Informatics Institute of Technology

Module Name -: Individual Report

Module Code -: 5COSC021C.Y

“ THRUVOX ”

Full Name: Thilakesan Vishnujan

UOW ID: w1957448

Group Members

Name	UoW number	IIT ID
Thilakesan Vishnujan	w1957448	20211403
Jegannathan Keyuran	w1989365	20222342
Mohan Nidwin Ranz	w1985749	20222240
Sophia Blesslyne Fernando	w1998871	20211203
Hafeelul Ilahi Ahamed Haseef	w1985566	20221789

Declaration Page

I solemnly declare that the content presented in this individual report is the result of my own work and research. The information provided is accurate and authentic to the best of my knowledge and it has not been submitted before nor is it being submitted for any other degree Programs. This report has been prepared as part of SDGP at University of Westminster, UK.

Abstract + Keywords

Start a language changing trip with our amazing app, carefully made for turning English into Tamil PDF and shortening it. This report shows the great things about technology. It explains how a tool works that easily moves past language problems and makes information simple.

Look at how new natural language processing (NLP) methods and top-notch machine learning models are combined to change our app's way of translation or summarizing. Beyond the complicated parts, explore a user-focused design that promises an easy and fun experience. Report shows how the combination of new ideas and everyday use can turn a complex PDF into simple translations.

Work with the comparison study that shows how good our app is, and it's better than what we have now. Be part of a big change in how we see and talk with many languages. We are opening up to a world where language is no longer an obstacle but connects people together. design creates a translation and summarization experience from English to Tamil that has never been seen before. This report calls you, asking to explore the ideas of innovation that promise to change how we understand and speak through writing.

Keywords: Language Technology, Translation Innovation, Summarization Excellence, NLP Advancements

Acknowledgement

First off we would like to thank our parents and family for their support, continuously inspiring us to be the best versions of ourselves and motivating us through all hardships. We would also like to express our sincere gratitude to Mr. Banuka Athuraliya and the entire Software Development Group Project module team for always making time to attend to each and every problem we have encountered throughout this module, the support they have given and for their teachings. We would also like to extend our gratitude to Miss. Sapna Kumarapathirage, our mentor for always providing us with valuable insight and always directing us the right way when we were facing challenges. A special thanks should also go out to all those who took their time to fill out the questionnaire and everyone who has listened and answered to our countless queries regarding different aspects of the project.

Table of Contents

Table of Figures

List of Table

Abbreviations Table

Table 1 Abbreviation table

Abbreviations	Explanation
NLP	Natural Language Processing
AI	Artificial Intelligence
NLTK	Natural Language Toolkit
FR	Functional Requirement
NR	Non-Functional Requirement
ANN	Artificial Neutral Network
API	Application Programming Interface
CNN	Convolution Neural Network
FDD	Feature Driven Development
GDPR	General Data Protection Regulation
ML	Machine Learning
OOD	Object-Oriented Design
OOP	Object-Oriented Programming
PRINCE	Projects in Controlled Environments

SDGP	Software Development Group Project
SDLC	Software Development Life Cycle
SRS	System Requirement Specification
UI	User Interface
UX	User Experience
WBS	Work Breakdown Structure

Chapter 4: System Requirements Specification (SRS)

4.1 Chapter Overview

This chapter is mainly focused on how the web application is expected to perform and the functionalities of the software. This chapter will include an analysis of the stakeholders which will be shown by using an onion model, the techniques which will be used for elicitation purposes, use case diagrams will be used in order to indicate the high-level functions and the scope of the system, functional and nonfunctional requirements in order to describe the functions performed by the web application and to describe the quality attribute of the software.

4.2. Stakeholder Analysis

4.2.1 Onion Model

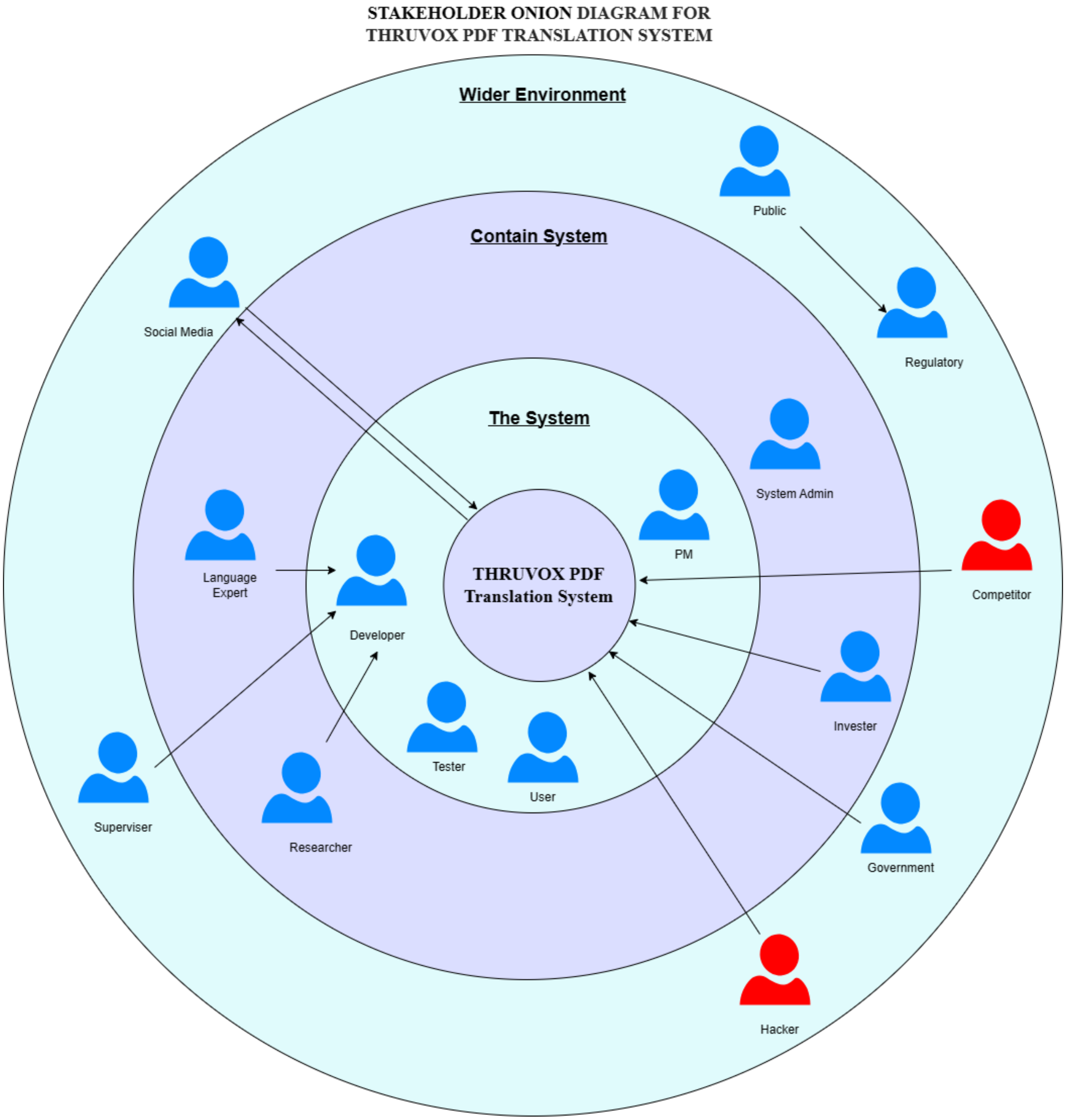


Figure 0.1 Onion Model diagram

4.2.2 Stakeholder Descriptions.

Table 2 Stakeholder description

Stakeholder	**Viewpoint**
Functional beneficiary	
User	Those who use the developed system for their needs. eg:- education, personal.
Financial beneficiary	
Investors	Invests funds for production and makes profits.
Social beneficiary	
Public	Public may use the product to get relevant information in a short time.
Operational beneficiary	
Project managers	Project managers communicate with other stakeholders to make sure the project is completed to the specified quality standards.
Product developers	Develop the system.
System Admin	Wants to monitor the web application and the system to maintain the standards and objectives of the project
Negative Stakeholders	
Competitors	Wants to design and implement a better web application with advanced features which will eventually reduce the demand for Thruvox app.
Hacker	Wants to invade the system and extract information from the system and damage the system

Regulatory	
Regulator	Wants to make sure the system works as expected and does not input any false information
Experts	
Language Expert	Experts ensure the quality and accuracy of translations.
Neighboring systems	
Social Media	Will expand the publicity of the application through sharing.

4.3 Selection of Requirement Elicitation Techniques/Methods

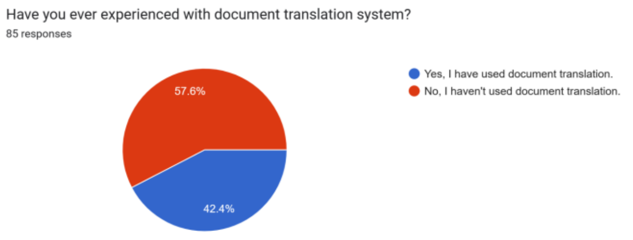
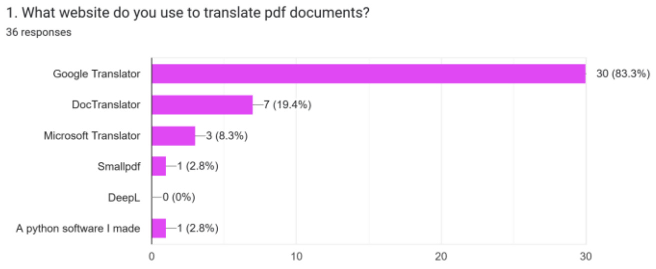
Requirement Elicitation Techniques	
Brainstorming	Method 01
Brainstorming sessions can be conducted with our own group members, colleagues and experts to collect ideas to expand the system and add extra features which will suit the customer for a user-friendly design. Friends, colleagues and experts participate in brainstorming sessions to consider new ideas on how to add features and improve the system. It is also a time-consuming process. a way to consume it, and can also be very effective; there are many giant ideas. go up when people pull together.	

