Android application and intended to be used by the users from diverse background knowledge. This requires that the interface of Summ-The-Search should have an easy learning curve for the user. Most of the important features should be visible to the user and no functionality should be hidden.

Please note that the interfaces provided are just for demonstration purposes. Actual interfaces may be different.

Home Page

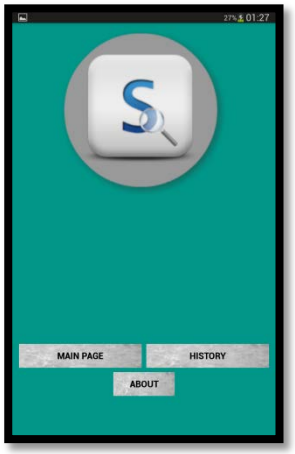


Figure 9: Home Page

Main Page

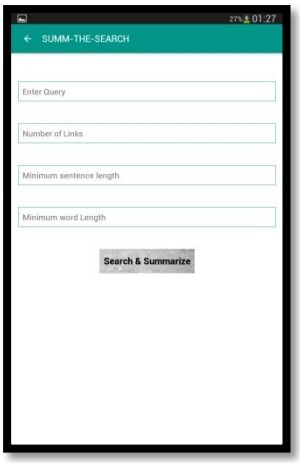


Figure 10: Main Page

Summarized Search

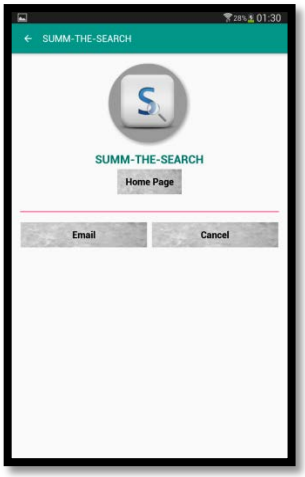


Figure 11: Summarized Search

Email

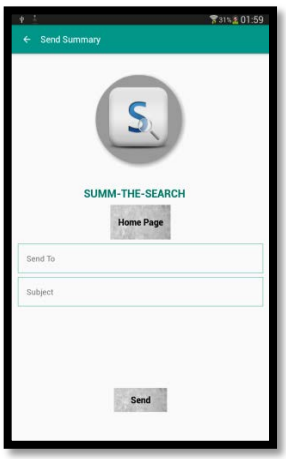


Figure 12: Email

History

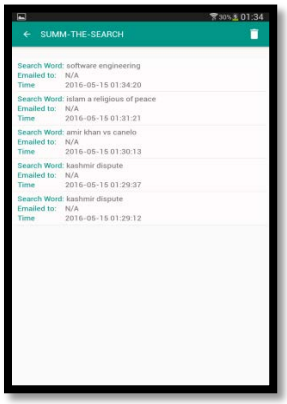


Figure 13: History

4.10. Summary

This chapter described the design of the software taking into consideration different assumptions and constraints that applied on the system because of its goals and requirements. Use-Case Diagram, Class Diagram, Activity Diagram and Sequence Diagram have been added to explain system functionalities. A few low level design diagrams and GUI have also been shown to ascertain the internal behavior of the system.

Chapter 5

5. Implementation Details

5.1. Introduction

Detailed design of the Summ-The-Search is discussed in the previous chapter. This design is transformed into an application by using various technologies. The implementation details are discussed in the following sections giving details of the system's internal working.

5.2. Tools and Technologies

It is an Android application, tools & technologies include Android Framework and Java as programming language. We have used data bases and to communicate (data retrieval) with server we have used volley library which internally handle HTTP API calls, request queues, request/response time. Email intents are being used to mail final result.Application works in android studio, supporting minimum jellybean (android 16). Jsoup for the parsing of data, shared preferences in order to store history of transaction which has been executed, itext lib for pdf conversion and volley for data search on web.

5.2.1. Java

Java is a dynamic programming language. It is most commonly used to develop android applications in android studio. This mix of features makes it a multi-paradigm language, supporting object-oriented, imperative, and functional programming styles.

5.2.2. Android SDK

The Android software development kit (SDK) includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows XP or later.

5.3. User Interface

Since Summ-the-search is a mobile application so it needs to be more interactive, user friendly and responsive. Navigation flow of the application should be self-driving so as to guide the user in completing his desired task/activity. Interface is an important part of any mobile application and it is treated as a separate module. The issues addressed in the user interface design are that no functionality is hidden from user and data should be presented in a clear way to end user so nothing is missed by him. The interface is developed using Android native Layouts, list views, buttons, action bars and drawer layout. The interface is kept separate from business layer by making fragments classes. Each fragment in project hierarchy represents a different screen.

