

Final Year Project (Mid Evaluation Report)

“SUMMIX”

AN AUTOMATIC TEXT SUMMARIZER

Team Members:

M. Muneeb Raza i16-0128

M. Shahzeb Ali i16-0246

Supervised By:

Dr. Omer Beg

**Department Of Computer Science
National University of Computer and Emerging Sciences
Islamabad, Pakistan**

Anti-Plagiarism Declaration:

Summix is the sole contribution of the authors and no part here of has been reproduced on as it is basis (cut and paste) which can be considered as Plagiarism. All referenced parts have been to argue the idea and have been cited properly. The work presented in the report is our own and it has not been submitted/presented previously to any other institution or organization. We will be responsible and liable for any consequence if violation of this declaration is determined.

Date: _____

Student 1

Student 2

M.Muneeb Raza (i16-0128)

M.Shahzeb Ali (i16-0246)

Signature: _____

Signature: _____

Supervisor (Faculty)

Dr. Omer Beg

Signature: _____

Table of contents

1.	Introduction.....	
1.1.	Executive Summary.....	3
1.2.	Project Vision.....	3
1.3.	Goals and Objectives.....	4
2.	Use cases Diagrams.....	4
2.1.	High level Use cases.....	4
2.2.	Use case diagram	6
2.3.	Expanded Use cases.....	6
3.	Domain Model.....	11
4.	Class Diagram.....	12
5.	Sequence Diagrams.....	13
6.	Proposed Architecture.....	17
7.	System Specifications.....	18
8.	Progress Till Mid Evaluation.....	19

1. Introduction

Executive Summary:

In this modern era, Technology has revolutionized our lives. There are millions of Websites and Apps which we use every day and each one of them has its **Privacy Policy** and **End User Agreements**. But rarely we read any and instead we just tick by clicking in a hurry and agree to the Privacy Policies without even knowing about what is written in the policy agreement. Due to this we are seriously at a risk of losing our personal data and in result we may agree to something which may become a threat to our data in later stages.

There are many Tools which summarizes long texts of Privacy Policies into shorter paragraphs but now the question arises that if a user is not wasting time on reading the long texts then will he even bother to read summarized paragraphs? Definitely No!

And moreover, neither of them make crisps of Bullet Points and Branch Tree Visualization for easy understanding of important aspects.

Summix is a self-contained product being developed to fulfill the needs of people to whom their privacy and data really matters. It will enable its users to automatically generate crispy, short, concise and precise **Bullet Points** (of the important information only) out of the Privacy Policies and User Agreements alongside preserving the true meaning of the Policies. Moreover, It will enable its users to **visualize** important aspects of the Privacy Policies in the form of a Branch Tree.

Project Vision:

To develop a Mobile App which can efficiently generate short, meaningful, precise and concise Bullet Points out of lengthy Policy Statements and End User agreements.

Goals and Objective:

To develop a Deep Learning Model of Natural Language Processing to Summarize Privacy Policies into short crisp Bullet Points.

2. Use Case Diagrams

High Level Use Cases:

Use case: Bullet Points Generation

Actor: User

Actor Type: Primary

Goal: To generate short and meaningful Bullet Points.

Description: Generating concise and precise Bullet Points, of the most important information, out of lengthy privacy policies and user agreements.

Use case: View Bullet Points

Actor: User

Actor Type: Primary

Goal: To view generated Bullet Points.

Description: User may be able to view and read the generated Bullet points and make an useful understanding about the most important information of the privacy policy.

Use case: Branch Tree Generation

Actor: User

Actor Type: Primary

Goal: To generate a Branch Tree from the generated Bullet Points.

Description: Generating a Branch Tree with the help of the Bullet Points, to make user enable to understand the important crux of the Privacy Policy in a pictorial tree form.