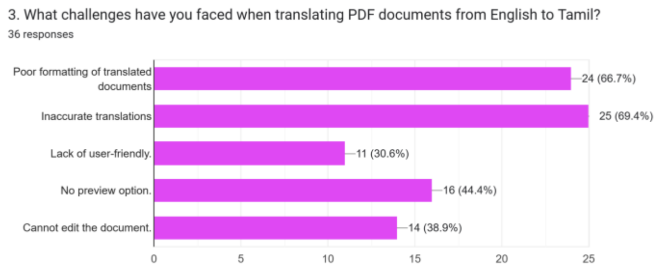
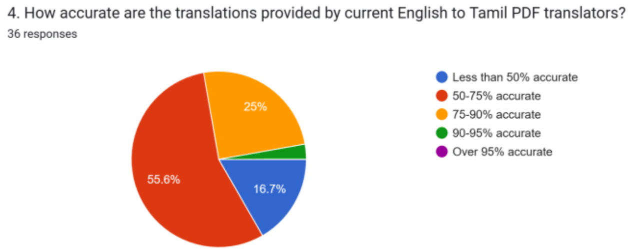
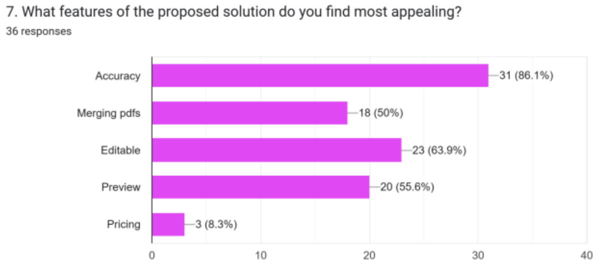
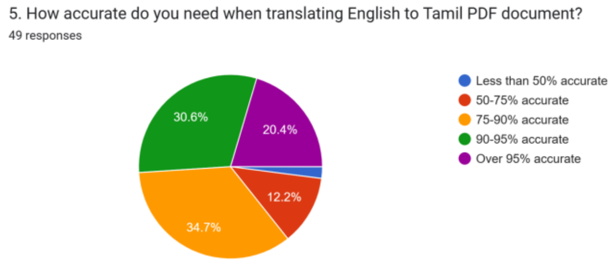
Literature Review	Method 02
Through the method of Literature review, concentration will be focused on comparing like systems. come within the purview of pdf translation. a valid library resource. Concern one that IEEE and other online resources can help to address. In the field, what features are offered by similar systems functionalities. with the strengths and weaknesses. This will provide a guide in developing Thruvox prototype what it should have and what is not suitable.	
Online questionnaire for General Public	Method 03
An online questionnaire will be sent out to the general public mainly aiming at the users who are expected to use the translation website after the implementation. Questionnaires are an efficient approach to gather requirements from a large number of stakeholders within a short duration of time. This questionnaire will help us to grasp the user expectations as well as the behavioral patterns of the user which will eventually be taken into consideration when implementing the THRUVOX. We can get to know if there's any other features that the users are expecting from this kind of web application.	
Online questionnaire for Domain Experts	Method 04
Having Interviews with Tamil professors and with domain translation experts can help to clarify the uncertainties before implementing the prototype application. Unlike other requirement elicitation techniques in an interview, we can query them instantly to get it clarified without much hustle. These personnel will get the opportunity to express and explain in detail the answers they provide in the questionnaire. This will be helpful to identify the view of students and the likelihood requirements of having certain functions in the system. A conversation with experts will give better direction and technical knowledge about the project. Depending on the situation, these interviews will be face to face.	

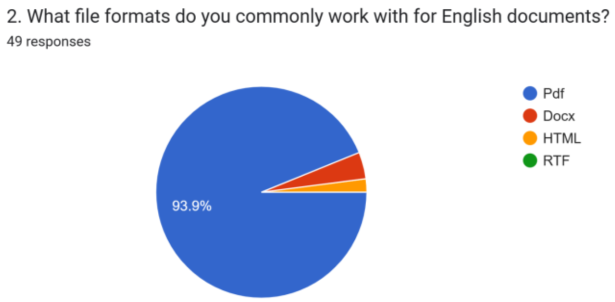
Observations	Method 05
The system can be enriched further by observing user and domain behavior, and also having the accuracy of its operation be at a standard where you could actually say it works at a higher level.	

4.4. Discussion/ Analysis of Results

- The questionnaire was distributed on November 25th, 2023 and 85 responses were collected over a period of two weeks. The following information presents a detailed analysis of the results of the questionnaire.

Questionnaire Result	Analysis
 	This survey reveals a clear divide in experience with document translation systems. Just over half (57.6%) of respondents have not yet used such a system, while a sizable minority (42.4%) report having used one. Most people have no past experience in using a document translation application, which means for most this application will be a fresh experience to witness. Half out of the people who have used a similar kind of an application have experienced a not good performance from the application as depicted from the results of the poll. This encourages us to implement an application which

	will give a better user experience for the public.
  	Despite 55.6% of users finding current English to Tamil PDF translations only "50-75% accurate" in a recent survey, revealing a clear gap in expectations, our project stands poised to surpass these standards. By tackling the complexities of nuanced language, cultural context, and idiomatic expressions, users wanted 75%-90% of accuracy in Tamil when translating and aim to deliver a level of accuracy that exceeds existing offerings, pushing the boundaries of cross-linguistic communication and exceeding user expectations. So, while the current landscape paints a picture of underperformance, our project promises to revolutionize the field, offering precise and culturally-aware translations that bridge the gap between English and Tamil with unprecedented accuracy.
	The survey reveals a near-unanimous preference for PDFs (93.9%) among

 2. What file formats do you commonly work with for English documents? 49 responses 	users working with English documents for translation. While our project's full potential extends beyond format limitations, this insight points towards a strong initial focus on PDF translation. By prioritizing the industry-standard format, we ensure immediate relevance and widespread adoption, exceeding early expectations by tackling the most commonly encountered document type in user workflows. This strategic direction positions us to become the go-to solution for PDF translation, streamlining cross-linguistic communication for a vast user base.

4.5. Use Case Diagram

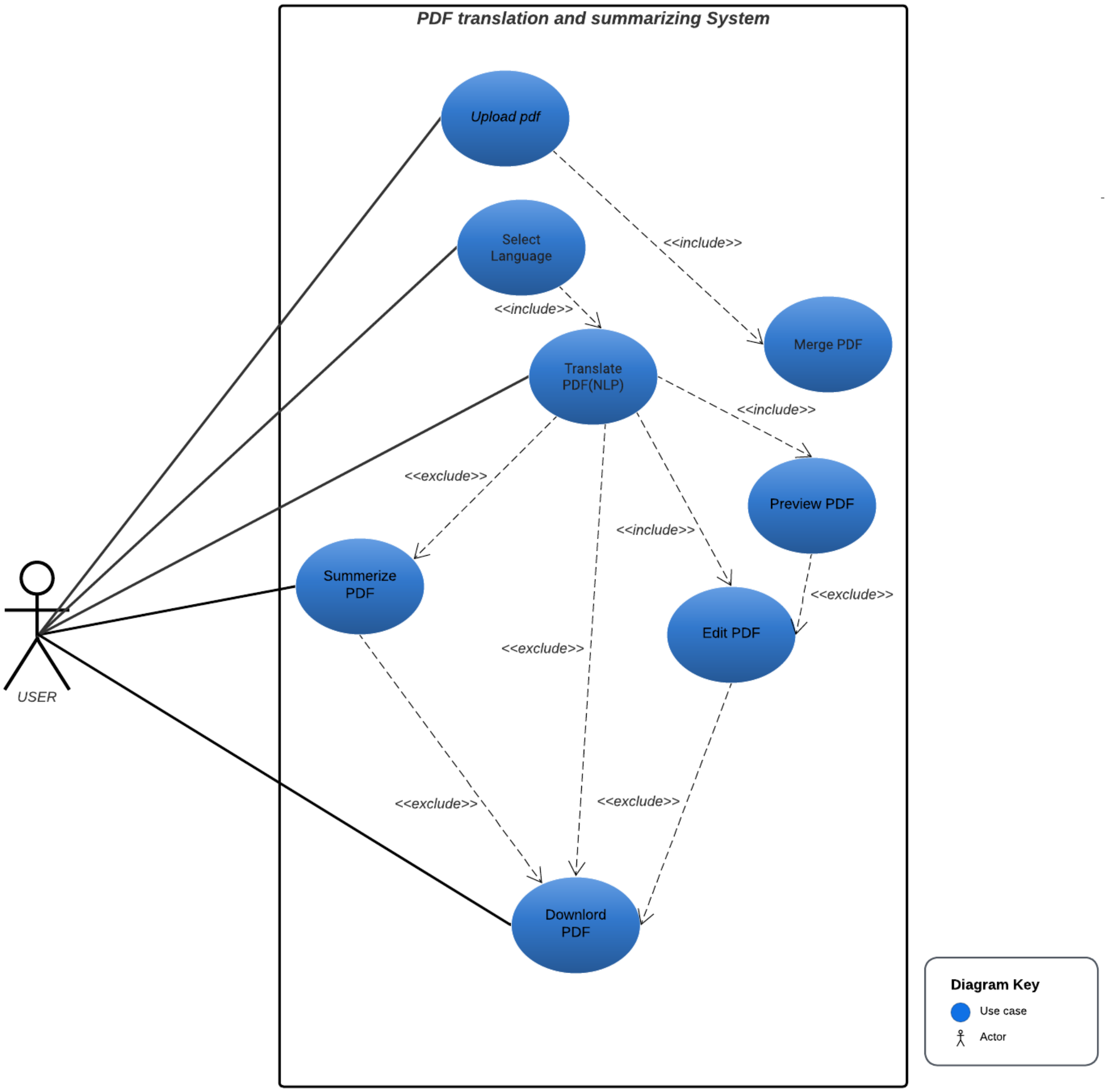


Figure 0.2 Use case diagram

4.6. Use Case Description

Use Case Name	Upload PDF	
Use Case ID	**UC-001**	
Description	Allows the user to upload a PDF file to the system for translation and summarization	
Priority	High	
Primary Actor	User/ Translator	
Supporting Actors	**-**	
Pre-Conditions	System must have access to a translation service	
Trigger	User clicks the 'Upload PDF' button	
Main flow	**Actors**	**System**
	User selects a PDF file to upload	System validates the file format and size.
Exception flow	**Actors**	**System**
	Invalid file format	Display an error message in system
Alternate flow	**Actors**	**System**
	User cancels the upload	–
Exclusions	Editing the uploaded PDF	
Post Conditions	Translation and summarization process initiated	