5.4. User Management Module

This module covers all functionalities associated with a user. It provides functionalities to search for summary about anything, getting to know the application, generating and emailing the summary to the respective address. Also manages to save/update the history.

5.5. Preprocessing Module

This module process the user inputs (query, no of links and summary length), connects to server, crawls the web pages data and removes the ads and images and references links and will generate the summary by applying different filters (irrelevant data, stemming, stopwards removal) out of it by applying different filters.

5.5.1. Images and Ads Removal

This is one of the sub modules of pre-processing in which images and ads would be removed from the extracted links.

5.5.2. Stemming

This is one of the sub modules of pre-processing in which words having different forms used in documents would be converted into simplest or first form such eating, ate and eaten into eat, eat and eat.

5.6. Extraction Module

5.6.1. Filters

This is one of the sub modules of extraction in which data would be extracted by applying different filters (Query Words, Cue-words, Position of Sentences, Keyword) after removing the ads and image removal.

5.6.2. Query Words

This is one of the sub modules of filters in which sentence would be selected on the basis of query word.

5.6.3. Cue-Words

This is one of the sub modules of filters in which we would be able to find out the query related sentences.

5.6.4. Position Of Sentence

This is one of the sub modules of filters in which we filter out the sentences on the basis of position of sentences normally in the documents important sentences are in the beginning or conclusion paragraph.

5.6.5. Generate Summary

This module will generate summarized file after pre-processing and applying different filters and that would be emailed.

5.7. Mail and History Module

5.7.1. This Email

This is the one of the modules of core system in which generated summary would be sent to given email address and user will also be able to select document type of that summarized file.

5.7.2. History Maintaining

This is the one of the modules of core system in which users entered query, email address, date and time of emailed file would be recorded and maintained herein and database would be used Shared Preferences.

5.8. Summary

Implementation details of Summ-The-Search are discussed in this chapter. Different functionalities and strategies to develop the system have also been pondered upon.A brief introduction to different tools and technologies employed is also given.

Chapter 6

6. Testing

6.1. Introduction

To ensure quality of the product, testing is conducted. Accuracy of functions performed by Summ-The-Search has to be tested and maintained to improve quality of software. Software testing techniques and results obtained are discussed in the coming sections.

6.2. Testing Levels

Separate modules are developed to provide different functionalities of Summ-The-Search. All of these modules are tested at different levels in their development and post-integration process. Different levels at which Summ-The-Search has been tested, and results obtained are described in this section.

6.2.1. Unit Testing

Unit testing involves the testing of each module at the completion and at times, during the very course of development of the module. It is a testing method by which individual units of source code, sets of one or more computer program modules together with associated control data, usage procedures, and operating procedures, are tested to determine whether they are fit for use. Summ-The-Search app has also been passed through unit testing process, wherein the various units have been tested in accordance with the anticipated output of each unit.

6.2.2. Integration Testing

Summ-the Search different modules which were developed and tested independently were also tested during integration to ensure system stability. Integration testing helped in ensuring that different modules when combined give complete functionality and nothing is missed or some functionality doesn’t give error when integrated with other modules.

6.2.3. System Testing

System testing was performed at the end of development and integration of Summ-the Search. Complete system was tested using sample data. All sub modules including user registration, publishing/ seeking the services, managing the credentials information, viewing/ editing the profile etc. were tested as a whole using the sample data.

6.3. Test Cases

Table 1: Test Case-01

Test Case Name	Main Page Button
Test Case Number	1
Description	Testing Main Page Button for application user.
Preconditions	Application should be open
Input	Click on Main Page Button
Steps	Select the Main Page Button from the Home Page
Expected output	Three Tests for Application User opens
Results	Three Tests for Application User opens

Table 2: Test Case-02

Test Case Name	History Button
Test Case Number	2
Description	Testing History Button for application user.
Preconditions	Application should be open and it must be Home Page
Input	Click on History Button
Steps	Select the History Button from the Home Page
Expected output	Three Tests for Application User opens
Results	Three Tests for Application User opens

Table 3: Test Case-03

Test Case Name	About Button
Test Case Number	3
Description	Testing the About Button for application user.
Preconditions	Application should be open
Input	Click on About Button
Steps	First Select the About Page Button from the Home Page
Expected output	Three Tests for Application User opens
Results	Three Tests for Application User opens