Use case: Branch Tree Visualization

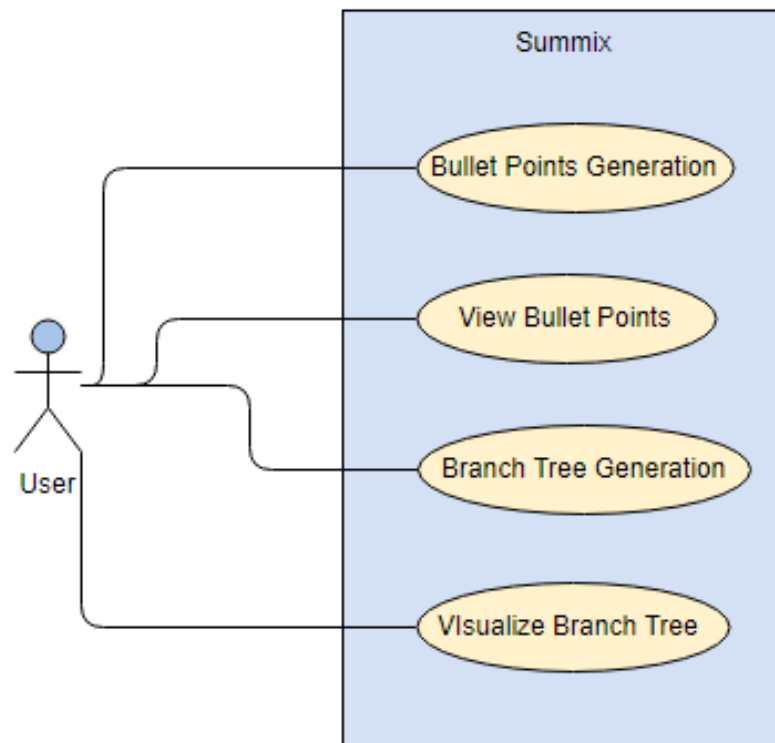
Actor: User

Actor Type: Primary

Goal: To visualize Bullet Points with the help of Branch Tree.

Description: User may get a pictorial visualization of the generated Bullet Points in the form of a Branch Tree.

Use Case Diagram



Extended- High Level Use Cases

Scope	Summix-An Automatic Text Summarizer
Use Case	Bullet Points Generation
Level	Primary Goal
Primary Actor	User,Summix
Goal	To generate short and meaningful Bullet Points.

Stakeholders and Interests	User: wants to view Bullet Points instead of long text. Summix: wants to generate short Bullet points.
Preconditions	<ul style="list-style-type: none"> Summix has to be well trained on ML,NLP models. There should be proper Training Data and Testing Data .
Success Guarantee	<ul style="list-style-type: none"> Bullet Points have been generated.

Actor Action	System Responsibility
1. Opens Summix	2. Prompts user to choose a privacy policy.
3. Chooses to Generate Bullet Points	4. Promptly generate Bullet Points.

Scope	Summix-An Automatic Text Summarizer
Use Case	View Bullet Points
Level	User Goal
Primary Actor	User
Goal	To view generated Bullet Points.
Stakeholders and Interests	User: wants to view the generated Bullet Points. Summix: wants to promptly display the Bullet Points.

Preconditions	<ul style="list-style-type: none"> • User knows how to use Summix • Summix has already generated the Bullet Points out of large privacy policies
Success Guarantee	<ul style="list-style-type: none"> • Bullet Points are displayed for the user.

Actor Action	System Responsibility
5. Opens Summix	6. Prompts user to choose a privacy policy.
7. Chooses to Generate Bullet Points	8. Promptly generate Bullet Points.
9. Chooses to view the generated points.	10. The Bullet Points are displayed to the user.

Scope	Summix-An Automatic Text Summarizer
Use Case	Branch Tree Generation
Level	Primary Goal
Primary Actor	User, Summix
Goal	To generate a Branch Tree from the generated Bullet Points.
Stakeholders and Interests	<p>User: wants to visualize Bullet Points in the form of a tree.</p> <p>Summix: wants to make a Branched Tree for as many Bullet Points as possible.</p>

Preconditions	<ul style="list-style-type: none"> • User has knowledge of Summix. • Summix knows how to make a Branch Tree Visualization. • Summix has already generated Bullet Points.
Success Guarantee	<ul style="list-style-type: none"> • Branched Tree is generated with the help of Bullet Points.