Use Case Name	Select Language	
Use Case ID	**UC-002**	
Description	Allows the user to choose the language for translation	
Priority	High	
Primary Actor	User/ Translator	
Supporting Actors	-	
Pre-Conditions	Users must be interacting with a feature that requires language selection.	
Trigger	User clicks a "Select language" button	
Main flow	**Actors**	**System**
	User selects the desired language.	System stores the selected language preference.
Exception flow	**Actors**	**System**
	User attempts to select an unavailable language	Display an error message and allow the user to retry.
Alternate flow	**Actors**	**System**
	–	
Exclusions	Translating content	
Post Conditions	System interface or content is displayed in the chosen language.	

Use Case Name	Translate PDF	
Use Case ID	**UC-003**	
Description	Allows the user to translate a previously uploaded PDF file into selected language.	
Priority	High	
Primary Actor	User/ Translator	
Supporting Actors	-	
Pre-Conditions	User must have selected a target language for translation.	
Trigger	User clicks the "Translate PDF" button	
Main flow	**Actors**	**System**
		System sends the PDF file and target language to the translation model
Exception flow	**Actors**	**System**
		System displays an error message and allows the user to troubleshoot or try again
Alternate flow	**Actors**	**System**
	–	
Exclusions	Editing the translated PDF	

Post Conditions	Translated PDF file is generated and available for viewing or download.

Use Case Name	Summarize PDF	
Use Case ID	**UC-004**	
Description	Allows the user to generate a concise summary of a previously uploaded PDF file.	
Priority	High	
Primary Actor	User/ Translator	
Supporting Actors	**-**	
Pre-Conditions	User must have uploaded a PDF file to the system.	
Trigger	User clicks the "Summarize PDF" button.	
Main flow	**Actors**	**System**
	.	Summarization service analyzes the PDF content and extracts keyinformation
Exception flow	**Actors**	**System**
	User attempts Summarization process fails.	Display an error message and allow the user to retry.

Alternate flow	**Actors**	**System**
	–	
Exclusions	Uploading the PDF file	
Post Conditions	Summarized text is generated and displayed to the user..	

Use Case Name	Preview PDF	
Use Case ID	**UC-005**	
Description	Allows the user to view a preview of a PDF file before download.	
Priority	Medium	
Primary Actor	User/ Translator	
Supporting Actors	**-**	
Pre-Conditions	User must have uploaded a PDF file to the system.	
Trigger	-	
Main flow	**Actors**	**System**
		System retrieves the PDF file from storage or generates a preview from the original file.

Exception flow	**Actors**	**System**
		PDF file is corrupt or invalid System displays an error message and does not generate a preview
Alternate flow	**Actors**	**System**
	–	
Exclusions	Editing the translated PDF	
Post Conditions	User has a clear visual representation of the translated PDF content.	

Use Case Name	Edit PDF
Use Case ID	**UC-006**
Description	Allows the user to modify the PDF file before download.
Priority	Medium
Primary Actor	User/ Translator
Supporting Actors	**-**
Pre-Conditions	User must have uploaded a PDF file to the system.

Trigger	User clicks the "Edit PDF" button or link associated with a PDF file	
Main flow	**Actors**	**System**
	User makes desired changes to the PDF content	Systemr saves the edited PDF file
Exception flow	**Actors**	**System**
		Editing tool encounters errors or crashes
Alternate flow	**Actors**	**System**
	–	
Exclusions	Previewing the PDF	
Post Conditions	Edited PDF file is saved and available for further actions.	

Use Case Name	Merge PDF
Use Case ID	**UC-007**
Description	Allows the user to combine multiple PDF files into a single file for streamlined translation
Priority	Medium

Primary Actor	User	
Supporting Actors	-	
Pre-Conditions	User must have uploaded multiple PDF files to the system	
Trigger	User selects the "Merge PDFs" option before the translation	
Main flow	**Actors**	**System**
	User selects the PDF files to be merged	System merges the PDF files into a single, combined PDF
Exception flow	**Actors**	**System**
		Incompatible PDF files
Alternate flow	**Actors**	**System**
	User adjusts merge order –	
Exclusions	Uploading PDF files	
Post Conditions	Merged PDF file is created and ready for translation.	

Use Case Name	Download PDF	
Use Case ID	**UC-008**	
Description	Allows the user to download.	
Priority	High	
Primary Actor	User	
Supporting Actors	-	
Pre-Conditions	System must have the PDF file stored or accessible	
Trigger	User clicks the "Download PDF" button or link associated with the desired file.	
Main flow	**Actors**	**System**
	user select a download location	System transfers the PDF file to the user's device
Exception flow	**Actors**	**System**
		File not found or inaccessible.
Alternate flow	**Actors**	**System**
	–	
Exclusions	Translating or summarizing PDF content	
Post Conditions	PDF file is downloaded and saved on the user's device.	

4.7. Functional Requirements (with prioritization)

- Critical – The requirements that are critically needed in the successful completion
- Desirable – The requirements that can add value, but are not required immediately
- Luxury – The requirements that would add luxury to the system

Requirements list		Priority Level	Description
FR1	Upload PDF	Critical	The user must be able to upload a PDF document to the system for translation and summarization.
FR2	Select Language	Critical	The user must be able to select the target language for translation (Tamil)
FR3	Translate PDF from English to Tamil	Critical	The app must accurately translate the text content of a PDF document from English to Tamil
FR4	Summarize translated language PDF	Critical	The system must generate a concise and accurate summary of the translated Tamil text.
FR5	Preview translated language PDF before download	Desirable	The user should be able to preview the translated PDF and summary before downloading.
FR6	Edit translated language PDF before download	Desirable	The system could allow the user to edit the translated PDF document before downloading.

Table 3 Functional requirements table

4.8. Non-Functional Requirements

.

Requirements list		Priority Level	Description
NF1	Accuracy	Critical	Output of the system should be accurate and valid.
NF2	Performance	Critical	Translation and summarization should be fast and efficient, with minimal processing time.
NF3	Usability	High	The app should be easy to use and navigate for users with varying technical expertise
NF4	Compatibility	Medium	The app should function across different devices, browsers, and operating systems

Table 4 Non-functional requirement table

4.9. Chapter Summary

As this chapter was focused on the system requirements, it looked at the appropriate stakeholders, the techniques used for elicitation purposes and execution of them , a use case diagram, use case description and the functional and non-functional requirements along with their priority levels and descriptions.

Chapter 5: Social, Legal, Ethical and Professional Issues

5.1. Chapter Overview

In the previous chapter the System Requirement Specifications were explained and depicted in detail. This chapter will elaborate how the main Social, Legal, Ethical and Professional Issues are addressed and mitigated by the group in order to complete this project for the Software Development Group Project.

5.2 Dataset Ethical Clearance.

The English to Tamil PDF translation and summary project obtained datasets from various platforms to ensure diverse and comprehensive datasets,

- **The dataset sourced from the GitHub repository**

The dataset used in this work is obtained from the GitHub repository by the authors of the paper "Neural Machine Translation from Tamil to English" (Jain, Punia, Hooda) and published in the Journal of Statistics and Management Systems in 2020. This repository, last accessed June 12, 2021 The updated, provides an important and high-quality corpus of 236,427 English-Tamil synonyms. Notably, the authors continued to improve the dataset by adding more sentences to ensure it was relevant and rich for training analysis.

Effectively formatted information is scaled into six files, making it simple and easy to share. In their experiment, the authors designed and tested two different architectures based on Encoder-Decoder for translating Tamil into English. Notably, the authors aimed to solve challenges such as the problem of polysemy after words, which made a valuable contribution to the field. In addition, the authors performed several experiments, including pre-trained hidden words and tuning hyperparameters to improve the translation quality

The research paper presents and discusses the training example, which shows a remarkable improvement by improving Google Translator by a significant difference of 7.5 BLEU scores. The qualitative research involves people research by three Tamil scholars, besides providing nuanced insights into the translation and accuracy of the translations, the paper provides translations by Google Translator and suggested Tamil translators provides a comparative analysis. Below are the citation of this dataset,

```
@article{jain2020neural,
  title={Neural machine translation for Tamil to English},
  author={Jain, Minni and Punia, Ravneet and Hooda, Ishika},
  journal={Journal of Statistics and Management Systems},
  volume={23},
  number={7},
  pages={1251--1264},
  year={2020},
  publisher={Taylor \& Francis}
```

- **The dataset from GitHub repository**

The datasets used in this exercise are obtained from the Papers with Code repository, in particular the XL-Sum dataset. This dataset is an important contribution to the abstract summarization field because it includes a multilingual collection of 1.35 million article-summary pairs The authors actively updated the dataset to include new definitions, using new languages ho, such as Traditional Chinese, and various aspects were enhanced

The repository provides two versions of the dataset: the old version reported in the associated document and the new version recommended for use. The final version boasts better structure, larger analytical splits, duplication, and an increased size of 1.35 million pairs, making XL-Sum the most extensive summary dataset available to the public there is It is ensured

Importantly, the dataset is only available for non-commercial research purposes, under the Creative Commons Attribution-Noncommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), users are prompted to they do not comply with licensing laws, and copyright in dataset content belongs to the original copyright holders.

