

A
Major Project Report
on
JAPANESE LANGUAGE TRANSLATOR

Submitted in Partial Fulfillment of
the Requirements for the Degree
of
Bachelor of Engineering

in
Computer Engineering
to

Kavayitri Bahinabai Chaudhari
North Maharashtra University, Jalgaon

Submitted by
Pranav Subhash Wadodkar
Bhumika Dilip Narkhede
Shivani Anil Mahajan
Purva Vishnu Patil

Under the Guidance of
Mrs. Nilima S. Ramteke



DEPARTMENT OF COMPUTER ENGINEERING
SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY,
BAMBHORI, JALGAON - 425 001 (MS)
2021 - 2022

**SSBT's COLLEGE OF ENGINEERING AND TECHNOLOGY,
BAMBHORI, JALGAON - 425 001 (MS)
DEPARTMENT OF COMPUTER ENGINEERING**

CERTIFICATE

This is to certify that the major project entitled *JAPANESE LANGUAGE TRANSLATOR*, submitted by

**Pranav Subhash Wadodkar
Bhumika Dilip Narkhede
Shivani Anil Mahajan
Purva Vishnu Patil**

in partial fulfillment of the degree of *Bachelor of Engineering in Computer Engineering* has been satisfactorily carried out under my guidance as per the requirement of Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon.

Date: April 21, 2022

Place: Jalgaon

Mrs. Nilima S. Ramteke
Guide

Dr. Manoj E. Patil
Head

Prof. Dr. G.K.Patnaik
Principal

Acknowledgements

We would like to express our deep gratitude and sincere thanks to all who helped us in completing this project report successfully. Many thanks to almighty God who gave us the strength to do this. Our sincere thanks to Principal Prof. Dr. Girish. K. Patnaik for providing the facilities to complete this Project report. We would like to express our gratitude and appreciation to all those who gave us the possibility to complete this report. A special thanks to Prof. Dr. Manoj E. Patil, Head of the Department, whose help, stimulating suggestions and encouragement, helped us in writing this report. We would also like to thank Mrs. Nilima S. Ramteke, Project Guide, who has given her full effort in guiding us and achieving the goal as well as her encouragement to maintain the progress in track. I am also sincerely thankful to Mr. Akash D. Waghmare, Incharge of Project, for his valuable suggestions and guidance. We would also like to appreciate the guidance given by other supervisor that has improved our presentation skills by their comments and tips. Last but not the least, we are extremely thankful to our parents and friends without whom it could not reach its successful completion.

Pranav Subhash Wadodkar
Bhumika Dilip Narkhede
Shivani Anil Mahajan
Purva Vishnu Patil

Contents

Acknowledgements	ii
Abstract	1
1 Introduction	2
1.1 Problem Definition	3
1.2 Scope	3
1.3 Objective	3
1.4 Organization of Report	3
1.5 Summary	4
2 Project Planning and Management	5
2.1 Feasibility Study	5
2.1.1 Economical Feasibility	6
2.1.2 Operational Feasibility	6
2.1.3 Technical Feasibility	7
2.2 Project Scheduling	7
2.3 Effort Allocation	7
2.4 Summary	7
3 Analysis	8
3.1 Requirements Collection and Identification	8
3.2 Hardware and Software Requirement	8
3.2.1 Hardware Requirement	8
3.2.2 Software Requirement	9
3.2.3 Product Features	9
3.3 Functional and Non-Functional Requirement	9
3.3.1 Functional Requirements	9
3.3.2 Non-Functional Requirements	9
3.4 Software Requirements Specification (SRS)	9
3.5 Summary	10

4	Design	11
4.1	System Architecture	11
4.2	Data Flow Diagram	12
4.3	UML Diagram	12
4.3.1	Use Case Diagram	12
4.3.2	Sequence Diagram	13
4.3.3	Class Diagram	13
4.3.4	State Diagram	14
4.4	Summary	14
5	Implementation and Coding	15
5.1	Algorithm/Steps	15
5.1.1	Algorithm Used	16
5.2	Require Software and Hardware For Development	17
5.2.1	Software Requirements	17
5.2.2	Hardware Requirements	17
5.3	Summary	17
6	Testing	18
6.1	Manual Testing	18
6.2	Automated Testing	19
6.3	Test Cases Identification and Execution	21
6.4	Summary	21
7	Result and Discussion	22
7.1	Result	22
7.2	Summary	23
8	Conclusion	24