Table 4: Test Case-04

Test Case Name	Home Button
Test Case Number	4
Description	Testing the Home Button for application user
Preconditions	Application must be open and it must be on Main Page
Input	Click on the Home Page Button
Steps	First go to the Main Page then click on Home Page Button
Expected output	It should return to the Home Page
Results	Test to return to the Home Page

Table 5: Test Case-05

Test Case Name	Enter Query
Test Case Number	5
Description	Testing the input query from user in Enter Query text box
Preconditions	Application must be open and it must be on Main Page
Input	Click on the text box and will input query
Steps	First go to the Main Page then click on Enter Query text box and write query

Expected output	On clicking, typing screen should appear and take the query correctly
Results	Test for Typing screen opens, cursor blinks and took query correctly

Table 6: Test Case-06

Test Case Name	No Of Links
Test Case Number	6
Description	Testing the No Of Links input from user in No Of Links text box
Preconditions	Application must be open and it must be on Main Page
Input	Click on the text box and will input no of links
Steps	First go to the Main Page then click on No Of Links text box and will enter the no of the links
Expected output	On clicking, typing number screen should appear and give the number of the links
Results	Test for Typing screen opens, cursor blinks and took input(number)

Table 7: Test Case-07

Test Case Name	Search &Summarize Button
Test Case Number	7
Description	Testing the Search &Summarize Button to generate the summary
Preconditions	Application must be open and it must be on Main Page
Input	Click on the Search & Summarize Button
Steps	First go to the Main Page then click on Search & Summarize Button
Expected output	Test should open for the summary
Results	Summary test opens

Table 8: Test Case-08

Test Case Name	Email Button
Test Case Number	8
Description	Testing the Email Button for sending mail of generated Summary
Preconditions	Application must be open and summary must be generated
Input	Click on the Email Button
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button
Expected output	Test to prompt email address should open
Results	Test opens to prompt email address

Table 9: Test Cace-09

Test Case Name	Cancel Summary Button
Test Case Number	9
Description	Testing Cancel Summary Button to cancel the generated Summary
Preconditions	Application must be open and summary must be generated
Input	Click on Cancel Summary Button
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Cancel Summary Button to cancel the generated summary
Expected output	Test cancel summary should open and it should return to Home Page
Results	Test opens for cancel summary and return to Home Page

Table 10: Test Case-10

Test Case Name	Home Button
Test Case Number	10
Description	Testing the Home Page Button for application user
Preconditions	Application must be open and it must be on Email Page
Input	Click on the Home Page Button
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button then to return to Home Page click on the Home Page Button
Expected output	Test to return to the Home Page should open
Results	Test to return to the Home Page opens

Table 11: Test Case-11

Test Case Name	To
Test Case Number	11
Description	Testing the input email address from user in the “To” text box
Preconditions	Application must be open and it must be on Email Page
Input	Click on the text box and enter the email address
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button then click on the text box to enter the email address
Expected output	Test to enter email address should open
Results	Test to enter email address opens

Table 12: Test Case-12

Test Case Name	Subject
Test Case Number	12
Description	Testing to enter the Subject of generated summary from user in the Subject text box
Preconditions	Application must be open and it must be on Email Page
Input	Click on the text box and enter the subject of the summary
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button then click on the text box to enter the subject of the summary
Expected output	Test to enter the subject should open
Results	Test to enter subject opens

Table 13: Test Case-13

Test Case Name	Radio Button
Test Case Number	13
Description	Testing to select the type of mailing summary from user by checking the Radio Button
Preconditions	Application must be open and it must be on Email Page
Input	Select the type of mailing summary by checking the Radio Button
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button then check the Radio Button to select the Doc Type of the mailing summary
Expected output	Test to check Radio Button should open
Results	Test to check Radio Button opens

Table 14: Test Case-14

Test Case Name	Send Button
Test Case Number	14
Description	Testing the Send Button to send the generated to provided email
Preconditions	Application must be open and it must be on
Input	Click on the Search & Summarize Button
Steps	First go to the Main Page then click on Search & Summarize Button to generate summary then click on the Email Button then enter the email address, subject and finally click Send Button
Expected output	Test should send summary
Results	Test to send summary opens

Table 15: Test Case-15

Test Case Name	Home Button
Test Case Number	15
Description	Testing the Home Page Button for application user
Preconditions	Application must be open and it must be on History Page
Input	Click on the Home Page Button
Steps	First go to the History Page then click on Home Page Button
Expected output	Test to return to the Home Page should open
Results	Test to return to the Home Page opens