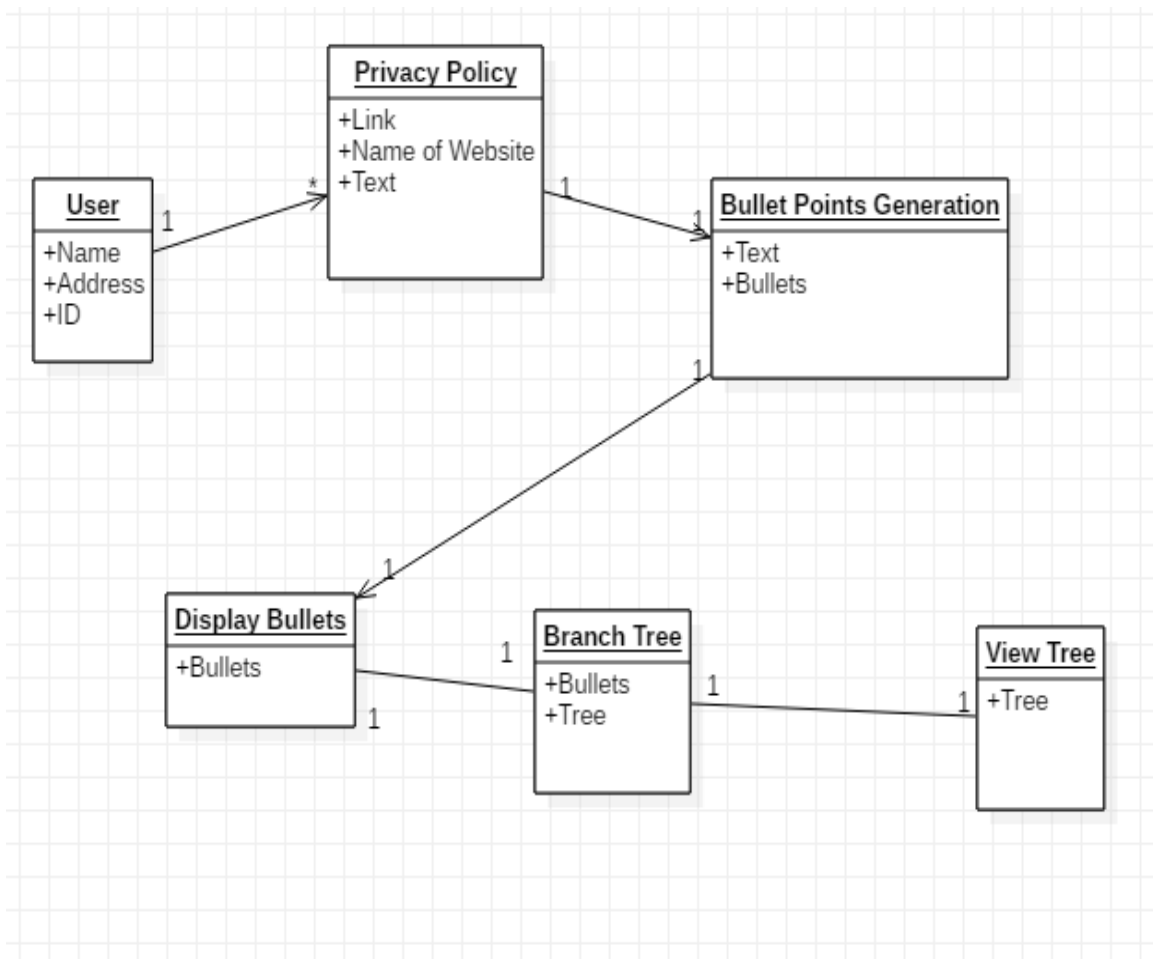
Actor Action	System Responsibility
11. Opens Summix	12. Prompts user to choose a privacy policy.
13. Chooses to Generate Bullet Points	14. Promptly generate Bullet Points.
15. Chooses to view the generated points.	16. The Bullet Points are displayed to the user.
17. Chooses To Generate Branch Tree.	18. Promptly generates Branch Tree.

Scope	Summix-An Automatic Text Summarizer
Use Case	Branch Tree Visualization
Level	User Goal
Primary Actor	User
Goal	To visualize Bullet Points with the help of Branch Tree.
Stakeholders and Interests	User: wants to view Important information of the privacy policy in the form of a Tree.