To demonstrate eligibility, dataset authors ask users to identify a related document that uses any part of the dataset, model, or code module. Quotes from the paper are provided as follows:

```
@inproceedings{hasan-etal-2021-xl,
    title = "{XL}-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages",
    author = "Hasan, Tahmid  and
      Bhattacharjee, Abhik  and
      Islam, Md. Saiful  and
      Mubasshir, Kazi  and
      Li, Yuan-Fang  and
      Kang, Yong-Bin  and
      Rahman, M. Sohel  and
      Shahriyar, Rifat",
    booktitle = "Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021",
    month = aug,
    year = "2021",
    address = "Online",
    publisher = "Association for Computational Linguistics",
    url = "https://aclanthology.org/2021.findings-acl.413",
    pages = "4693--4703",
}
```

Also Contents of this repository are restricted to only non-commercial research purposes under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0).

- **The dataset obtained from Kaggle**

In the dataset from Kaggle (https: 2000). //www.kaggle.com/code/bagavathypriya/english- to-tamil-translation). Clear evidence of ethical clearance was not explicitly provided, and no specific License information is provided by the dataset provider. The dataset contains several English sentences and their Tamil translations, providing a valuable resource for machine translation using sequence-sequence model with encoder and decoder, due diligence and proper attribution will be done to this data set of any further use, to determine the possible contribution of the data set to language translation research.

5.3 SLEP Issues and Mitigation

During the second-year project, it is important to ensure that all social, legal, ethical and professional issues are addressed. The social, legal, ethical and professional issues related to "Thruvox" is detailed below,

5.3.1 Social Issue

There are many important social factors that need to be deeply considered when developing an app for summarizing and translating documents from English to Tamil. It is important that the translation represents culture correctly and doesn't miss out on crucial details. For this reason, the importance of being culturally sensitive is highlighted here. In order to make the app appeal to all ages within the Tamil community, it needs to be built with an underlying structure that includes a diverse set of dialects and varying regional tone. To maintain gender equality while translating documents, it is important that there are no gender-specific terms present to avoid reinforcing stereotypes. Issues are likely to be spread out since users might put classified information. Therefore, precautions need to be taken. For an app to be deemed user-friendly there has to be made drastic considerations just how accessible that the application is for one with varying digital needs within their levels of abilities and skill. To maintain precision and avoid misinformation there should be quality control measures so that complex content can also be assessed by humans. For a user to better understand the interface and how it works guidelines provided by developers. In order to make a summarized application and translation from English to Tamil, by solving these social issues we can include everyone.

5.3.2 Legal Issue

Our English-to-Tamil translation and summarization application focuses on user privacy alongside responsible tech practices. No personal information gets stored, processing takes place locally and temporary text is quickly removed. We use open-source libraries, and we respect intellectual property by crediting the resources utilized and building our core functionality ourselves. We used a translation model we either developed ourselves or ethically borrowed, and the resources of Tamil language we employed are open-source donations or legally purchased. Key aspects include bias awareness and transparency; we work for equitable results with different text kinds, inform users of possible limitations, and reach out to receive feedback in order to keep refining our service always. By focusing on these ethical and legal issues, we can provide a reliable tool that will enable cross-language communication as well as access to information. A Kaggle dataset will be used in the Thruvox project. Data protection laws will be carefully considered, as will the terms and conditions of use. Since the dataset was published under the Creative Commons public license, we intend to download it. Under no circumstances will the data be mishandled. The software will also be given top priority. We will be planning to use the software legally provided by the university, such as Microsoft Office, Adobe XD, Figma and so on. Open Source tools will be used such as Rasa which is a community version. The questionnaire was submitted, and it did not gather any personal data of the users. The identities of those who completed the survey were kept anonymous. The privacy of the user is protected throughout the project.

5.3.3 Ethical

The experts responsible for conducting interviews to our translate and summarize application were made about the purpose of these surveys, and how such information was going to be used. It is essential to mention that all information gathered throughout the process will be used only within the boundaries of our project and won’t reveal them to representatives of any other organizations. We also ensure that all the software used in this project is acquired through legitimate means, and we never use any cracked or pirated versions.

When we were procuring the data set for our project, ethics took precedence. The dataset in Kaggle under the Creative Common Public License is made available to the public for various purposes as per permitted by the original author. We also ensure that we handle the dataset with integrity and do not try to make any modification they can compromise its authenticity.

The privacy of individuals was something that should not ever be bypassed during the creation process for our Thruvox application. The feedback process did not collect any personal data, such as names or email addresses. User feedback and survey data will be collected only for the purpose of application development, while being utilized in a way that does not result in conflicts between participants or other individuals.

We keep high ethical standards through the development of our application. Our team ensures that the ideas we use are unique and never plagiarized. Ethical boundaries are observed, ensuring that the development process aligns with the highest ethical standards.

5.4 Chapter Summary

This chapter focuses on the challenges presented initially and what one needs to consider when deploying technology meant to summarize documents in PDF format. The aspect of professionalism focuses on abiding by ethical standards, and upholding a responsible communication method both the capabilities and limitations. This chapter provides details about the adaptation of PDF files. It gives readers an understanding of how to change and summarize data in a documented file.

Chapter 6: System Architecture & Design

6.1. Chapter Overview

This chapter will mainly be focused on the conceptual model Diagrams and images used in this chapter to further explain the project. Utilizing these can contribute to a somewhat improved understanding of the project's final result. This chapter includes a variety of components, including architecture diagrams, class diagrams, sequence diagrams, UI design, and activity diagrams.

6.2. System Architecture Design

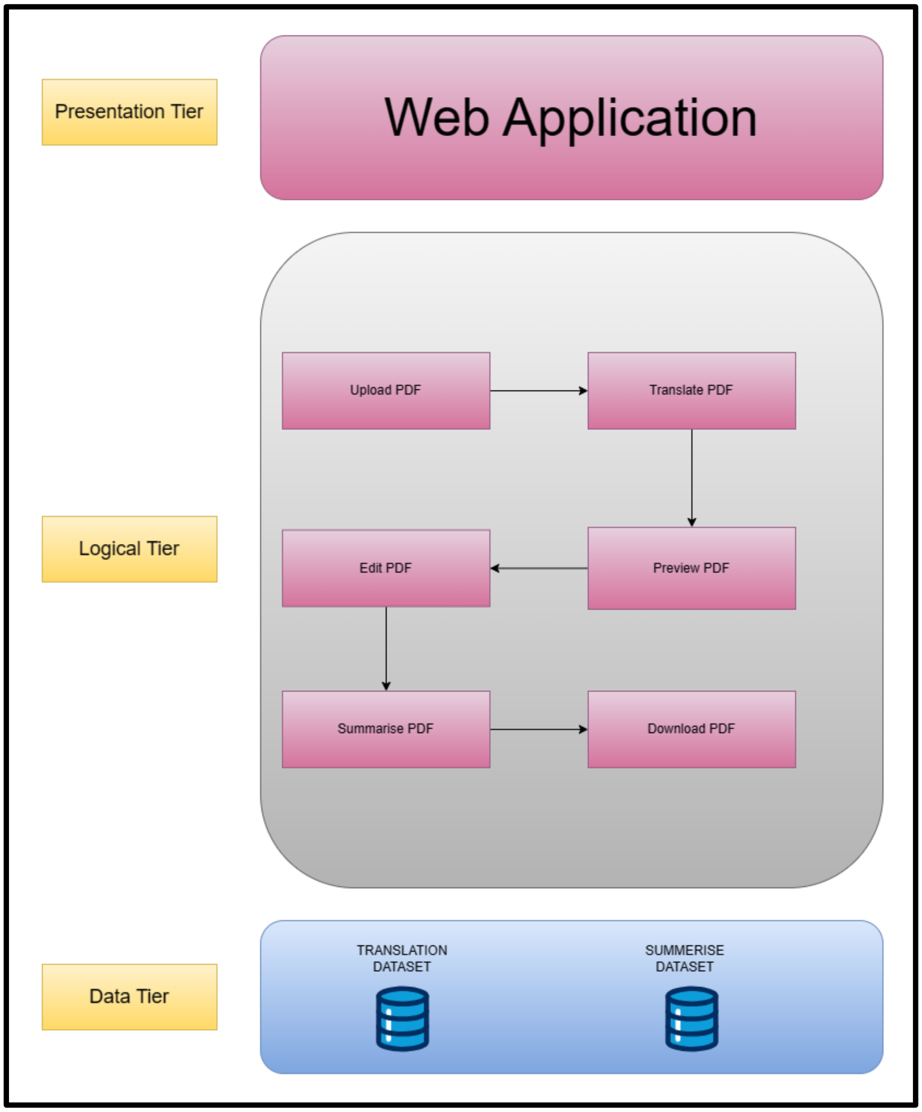


Figure 0.1 System architecture diagram

This architecture appears to be a traditional three-tier web application, likely for translating and summarizing PDFs. The user interface sits on top, potentially a web page or mobile app, which interacts with the middle tier, a business logic layer responsible for handling user requests and orchestrating the translation and summarization tasks. Finally, the bottom tier consists of data storage and external services, including translation and summarization engines.