Table 16: Test Case-16

Test Case Name	Clear History Button
Test Case Number	17
Description	Testing Clear History Button to clear the history
Preconditions	Application must be open and at least one summary must be present
Input	Click on Clear History Button
Steps	First go to the History Page then click on Clear History Button to clear the history
Expected output	Test for clear history should open
Results	Test for clear history opens

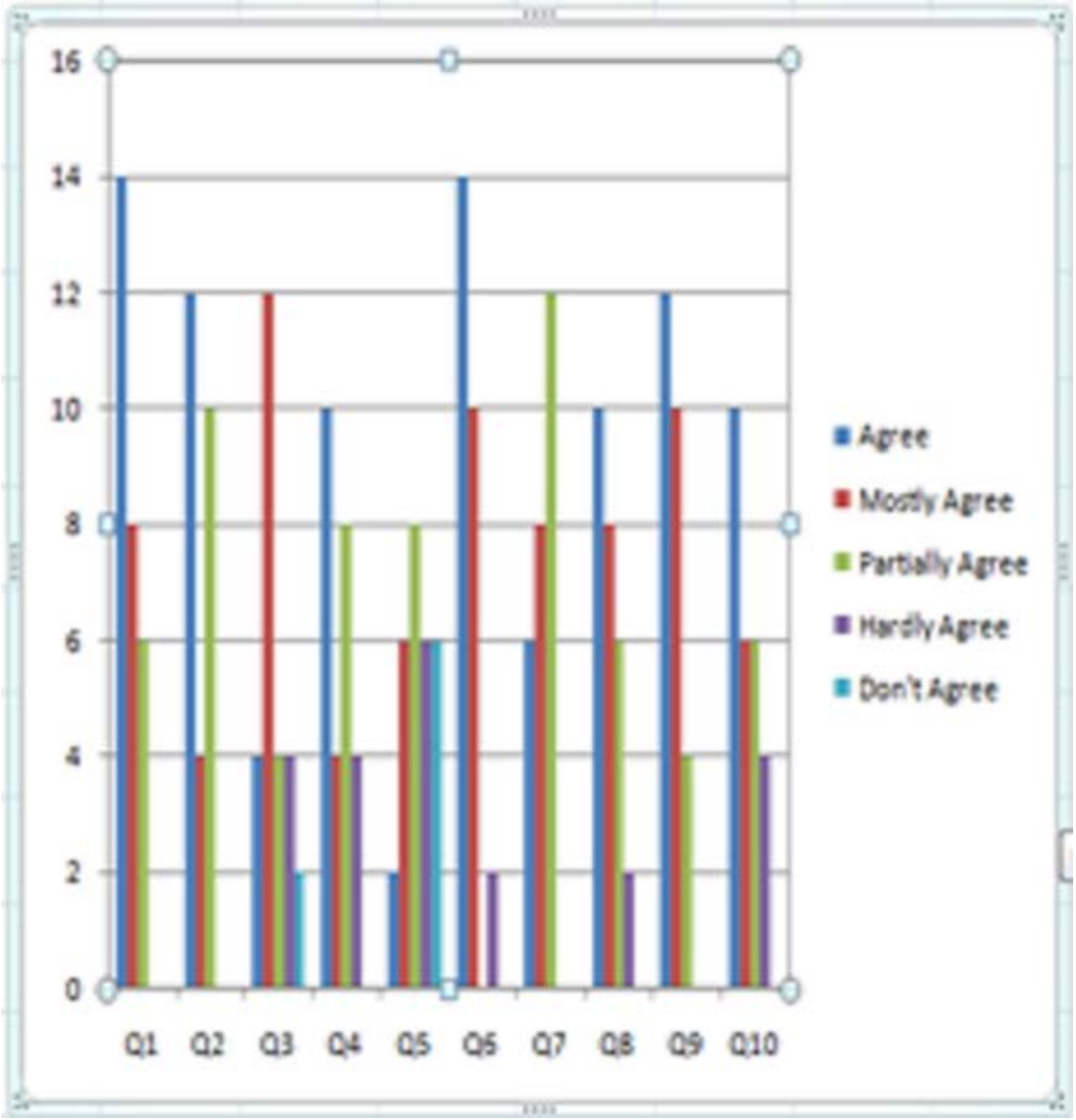
Table 17: Test Case-17

Test Case Name	Home Button
Test Case Number	18
Description	Testing the Home Page Button for application user
Preconditions	Application must be open and it must be on About Page
Input	Click on the Home Page Button
Steps	First go to the About Page then click on Home Page Button
Expected output	Test to return to the Home Page should open
Results	Test to return to the Home Page opens

6.4. Testing And Evaluation Graph

This is the graph which we generated after taking feedback from 26 users and the questionnaire which we used for evaluation that is attached as “Appendix F”

Figure 14:Evaluation Graph



6.5. Summary

Testing not only maintains the software quality but also improves overall usability of the project. At different stages of development suitable testing techniques were used to ensure product works accurately and efficiently. All errors detected during testing were removed and the test cases were prepared and made part of this document for the future compliance.