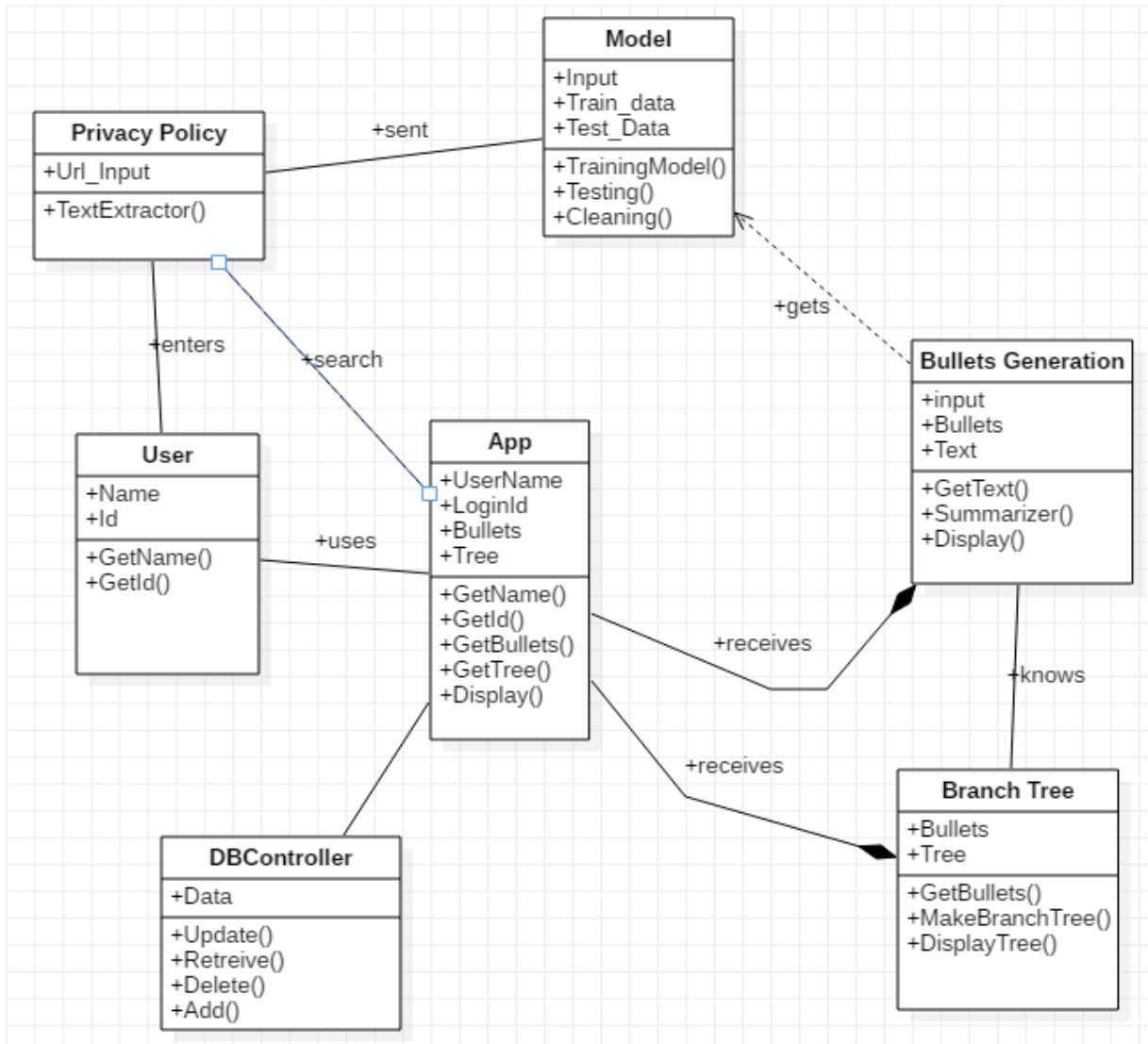
Preconditions	<ul style="list-style-type: none"> • User knows how to visualize Tree on the App(Summix) • Summix has generated Bullet Points and its Branched Tree
Success Guarantee	<ul style="list-style-type: none"> • User Visualizes Bullet Points in the form of a Branch Tree.

Actor Action	System Responsibility
19. Opens Summix	20. Prompts user to choose a privacy policy.
21. Chooses to Generate Bullet Points	22. Promptly generate Bullet Points.
23. Chooses to view the generated points.	24. The Bullet Points are displayed to the user.
25. Chooses To Generate Branch Tree.	26. Promptly generates Branch Tree.
27. . Chooses to view branched tree.	28. The Bullet Points are displayed to the user in the form of a Branch Tree.