6.3. System Design

6.3.1. Class Diagram

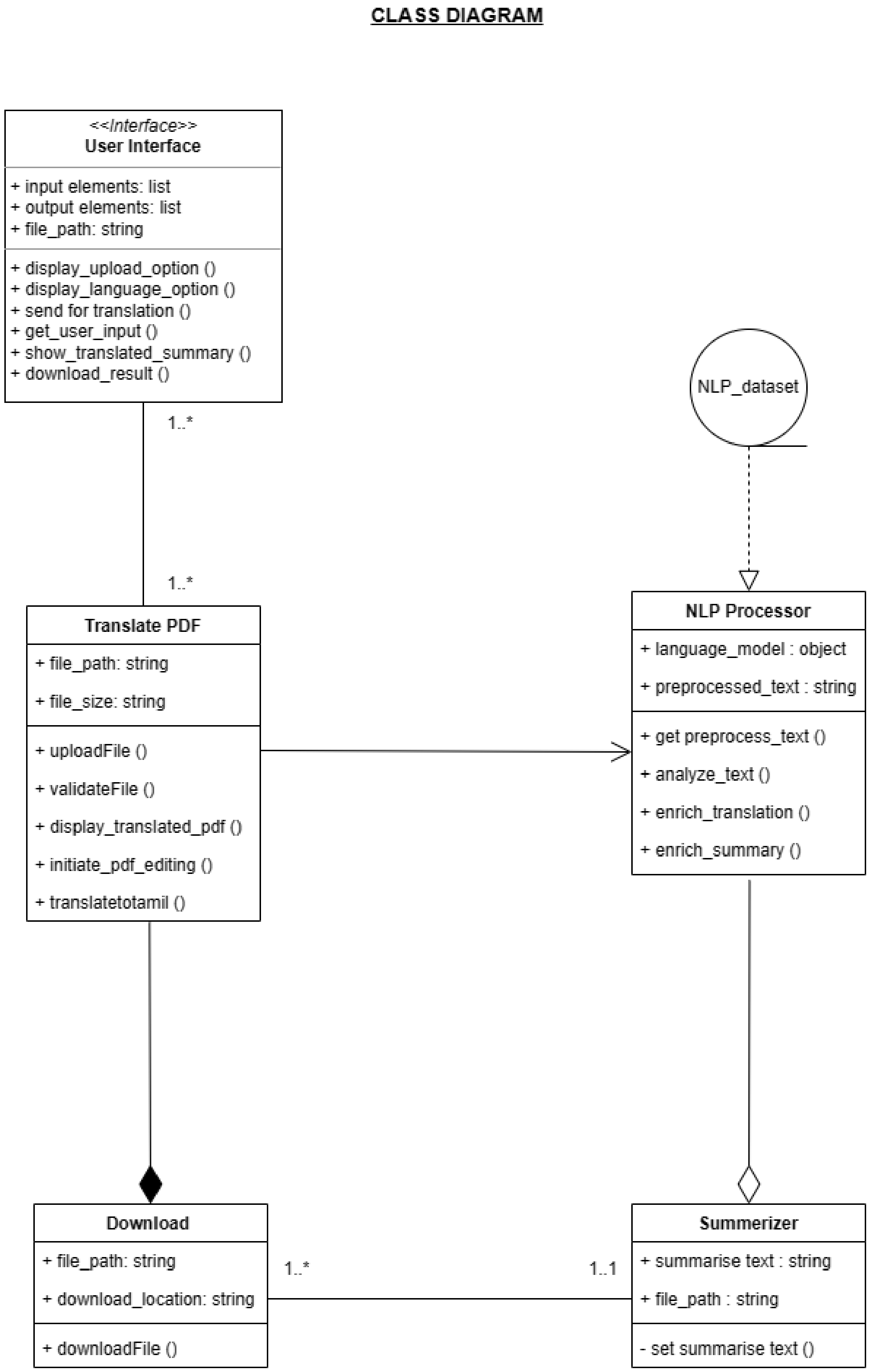


Figure 0.2 Class diagram

A class diagram is a visual representation of the classes, attributes, and relationships in a system. It is an important tool in software development as it helps to understand the structure of a system before it is built. In this class diagram, the class diagram portrays a sophisticated software system designed for PDF translation and summarization, encompassing four core classes: UI Download, T PDF Translate, NLP Processor etc. The User Interface coordinates user actions and controls the presentation of upload alternatives, linguistic options, as well as download sequence. With the help of Translate PDF, you can perform critical tasks like validating a PDFyou are working on , present content to readers in an attractive manner and initiate editing processes as wellas translations from text-to-HTML . NLP Processor leverages powerful language models to enhance translation and summarization. The Download class makes it easy to get the final output. Is it elegant, then when classes harmonize in the very concept of PDF translation and summarization that is user-oriented as well as effective.