Chapter 7

7. Conclusion and Future Work

7.1. Conclusion

The community service mobile app Summ-The-Search has four modules, implemented as data preprocessing; Summary generation, email, and data base history. Our product “Summ-The-Search” is an android Application that provides an online platform for generating text summarization. Users can search their desired services in different domains. By combining Cloud Computing technology, our android application will be facilitating the users in bridging the gap among online community.

7.2. Future Work

As this is a research project, so there is always room for improvement. An extension to this project can be added by taking it to the cloud computing platform. This will reduce the amount of processing on the user end. After doing all of the calculation on the server just a simple a response file is given to the user. It will also make a good impression on hardware where user has no more machine, dependency and capability issues. It can be helpful and leading stone toward abstract summary generation. As we just need to introduce the language sense in the application, so that self construction of sentences can happen in reality.

Appendix A

Use Cases Description

The various user classes identified the following Use Cases and primary actors for the Summ-The-Search :

Table 18: Actor-Use Cases

Actors	Use Cases
Android User	Enter Email Generate Summary Send Mail Maintaining History Update and Saving History Displaying About Enter Query Enter Number Of Pages Select Document Type Delete History

Table 19: Enter Email Id

Use Case ID:	1
Use Case Name:	Enter Email ID
Actors:	User
Description:	System has to display the home page at the opening of Application.
Preconditions:	Nil
Post conditions:	It will show options to go to any of 3 pages.(i)Main Page (ii)History Page (iii)About
Normal Flow (primary scenario):	User will open the application and will prompt home page having 3 choices.
Alternative Flows:	If displaying home is not prompting then user will close the application and run it again or will reinstall the application.

Table 20: Generate Summary

Use Case ID:	2
Use Case Name:	Generate Summary
Actors:	End User, Search Engine
Description:	System has to create the summary of the query entered by user depending upon no of links.
Preconditions:	Query, no of links and length should be provided and understandable.
Post conditions:	Summary will be generated in the form of text which can be further email(as a pdf format)or cancel.
Normal Flow (primary scenario):	User will enter query, no of links and length and summary will be generated in the form of text which can be further email(as a pdf format)or cancel.

Alternative Flows:	If any of given fields missing then system will prompt user to provide it again.

Table 21: Send Mail

Use Case ID:	3
Use Case Name:	Send Mail
Actors:	User
Description:	System will send that summarized file in the pdf format.
Preconditions:	Summary generated text.
Post conditions:	Summarized file would be sent and system will return to home page.
Normal Flow(primary scenario):	Summary text is generated and now that summarized file would be email after that system will return to home page.
Alternative Flows:	If user is not satisfied with results then it can cancel that and will give new query.

Table 22: Maintaining History

Use Case ID:	4
Use Case Name:	Maintaining History
Actors:	User Database
Description:	System has to display the history that user has generated the summaries through this application.
Preconditions:	Nil
Post conditions:	It will be showing the history.
Normal Flow (primary scenario):	After opening the application (home page) from where user will select history option and system will show history (all summaries).

Table 23: Update And Saving History

Use Case ID:	5
Use Case Name:	Update And Saving History
Actors:	User Database
Description:	System has to update and save history after every newly generated summary or if any history is deleted.
Preconditions:	At least one or more summaries must be present in the history.
Post conditions:	It will show the updated history and will save its state.
Normal Flow (primary scenario):	Summaries must be generated by user and user will delete or clear the present history and after system will show the updated history and will save its current state.
Alternative Flows:	If user is not deleting or clearing the complete history the system will save its previous state.

Table 24: Displaying About

Use Case ID:	6
Use Case Name:	Displaying About
Actors:	User
Description:	System will display the About page which will be giving the information about the application and developers.
Preconditions:	Nil
Post conditions:	It will be giving the information about the application and developers and afterwards user can return to home page.
Normal Flow (primary scenario):	User will open the application and will select the about page from home page and system will display that page which will be giving the information about the application and developers.