3. Domain Model



4. Class Diagram



5. Sequence Diagrams

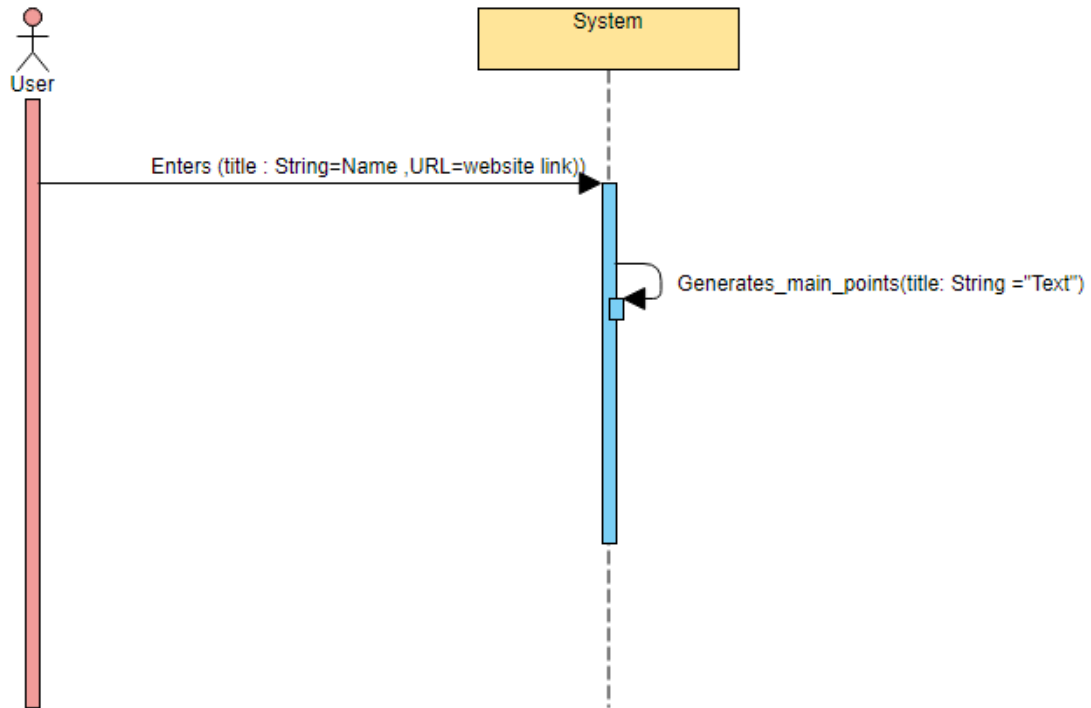


Fig 1.5 Bullet Points Generation

Preconditions	<ul style="list-style-type: none">• System has to be well trained on ML,NLP models.• There should be proper Training Data and Testing Data .
Success Guarantee	<ul style="list-style-type: none">• Bullet Points have been generated.

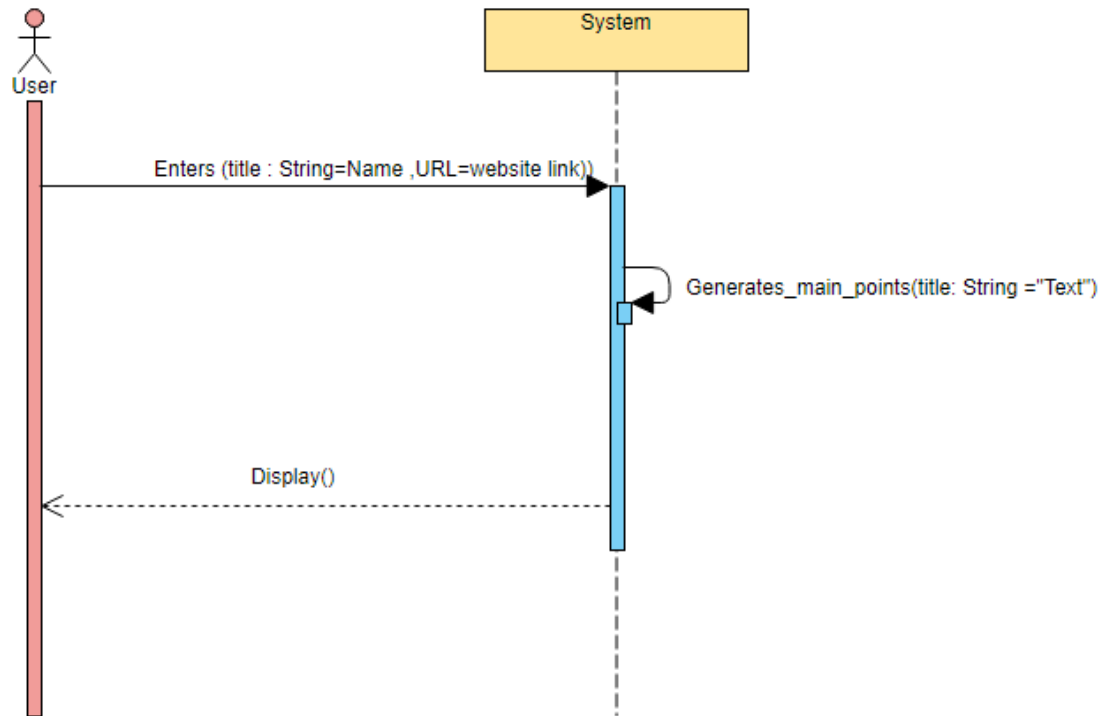


Fig 1.6 Display Bullet Points

Preconditions	<ul style="list-style-type: none"> • User knows how to use System • System has already generated the Bullet Points out of large privacy policies
Success Guarantee	<ul style="list-style-type: none"> • Bullet Points are displayed for the user.