6.3.2. Sequence Diagram

Upload PDF

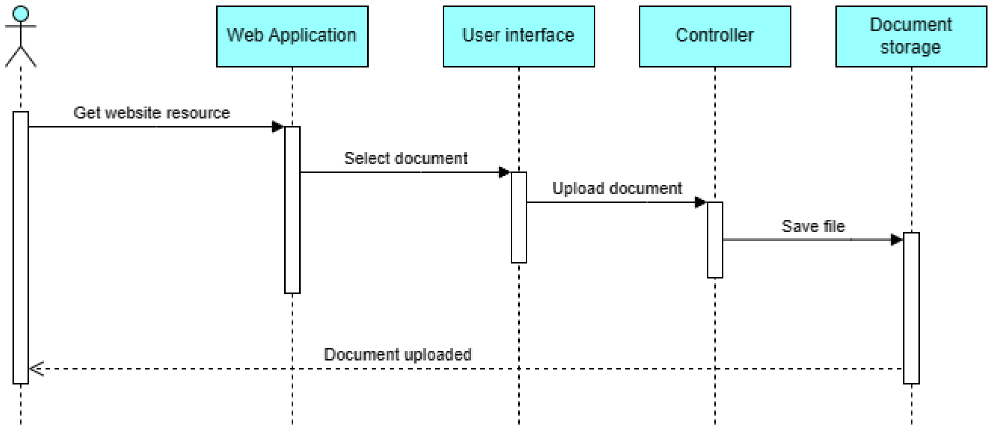


Figure 0.3 Upload sequence

Select Translation Language

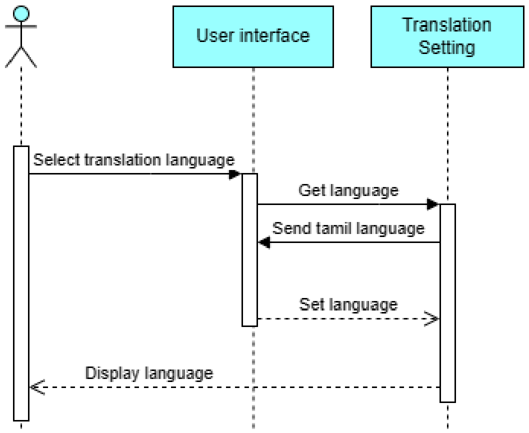


Figure 0.4 select translation language sequence

Merge PDF

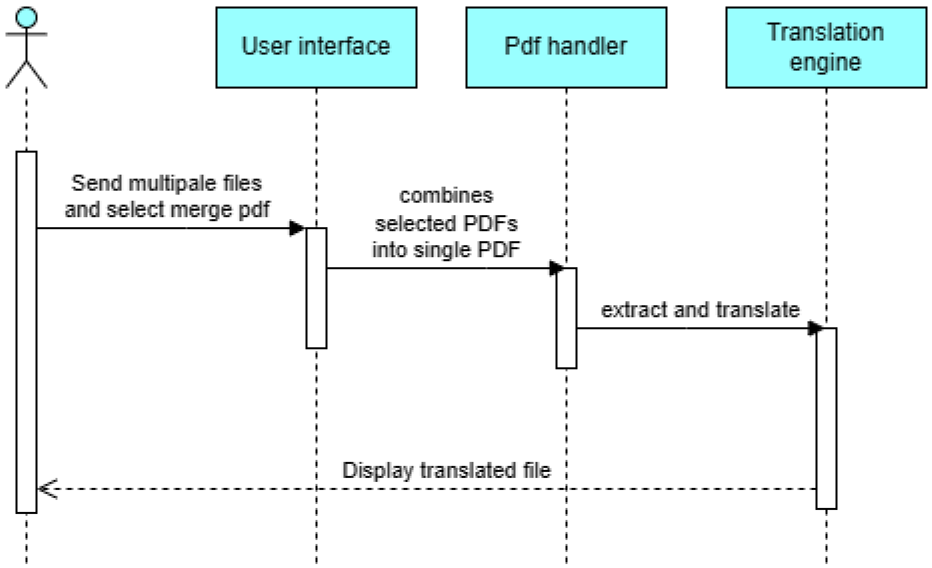


Figure 0.5 Merge sequence

Translate PDF

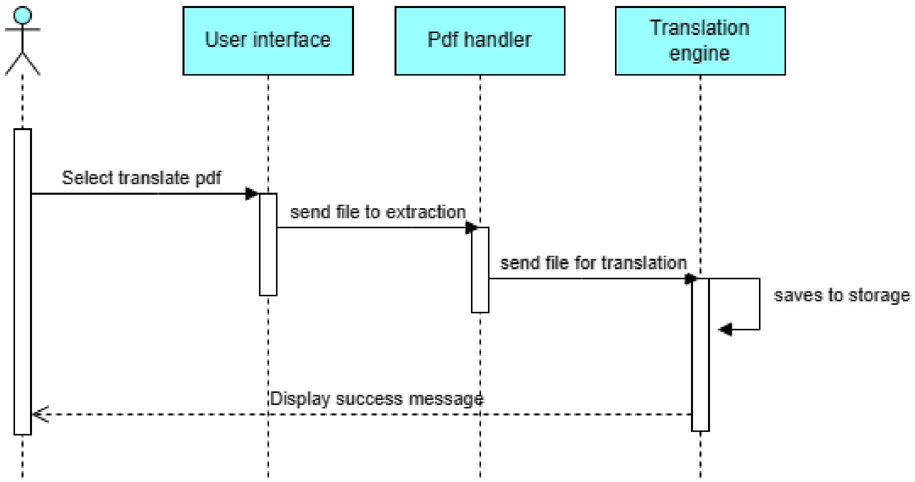


Figure 0.6 Translate sequence

Preview PDF

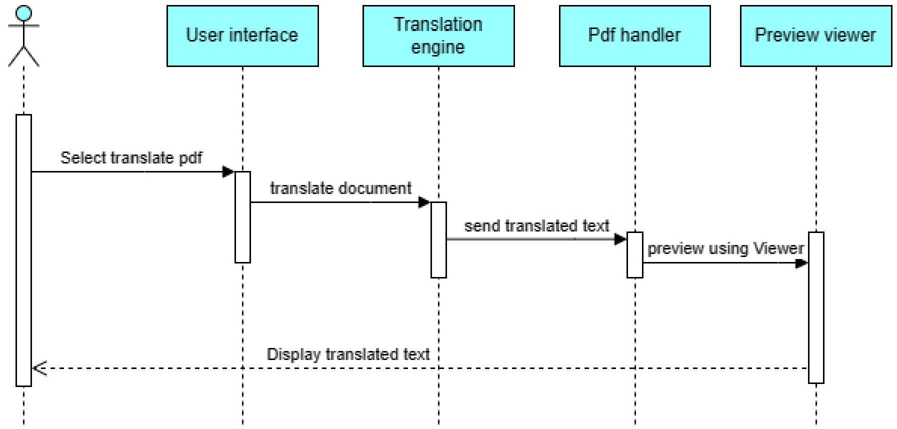


Figure 0.7 Preview sequence

Edit PDF

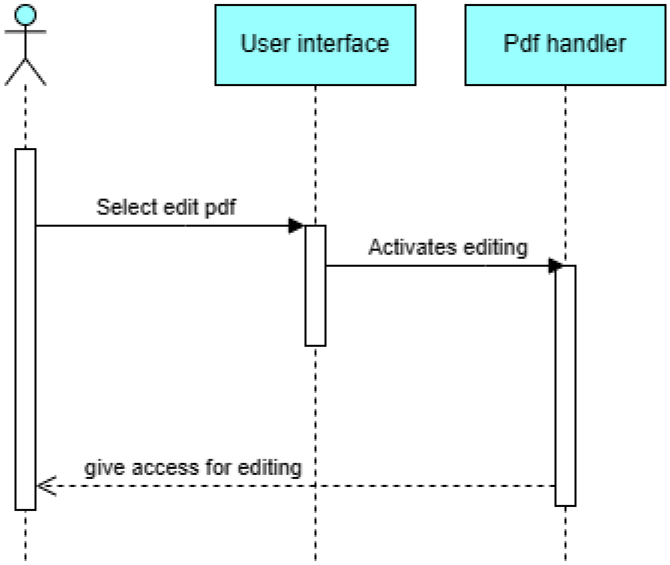


Figure 0.8 Edit sequence

Summerization

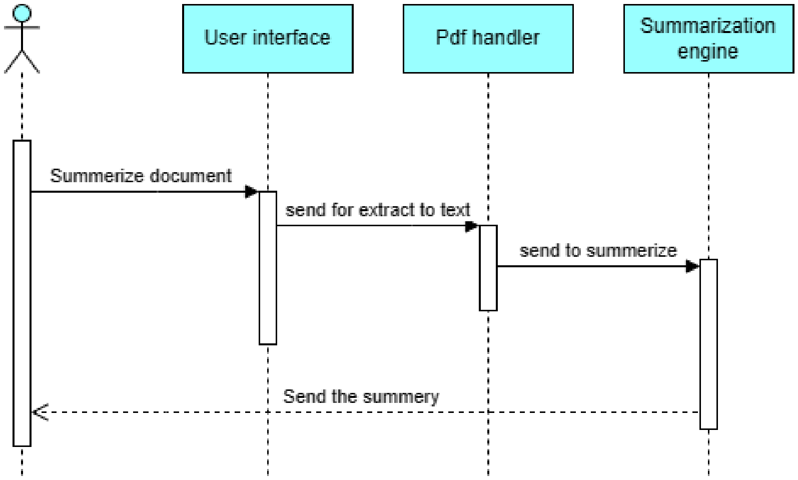


Figure 0.9 Summarize sequence

Download PDF

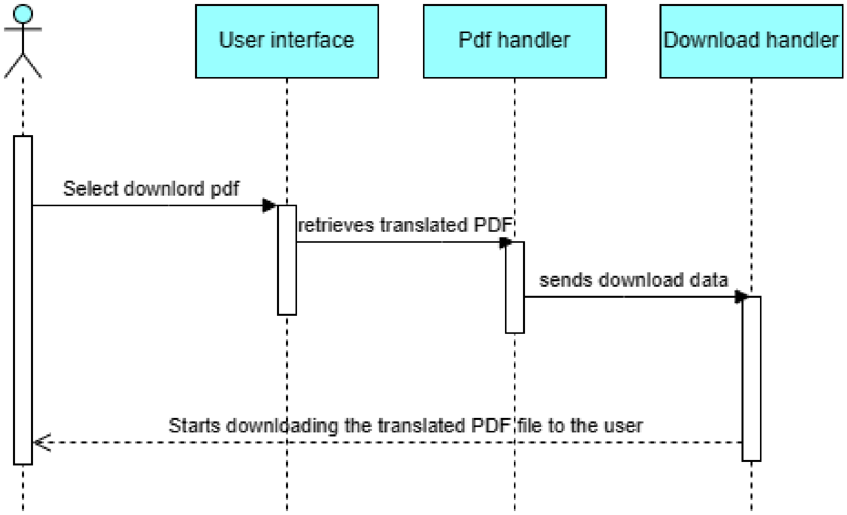


Figure 0.10 Download sequence

6.3.3. UI Design and mockups – des/high fidelity prototype

UI Design and mockups des/high fidelity prototype has been moved to Appendix Section for better clarity.

6.3.4. Activity Diagram

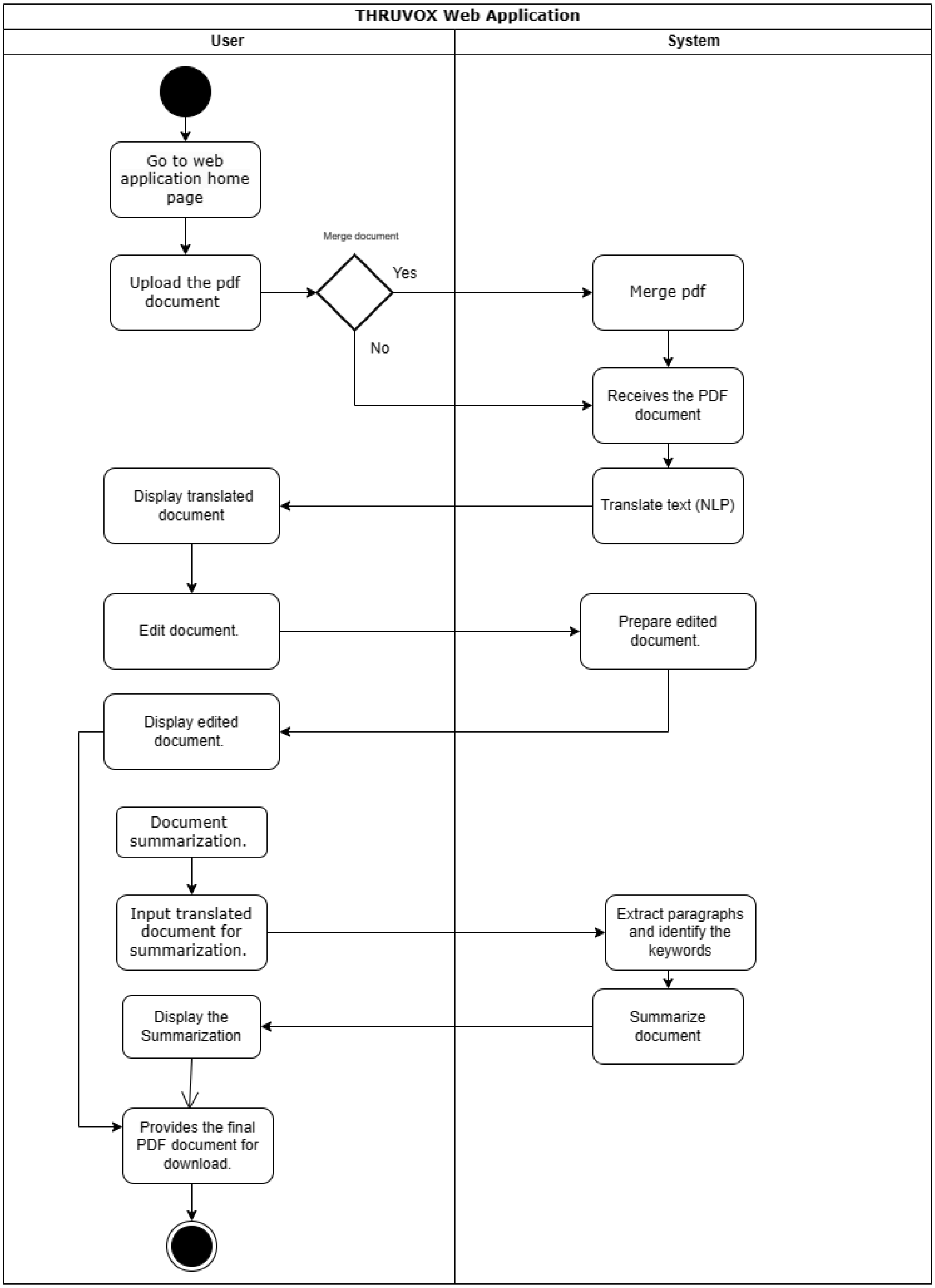

THRUVOX Web Application
User
System
Go to web application home page
Upload the pdf document
Merge document
Yes
No
Merge pdf
Receives the PDF document
Translate text (NLP)
Display translated document
Edit document.
Prepare edited document.
Display edited document.
Document summarization.
Input translated document for summarization.
Extract paragraphs and identify the keywords
Summarize document
Display the Summarization
Provides the final PDF document for download.

Figure 0.11 Activity Diagram

6.4. Chapter Summary

This chapter discussed the high level and low-level designs of the THRUVOX application including the system architecture design and system design. System design is outlined with the class diagram, sequence diagram including UI Design and mockups . Class diagram of the system design was included with all the entities, attributes and relationships of the system. Sequence diagram includes the order of the application which the user will be handling. The process flow diagram is depicted and explained in detail. This chapter concludes the main content of the System Requirements Specification.