Table 25: Enter Query

Use Case ID:	7
Use Case Name:	Enter Query
Actors:	User Database
Description:	User will enter his query for which he want to generate the summary.
Preconditions:	User must be on main page; as at opening user is at home page from where he will select main page for entering the query.
Post conditions:	As user provided his query then he will provide the no of pages of summary he want.
Normal Flow (primary scenario):	At opening of the application user will be at home page from where he will select main page for entering query, when user done with query then he will provide the no of pages of summary he wants.
Alternative Flows:	If user misses the query then summary would not be generated and message would be prompt.

Table 26: Enter No Pages

Use Case ID:	8
Use Case Name:	Enter No Of Pages
Actors:	User Database
Description:	User will enter the no of pages for summary.
Preconditions:	User must have entered the query.
Post conditions:	After the user has provided his query and no of pages then summary would be generated.
Normal Flow(primary scenario)	At opening of the application user will be at home page from where he will select main page for entering query and no of pages, when user done with that then summary would be generated.

Alternative Flows:	If user misses the no of pages then summary would not be generated and message would be prompt.

Table 27: Delete History

Use Case ID:	10
Use Case Name:	Delete History
Actors:	User Database
Description:	User will delete the complete history.
Preconditions:	At least one or more than one searches must be made in order to delete the history.
Post conditions:	History would be deleted completely and updated history would be saved.
Normal Flow (primary scenario):	User will go to the history page then will delete completely history.
Alternative Flows:	If nothing record of searches is present in the history then user will go back to home page.

Appendix B

Class Diagram Description

Pre-Processing

This class contains 4 functions AdsRemoval(), TableRemoval(), ImageRemoval() and ReferenceceRemoval().AdsRemoval() will remove the all Ads appearing in the links, ImageRemoval() will remove the all images appearing in the links, ReferenceceRemoval() will remove the all references appearing in the links TableRemoval() will remove the all tables appearing in the links.

Extraction

This class contains 4 functions keywordCheck(),CheckLength(), SenPosition() and TermFrequency().In all above functions has to extract data, every functions has its own way to extract data as SenPosition() ;on the basis of position of sentences;normally in the documents important sentences are beginning or in conclusion paragraph.

Summary

This class contains 2 functions CreateSummary(),CancelSummary(). CreateSummary() will make a file of generated/extracted data through input from the class extraction and Pre-Preprocessing.

Email

This class contains 3 functions SendMail(),CheckDocType() and CheckSub().SendMail() will send mail as per given email address.CheckDocType() for checking document type otherwise message will be prompt. CheckSub() will check subject is provided or not if not message will be prompt.

History

This class contains 2 functions SaveAndUpdate(), DeleteHistory().SaveAndUpdate() will enter that query in the history and will also be updated if user send that file to different email address.DeleteHistory() will delete the history in a way that deletes the whole history.

Appendix C

Activity Diagram Description

By opening the App System will display Home screen which will consists of three options About, History and Main Page, User will select either of the options.

If user selects “About” option then system will give the information about the app by Displaying About.

If user selects “History” option then system will give the History of all the searched queries by performing the Display History action. User can delete his searched history which will be saved and updated by system.

If user selects “Main Page” Option then system will display Main Page after that user will enter the query, Number of pages and the length of the summary and upon clicking the Search and Summarize by the user, system will generate the Summary. User can either cancel or email the summary. If user cancels the summary then system will move to Display Home Page. In case of Email Summary system will show Email Page then user will enter the Email ID, Subject. In case of incorrect Email ID User will enter the ID again otherwise system will not send the mail or will direct to another (if email address was added of some other person).

Appendix D

Sequence Diagram Description

Summary object will call its own function CreateSummary(), CreateSummary() will make a file of generated/extracted data through input from the class extraction and Pre-Preprocessing.

Pre-Processing has 4 functions AdsRemoval(), TableRemoval(), ImageRemoval() and RefereneceRemoval().AdsRemoval() will remove the all Ads appearing in the links, ImageRemoval() will remove the all images appearing in the links, RefereneceRemoval() will remove the all references appearing in the links TableRemoval() will remove the all tables appearing in the links and a these all functions would be called by Summary object.

Whereas Extraction has 4 functions keywordCheck(), CheckLength(), SenPosition() and TermFrequency().In all above functions has to extract data and every functions has its own way to extract data as SenPosition() ; on the basis of position of sentences; normally in the documents important sentences are in the beginning or in conclusion paragraph and these all functions would be called by Summary object.

Email object will call its 2 functions CheckSub() and SendMail().