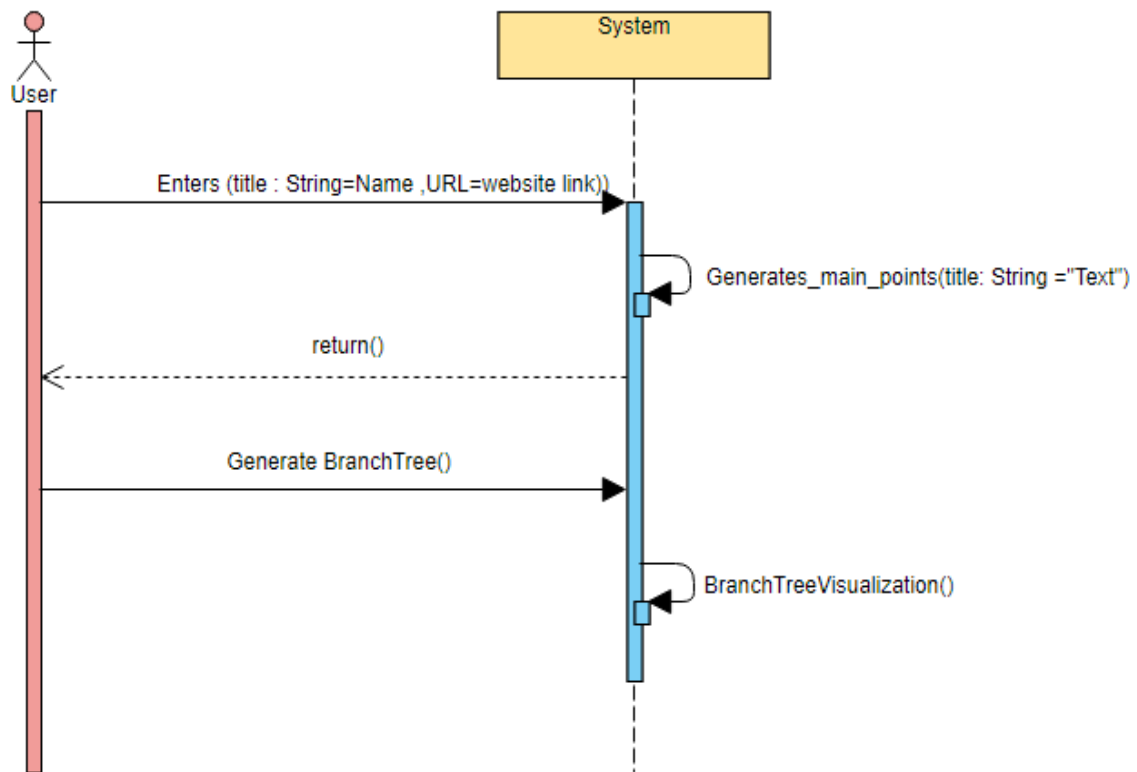


Fig 1.7 Branch Tree Generation

Preconditions	<ul style="list-style-type: none"> • User has knowledge of System. • System knows how to make a Branch Tree Visualization. • System has already generated Bullet Points.
Success Guarantee	<ul style="list-style-type: none"> • Branched Tree is generated with the help of Bullet Points.