Reference

Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. Proceedings of the second international conference on Human Language Technology Research, 138-145.

Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. Text Summarization Branches Out, 74-81.

Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 3075-3081.

W3C. (2019). Web Content Accessibility Guidelines (WCAG) 2.1. Retrieved from https://www.w3.org/TR/WCAG21/

Fowler, M. (2002). Patterns of enterprise application architecture. Addison-Wesley Professional. (Chapter 3: The Three-Tier Architecture)

Rao, P., & Sankar, A. (2014). Cultural diversity and translation: A critical analysis. Language and Intercultural Communication, 14(2), 159-175.

Subramonian, V. (2013). The Dravidian languages and the Indian sociolinguistic scene. The Routledge handbook of sociolinguistics in India, 43-63.

Disability and Development: A Human Rights Approach (https://www.un.org/development/desa/dspd/2019/04/un-disability-and-development-report-realizing-the-sdgs-by-for-and-with-persons-with-disabilities/)

Brundage, M., et al. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint.

Larman, C. (2004). Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development (3rd ed.). Pearson Education.

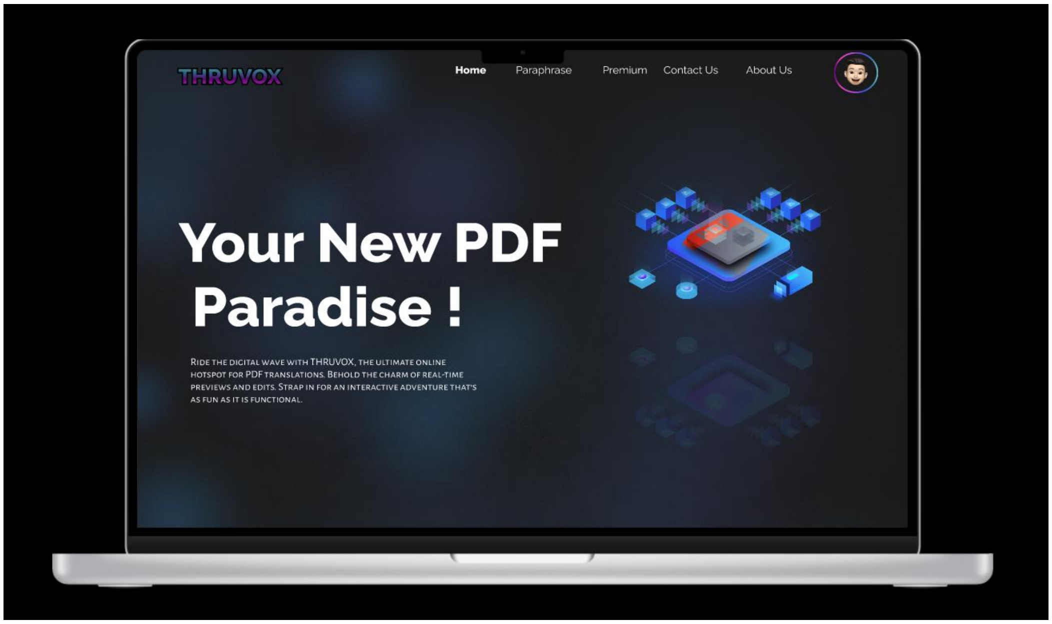
Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. Proceedings of the First Workshop on Neural Machine Translation, 28-39.

Rao, P., & Sankar, A. (2014). Cultural diversity and translation: A critical analysis. Language and Intercultural Communication, 14(2), 159-175.

van der Voort, H., & Dankbaar, B. (2014). Stakeholders in online information services: A literature review. Online Information Review, 38(2), 309-328.

Bhattacharjee, A., & Newman, M. (2007). Stakeholder analysis for information technology projects in developing countries. The Journal of Development Studies, 43(8), 1249-1275.

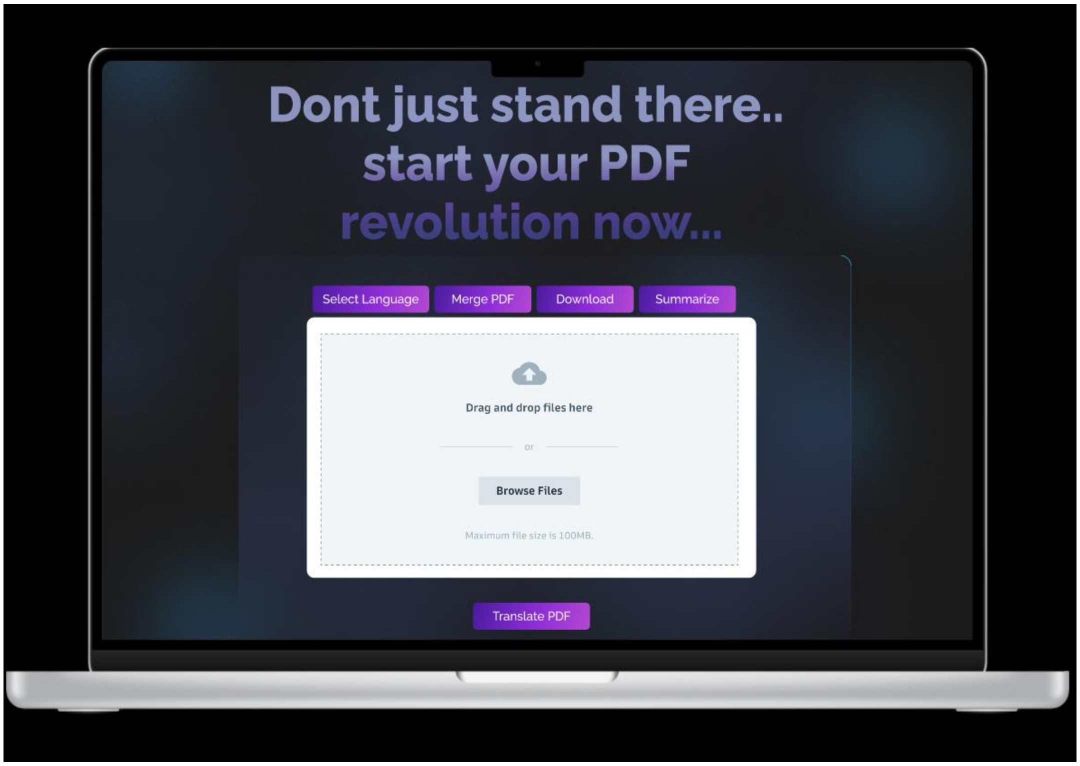
Appendix



Your New PDF Paradise !

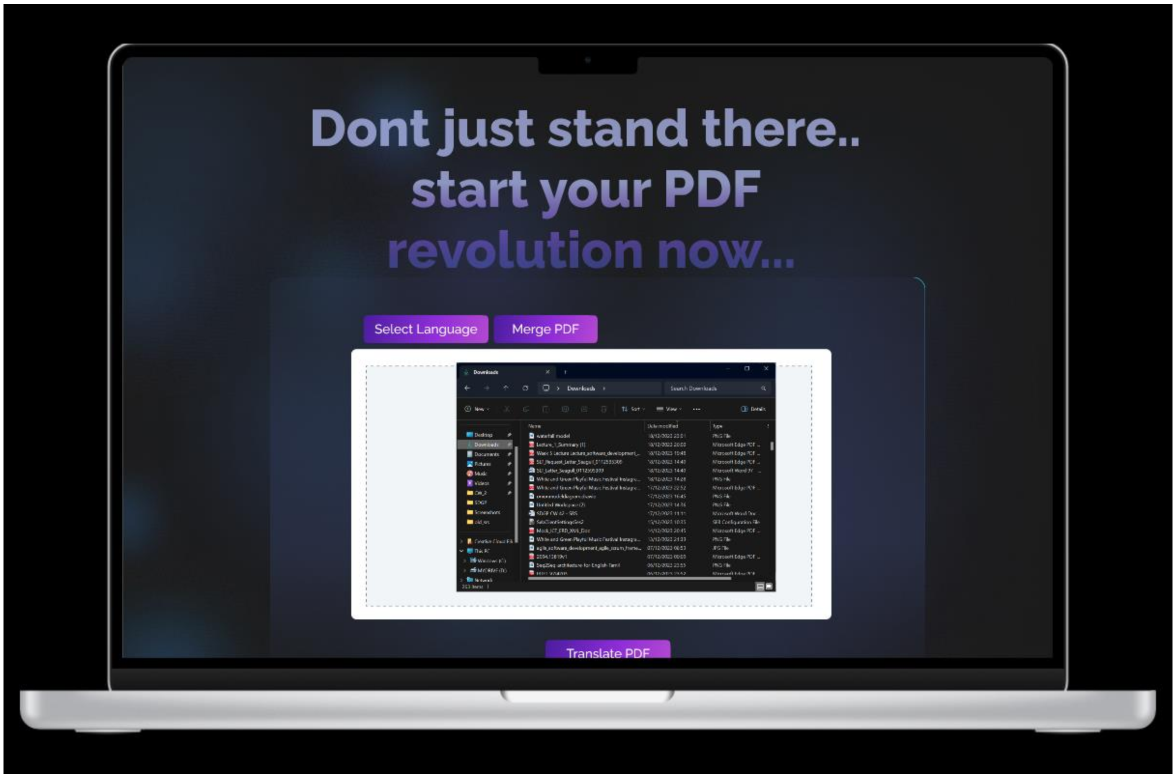
Ride the digital wave with THRUVOX, the ultimate online hotspot for PDF translations. Behold the charm of real-time previews and edits. Strap in for an interactive adventure that's as fun as it is functional.