History Object has 2 functions SaveAndUpdate(), DeleteHistory(), which are called by the Summary Object. SaveAndUpdate() will enter that query in the history and will also be updated if user send that file to different email address. DeleteHistory() will delete the history in a way that deletes the whole.

Appendix E

User Manual

Overview of User Manual for Summ-The-Search

This is the user manual for mobile app “Summ-The-Search”. Summ-The-Search is an android application which summarizes the multiple web pages and generates the one precise summary.

Following are the main functionalities of this application:

1) Summary generation
2) Emailing the summarized summary
3) History maintaining

Details:

1) **Opening of the application**

   Run the application, and at the very first glance this “Home Page” will appear

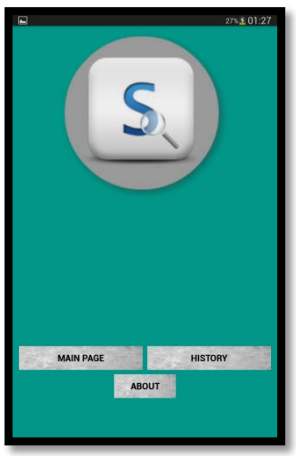


Figure 15: User Manual-Home Page

Home Page will consists of three options:

I. Main Page
II. History
III. About

2) Generation of the Summary

To generate the summary select Main Page. And this “Main Page” will appear.

:

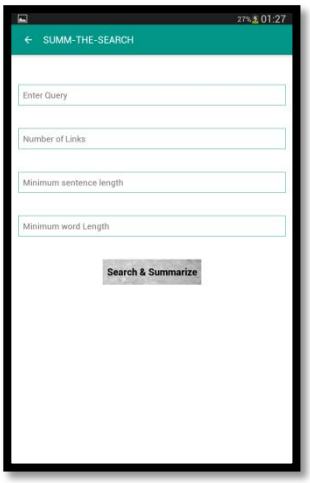


Figure 16: User Manual-Main Page

And do the following:

I. Enter query (query of which you want summary).

II. Provide the number of links to generate the summary.

III. Percentage of the Summary (how much of total data taken from web pages).

IV. Click the button “Search & Summarize” and it will show the generated summary from there user can email it (if he likes) otherwise cancel it.

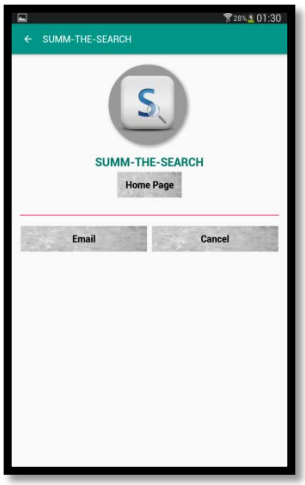


Figure 17: User Manual-Summary View Page

3) Mailing the generated Summary

If user cancels, it will go back to “Main Page”, if user presses the button “Email” then this “Email Page” will appear.

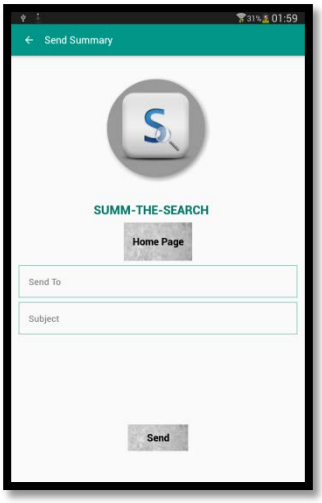


Figure 18: User Manual-Email Page

User can do the following :

I. Go back to Home Page by pressing Button “Home Page”.

II. Give recipient email address (To whom you want to send it).

III. Suitable subject for the summary.

After providing above info, now click on the Send Button to send summary as an email to recipient in pdf format and it will show as.

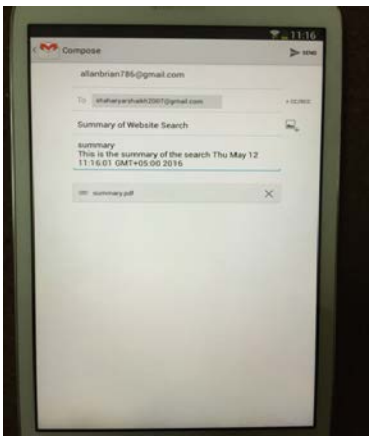


Figure 19:User Manual-Email sent view

4) History Checking

To check the history of your searched queries select History Button from “Home Page”

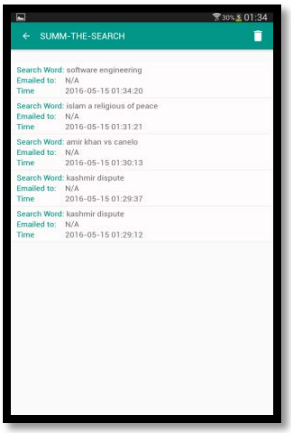


Figure 20: User Manual-History Page

On History Page you can keep last 10 searches and can delete complete history

5) About

To know about the App select Button “About” from “Home Page” and it will look like this and from this you can read about Summ-The-Search.