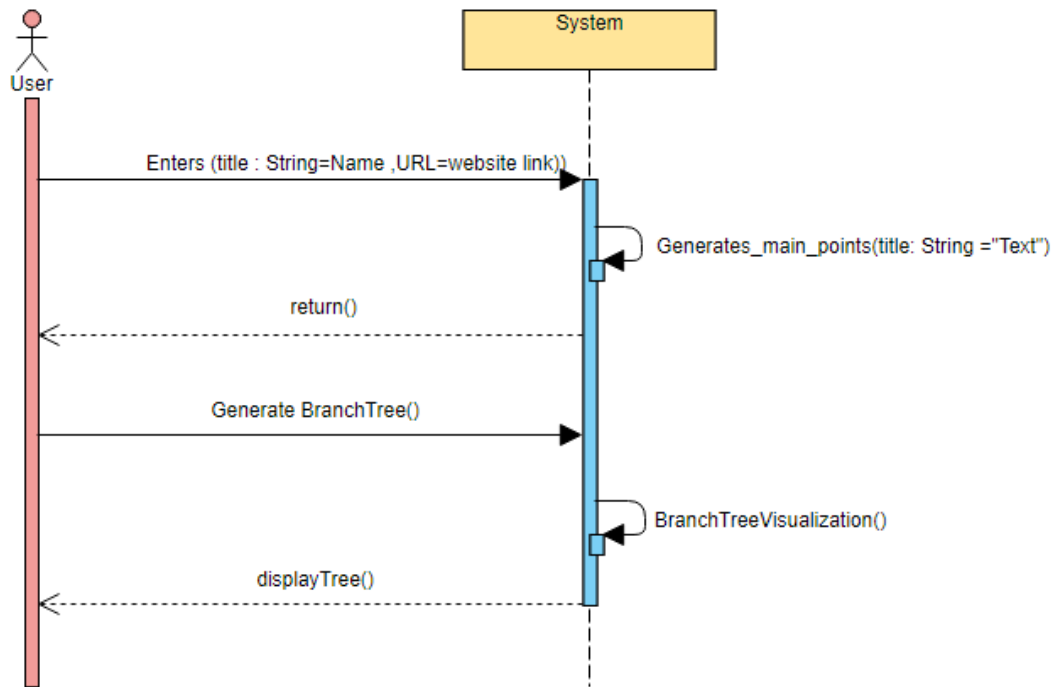


Fig 1.8 Branch Tree Visualization

Preconditions	<ul style="list-style-type: none"> • User knows how to visualize tree on the System • System has already generated Bullet Points and its Branched Tree
Success Guarantee	<ul style="list-style-type: none"> • User Visualizes Bullet Points in the form of a Branch Tree.

6. Proposed Architecture and Work Flow

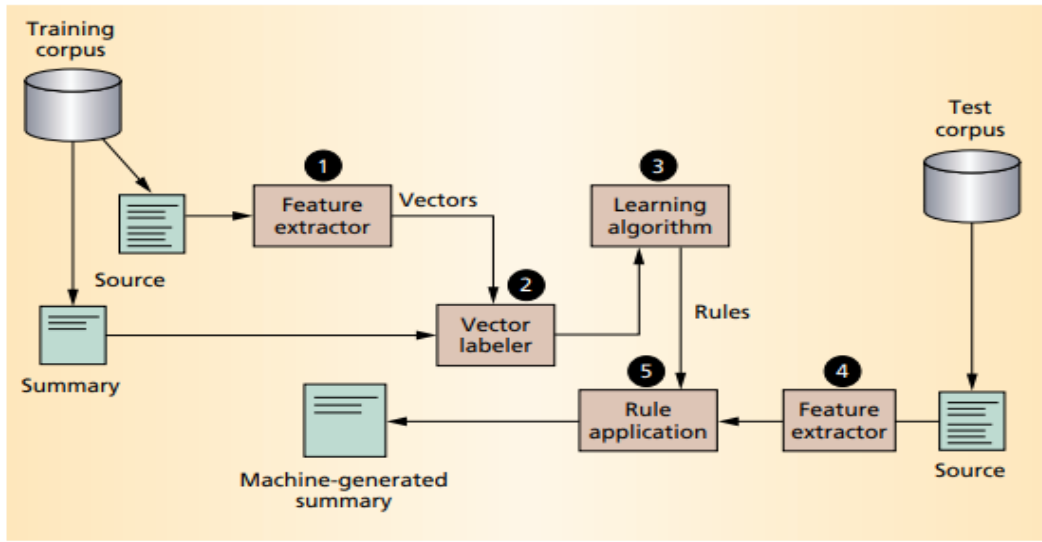
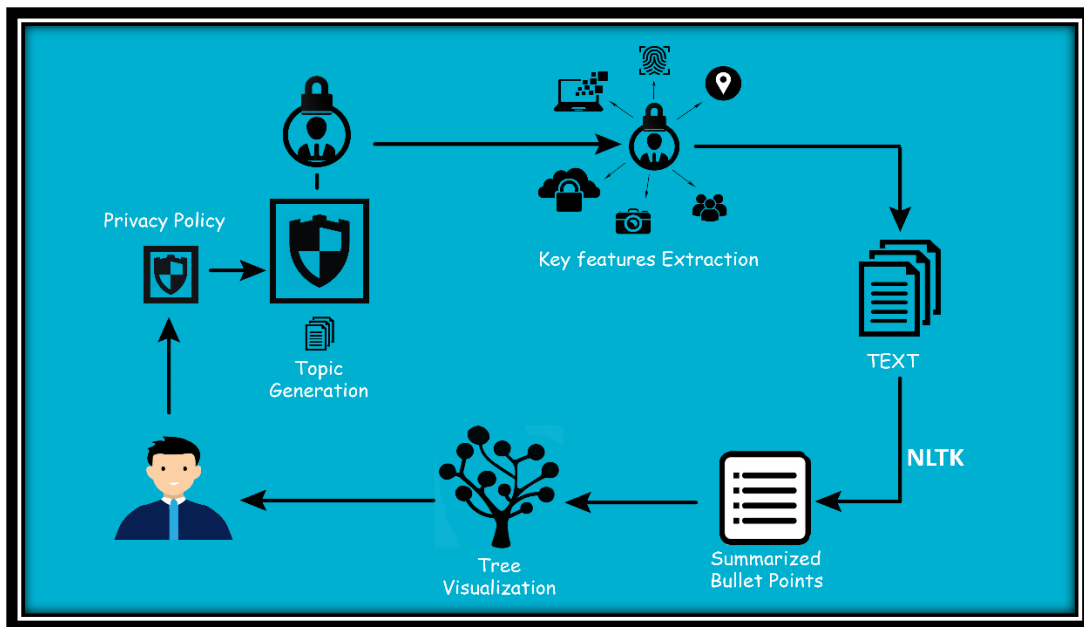