Figure 6.3.3.1 Home page of THRUVOX



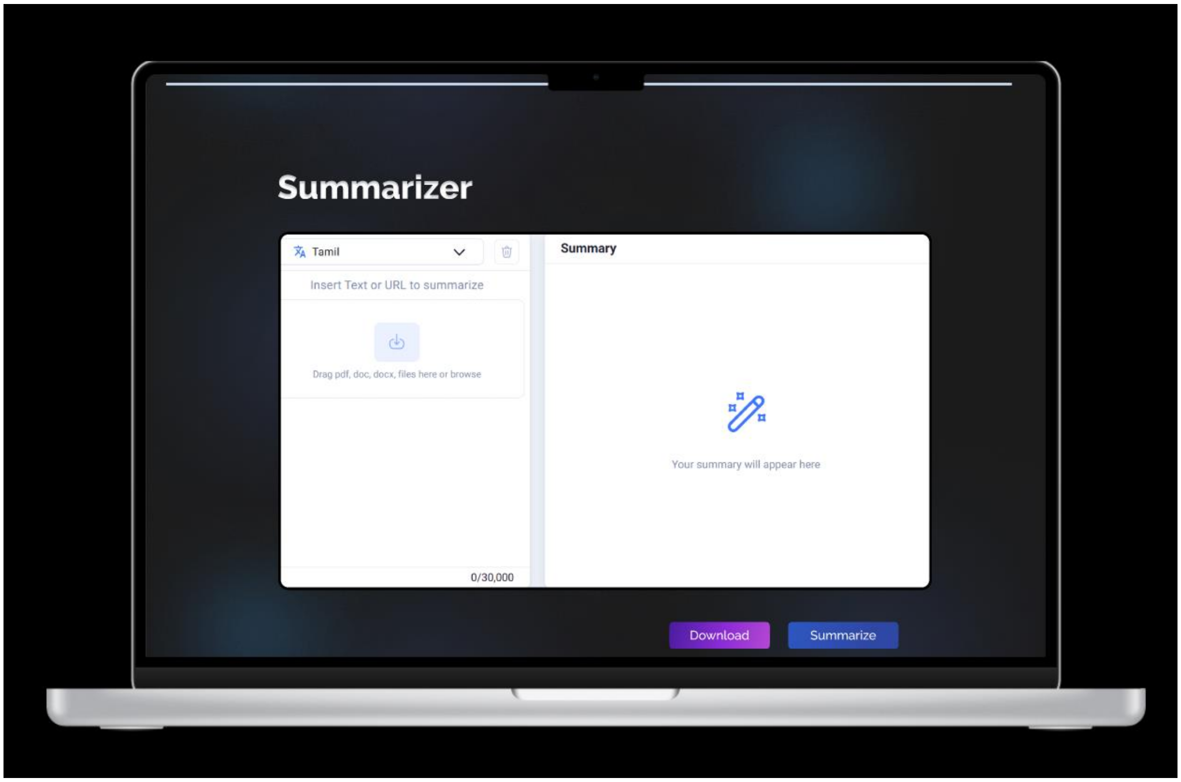
Dont just stand there.. start your PDF revolution now...

Figure 6.3.3.2 PDF upload page of THRUVOX



Dont just stand there..
start your PDF
revolution now...

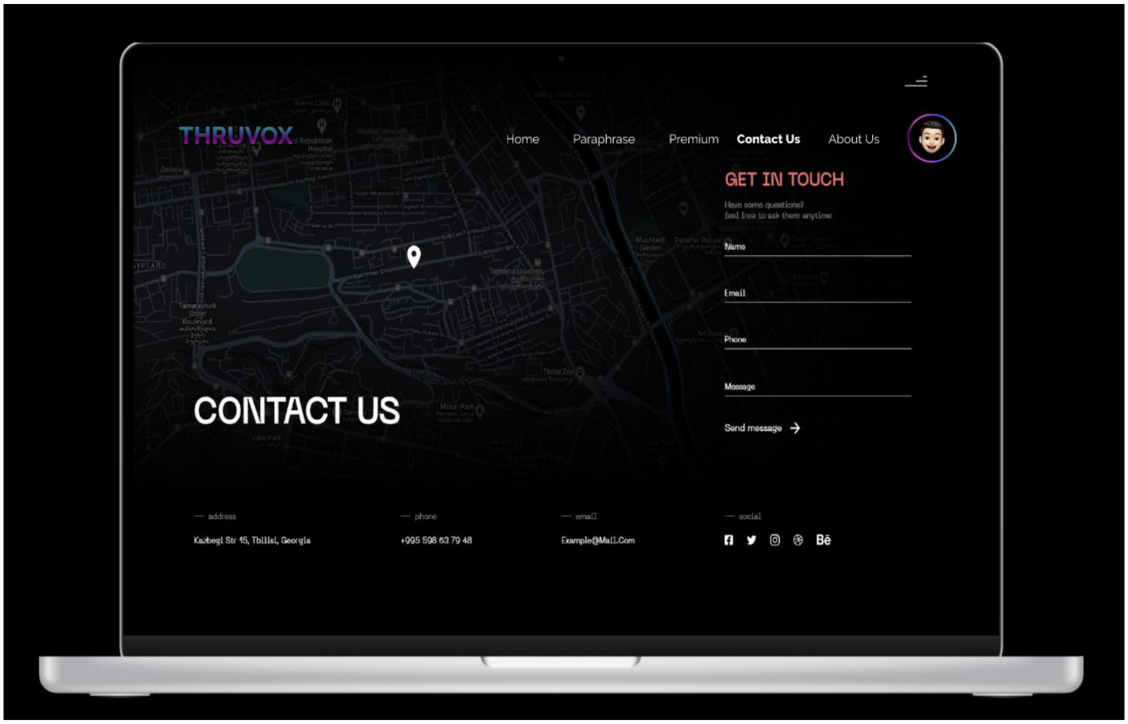
Figure 6.3.3.3   File select page of THRUVOX



Summarizer

Tamil
Insert Text or URL to summarize
Drag pdf, doc, docx, files here or browse
0/30,000
Summary
Your summary will appear here
Download
Summarize

Figure 6.3.3.4   Summarize page of THRUVOX



GET IN TOUCH

Have some questions?
feel free to ask them anytime

Name

Email

Phone

Message

CONTACT US

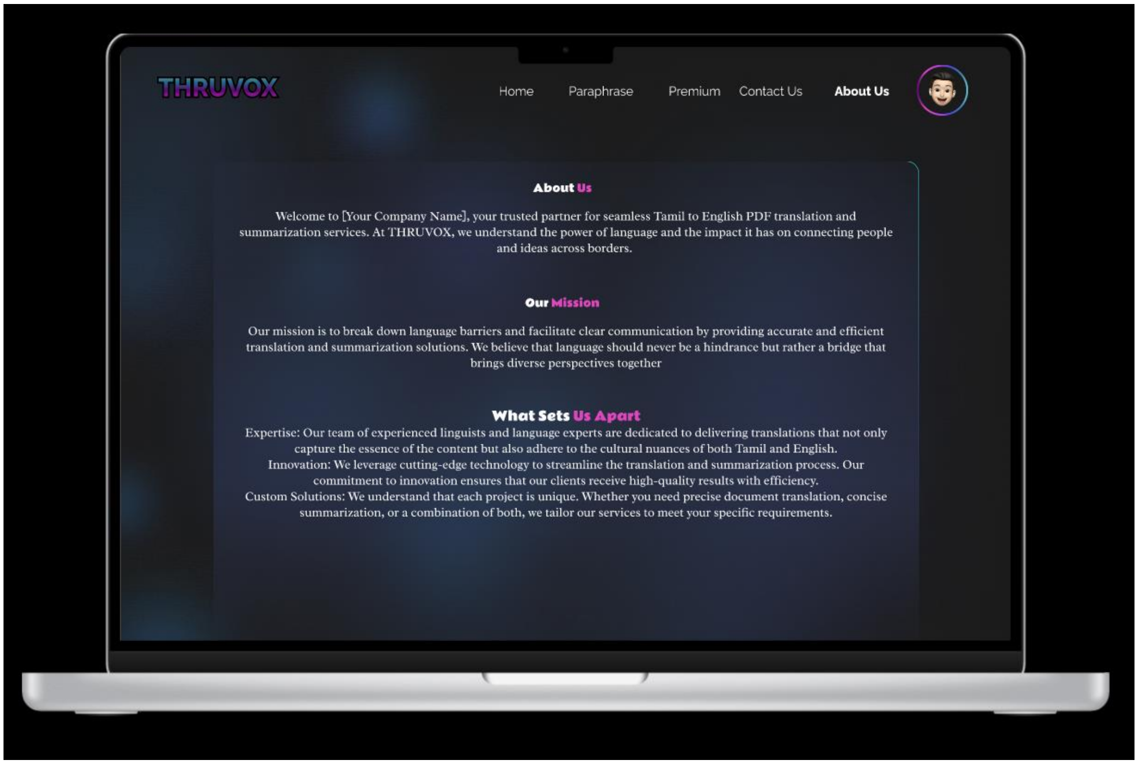
— address
Kazbegi Str 15, Tbilisi, Georgia

— phone
+995 598 63 79 48

— email
Example@Mail.Com

— social

Figure 6.3.3.5   Contact Us page of THRUVOX



About Us

Welcome to [Your Company Name], your trusted partner for seamless Tamil to English PDF translation and summarization services. At THRUVOX, we understand the power of language and the impact it has on connecting people and ideas across borders.

Our Mission

Our mission is to break down language barriers and facilitate clear communication by providing accurate and efficient translation and summarization solutions. We believe that language should never be a hindrance but rather a bridge that brings diverse perspectives together

What Sets Us Apart

Expertise: Our team of experienced linguists and language experts are dedicated to delivering translations that not only capture the essence of the content but also adhere to the cultural nuances of both Tamil and English.
Innovation: We leverage cutting-edge technology to streamline the translation and summarization process. Our commitment to innovation ensures that our clients receive high-quality results with efficiency.
Custom Solutions: We understand that each project is unique. Whether you need precise document translation, concise summarization, or a combination of both, we tailor our services to meet your specific requirements.

Figure 6.3.3.6   About Us page of THRUVOX