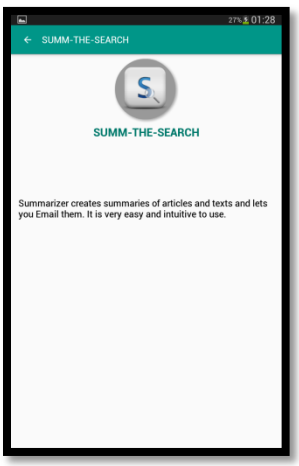


Figure 21:User Manual-About

Appendix F

Questionnaire

Summary Evaluation Questionnaire

Name: _____

	Quality Evaluation Question	agree	mostly agree	partially agree	hardly agree	don't agree
1.	To what degree of satisfaction does the resultant summary capture the important information from the source text?					
2.	To what degree of satisfaction can the resultant summary be considered as effective?					
3.	How readable is the summary? (Dangling references, incomplete sentences, premature sentence termination and redundancy).					
4.	How comprehensive is the summary? (Ability to interpret the meaning).					
5.	How much does the summary cover the unimportant content of the original text?					
6.	What is the over all impression of the summary?					

	Grammaticality, Cohesion and Coherence Evaluation Questions	agree	mostly agree	partially agree	hardly agree	don't agree
1.	Do sentences in the summary fit in with their surrounding sentences?					
2.	Is the content of the summary expressed and organized in an effective way?					

	Performance Evaluation Questions	agree	Mostly agree	Partially agree	hardly agree	Don't agree
1	Product is easily useable by common users.					
2	Response time of the application is appropriate.					

Remarks:

_____

_____

_____

Glossary

Android	An operating system designed for mobile devices (i.e. cell phones, tablet computers) by Google, Inc
App	Application
API	Application Programming Interface
Black box Testing	Testing emphasizes on the external behavior of the software entity
CO	Constraints
DBMS	Database Management System
EMAIL	Electronic mail
FRs	Functional Requirements
GUI	Graphical User Interface
HTML	Hyper Text Markup Language
MCS	Military College of Signals
NFRs	Non Functional Requirements
NUST	National University of Science and Technology
N/A	Not Applicable
OS	Operating System
PDAs	Personal Digital Assistants
Summ-the-Search	Name of the Application
SRS	Software Requirements Specification
UML	Unified Modeling Language
White Box Testing	Testing emphasizes on the internal behavior of the software entity
pdf	portable document format
24/7	All the time(24 hours 7 days)

Table 28: Glossary

Bibliography

[1] Sanda Harabagiu and Finley Lacatusu, “Using Topic Themes for Multi-Document Summarization”, University of Texas at Dallas.

[2] Suneetha Manne and S. Sameen Fatima“A Feature Terms based Method for Improving Text Summarization with Supervised POS Tagging” , International Journal of Computer Applications (0975 – 8887), June 2012, Volume 47– P.No.23

[3] Chin-Yew Lin and Eduard Hovy, “From Single to Multi-document Summarization:”, Proceedings of the ACL conference, PA. 2002 , pp. 457–464.

[4] Adam L. Berger, Pittsburgh, Vibhu O. Mittalt, “OCELOT: A system for summarizing web pages”.

[5] Orkut Buyukkokten ,Hector Garcia-Molina and Andreas Paepcke, “Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices” Stanford University.

[6] The Unified Modeling Language Reference Manual. James Rumbaugh, Ivar Jacobson, AND Grady Booch. 1998. P.81. **ISBN:** 0-201-30998-X.

[7] The Unified Modeling Language Reference Manual. James Rumbaugh, Ivar Jacobson, AND Grady Booch. 1998. P.87. **ISBN**: 0-201-30998-X.

[8] The Elements of UML™ 2.0 Style. Scott W. Ambler. May 2005. **ISBN:** 9780521616782

[9] Anton Leuski, Chin-Yew Lin, Eduard Hovy University of Southern California

[10] Álvaro Mendes Barbosa Pompeu Fabra University “Overview of text summarization in the context of information retrieval and interpretation: Applications for web pages summarization “ Audiovisual Institute research paper.