Figure 2. A classifier that learns how to summarize. (1) During training, a vector of features represents each sentence in the source. (2) The classifier labels each vector according to similarity of content between the sentence and the abstract. It then feeds the set of training examples to (3) a learning algorithm that learns the classification rules for determining whether or not a sentence should be part of the summary. During testing, the summaries are absent. Instead, the classifier (4) turns each sentence from the test corpus into a feature vector and (5) matches it against the learned rules to generate the extract.



7. Supplementary Specifications

Introduction:

The Supplementary Specification lists the requirements that are not readily captured in the use cases of the use case model. The supplementary specifications and the use case model together capture a complete set of requirements on the system.

Functional Requirements:

- The system will generate concise, short and meaningful crisps of Bullet Points out of long Privacy Policies.
- The system will enable users to visualize the Bullet points in a Branch Tree form.

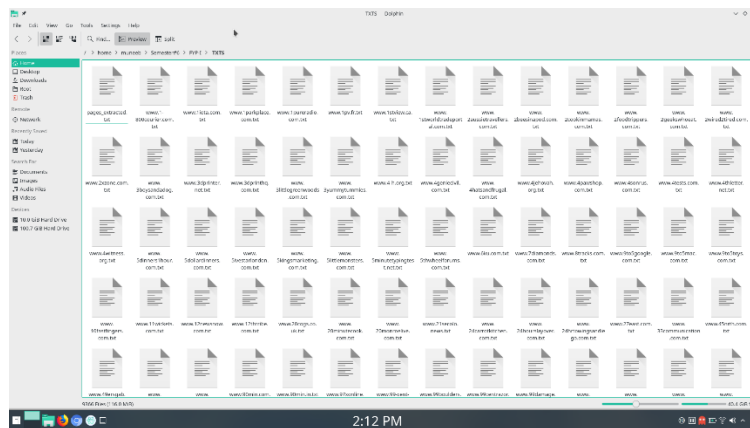
Non Functional Requirements:

- **Security**
 - All usage require user authentication. The users will have limited access to the Data Base of the system.
- **Robustness**
 - The system will ensure fast and robust response to the user ensuring minimal lags.
- **Usability**
 - A user friendly interface, easy to interact.
 - Help section entertaining queries of users.
 - Users Feedback
- **Reliability**
 - The application will ensure minimal failures.
 - If somehow the app crashes, recoverability will be ensured.
 - If you have downloaded the app, you can use it from wherever you want.

8. Progress till Mid Evaluation

Since there is no open-source dataset of privacy policies over the internet except of only four datasets. But they comprises of very small amount of privacy policies. i.e OPPO-115 contains only 115 privacy policies of websites, APP-350 contains only 350 privacy policies of android apps and Since we need thousands of privacy policies of different apps/websites therefore we had to scrap data and make our own data set.

- We have built our own scraper and till now it has scraped more than 30 thousand privacy policies. Moreover, also we are cleaning the scraped data so that it increases accuracy, becomes garbage free and usable etc.



- We have started summarizing the privacy policies using NLTK but we are not that much successful till now as our results are in paragraphs and also summarizing is not performed well but we are working on it and we are learning and studying models to increase our accuracy.
- We have build the Prototype of our Mobile App-“Summix”