List of Tables

6.1	Test Case For Translation	21
7.1	Confusion Matrix	22

List of Figures

4.1	System Architecture	11
4.2	Data Flow Diagram	12
4.3	Use Case Diagram	12
4.4	Sequence Diagram	13
4.5	Class Diagram	13
4.6	State Diagram	14
5.1	Algorithm	15
6.1	Manual Un-Translated text	18
6.2	Manual Translated text	19
6.3	Auto Un-Translated text	19
6.4	Auto Translated text	20

Abstract

Language Translation is the process of detecting the language from any kind of Text, Text File, or an Image and then translate it into target language. Implementation of a Natural Language Processing model for Language Translation is to be carried out in this project. The major task is to identify those features or parameters which could be used to clearly distinguish the language and translate it. This model makes use of Machine Learning, Artificial Intelligence, and Natural Language Processing. The project aims at detecting Japanese and translate it into English.

Experiments were conducted by forming text samples obtained from online articles and social media. This corpus comprises of general articles, each of them spanning over at least 100 words. The entire corpus is split into two sets, larger unit as the training data-set and a smaller set as the test set.

Chapter 1

Introduction

The phenomenon of globalization has brought together people from around the world. However one barrier to this increase in global communication is that many people communicate in different languages and effectively lack a common communication medium. That is, in order to communicate effectively a language which is mutually understandable by both parties is required. Language Translation offers a means for providing this medium

Language as a communication system is thought to be fundamentally different from and of much higher complexity than those of other species as it is based on a complex system of rules relating symbols to their meanings, resulting in an indefinite number of possible innovative utterances from a finite number of elements.[1] Among the various factors that define different cultures and communities, an important factor is language. The importance of language for human to human communication can be over emphasized.

Language Translation is an important task for information retrieval services. This paper presents the implementation of a platform for language translation in documents on web. The platform consists of modules to achieve the tasks automatically. Furthermore, artificial neural networks were used for the translation of language in documents. Results for Japanese language translation to English are present. The major benefit of the approach is that the ANN based language translation system could meet the expectations in real-time language translate accuracy with the help of a developed system.

1.1 Problem Definition

Given an utterance that belongs to the vocabulary of a language, the language translation systems should identify the language irrespective of accents or pronunciations using an ANN model, which extracts suitable features from text samples across different languages. The language set considered in our project includes English and Japanese. The system should evolve over a time with better accuracy and use continuous learning mechanism by incorporating machine learning techniques.

1.2 Scope

The sole purpose of this project is to correctly identify and translate the utterance of Japanese language available in the training set. The scope of Language Translation systems lies in making the Detection as the preprocessing stage for Language translation. This could be a contribution to the field of translation and many applications can make use of it as a more dominant input to machine.

1.3 Objective

To device a system confined to identify language of utterances Japanese. The system will not be restricted to limited vocabulary in other words, the detection is independent of the content of text. The system will not depend on any prosodies feature like intonation, stress and punctuation.

1.4 Organization of Report

The report is described in following way.

- Chapter 1 titled Introduction, describes, Problem Definition, Scope and Objective.
- Chapter 2 titled Project planning and Management, which presents Proposed system and Feasibility study.
- Chapter 3 titled System Analysis, which presents Software, Hardware, Functional, Non-functional, requirements and Software requirement specification.
- Chapter 4 titled System Design presents System Architecture and UML diagrams.
- Chapter 5 titled Implementation and Coding
- Chapter 6 titled Testing

- Chapter 7 titled Result Discussion.
- Chapter 8 titled Conclusion.

1.5 Summary

In this chapter, the introduction of the project is described. In next chapter, the project Planning and management is described.

Chapter 2

Project Planning and Management

The chapter includes details regarding Feasibility study, Risk Analysis, Project Scheduling, Effort Allocation and Cost Estimation as well as a short summary. Section 2.1 includes details regarding feasibility that includes Operating, Economical and Technical feasibility. Details regarding Risk Analysis like Commercial risks, Design risks and other risks as well are discussed in Section 2.2. provides explanation about Project Scheduling while section 2.3 describes the effort allocation i.e effort taken by each group member. Section 2.4 provides information about cost estimation of the project

2.1 Feasibility Study

The feasibility of an approach to automatic language translation that combines recent advances in computer speech recognition and artificial neural networks is discussed. It is shown that artificial neural networks can be used as pattern classifiers that use information about distributions of broad phonetic categories to translate languages. Using artificial languages that differ only by their distribution of stop consonants, feature vectors were extracted from varying amounts of speech from each language. These feature vectors were then used to train an artificial neural network using the back-propagation algorithm. Classification results for two different sets of artificial languages are presented. Feasibility is the analysis of risks, costs and benefits relating to economics, technology and user operation. There are several types of depending on the aspects feasibility they cover. Some important feasibility is as follows:

- Economical Feasibility
- Operational Feasibility
- Technical Feasibility

2.1.1 Economical Feasibility

More commonly known as cost/benefit analysis the procedure is to determine the benefits and savings that are expected from system and compare them with costs, decisions is made to design and implement the system. The part of feasibility study gives the language identification of the economic justification for the language translator.[2] A simple economic analysis that gives the actual comparison of costs and benefits is much more meaningful in such cases. In the system, the organization is most satisfied by economic feasibility. The purpose of an economic feasibility study (EFS) is to demonstrate the net benefit of a proposed project for accepting or disbursing electronic funds/benefits, taking into consideration the benefits and costs to the agency, other state agencies, and the general public as a whole.

2.1.2 Operational Feasibility

Operational feasibility is the measure of how well a proposed system solves the problems, and takes advantage of the opportunities translate during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. The operational feasibility assessment focuses on the degree to which the proposed development project fits in with the existing business environment and objectives with regard to development schedule, delivery date, corporate culture and existing business processes.[3] To ensure success, desired operational outcomes must be imparted during design and development. These include such design-dependent parameters as reliability, maintainability, support-ability, usability, productivity, disposability, sustainability, affordability and others. These parameters are required to be considered at the early stages of design if desired operational behaviours are to be realised. A system design and development requires appropriate and timely application of engineering and management efforts to meet the previously mentioned parameters. A system may serve its intended purpose most effectively when its technical and operating characteristics are engineered into the design. Therefore, operational feasibility is a critical aspect of systems engineering that needs to be an integral part of the early design phases.

2.1.3 Technical Feasibility

This assessment is based on an outline design of system requirements, to determine whether the company has the technical expertise to handle completion of the project. When writing a feasibility report, the following should be taken to consideration:

- A brief description of the business to assess more possible factors which could affect the study
- The part of the business being examined
- The human and economic factor
- The possible solutions to the problem

At this level, the concern is whether the proposal is both technically and legally feasible (assuming moderate cost). The technical feasibility assessment is focused on gaining an understanding of the present technical resources of the organization and their applicability to the expected needs of the proposed system. It is an evaluation of the hardware and software and how it meets the need of the proposed system

2.2 Project Scheduling

The sole purpose of this project is to correctly identify and translate the utterance in Japanese language available in the training set. In this project this project predict the natural language of a piece of text. There are several different approaches to language translation Before we can fit a model, we have to transform our dataset we need to extract features from our list of sentences to create a feature matrix. We do this using character which are sets of n consecutive characters

2.3 Effort Allocation

This project aims at translate language that predict the natural language of a piece of text NLP system is capable of translate Japanese language. Also, the system is independent of the input file format as well as text image.

2.4 Summary

In this chapter, the Project Planning and Management of the project is described. In next chapter, the project Analysis is described.

Chapter 3

Analysis

A set of requirement is used as inputs into the design stages of software development. Requirements are also an important input into the verification process. The various analysis regarding project is given in this chapter.

The organization of chapter is as follows. Section 3.1 presents Requirement Collection and Identification. Section 3.2 presents Hardware Requirements and Software Requirements. Section 3.3 presents Functional Requirements and Non Functional. Section 3.4 presents Software Requirements Specification(SRS). Section 3.5. presents Summary.

3.1 Requirements Collection and Identification

Requirement analysis help the software engineer to better understand the problem they will work to solve. It includes that set of tasks that lead to an understanding: -

3.2 Hardware and Software Requirement

Requirement analysis help the software engineer to better understand the problem they will work to solve. It includes that set of tasks that lead to an understanding: -

3.2.1 Hardware Requirement

Server side:

- Intel core due 3, 5, 7 processor
- Window server environment with python Requirement
- 2 gigabyte of random access memory

3.2.2 Software Requirement

Client side: A user should browser to access the language translation services.

Server side: Server should have support vector support machine library to identify language using NLP.

3.2.3 Product Features

1. User Friendly:-The product will be easy to use.
2. Fast:-It will be quick responsive as latest technologies will be used in it
3. Secured:-Product will be secured as it uses advance security protocols.
4. Portable:- It can be installed on any system.

3.3 Functional and Non-Functional Requirement

3.3.1 Functional Requirements

- The service expected of system to translate language the given input, the system recognizes the language irrespective the type of given input.
- Given text sample system should extract NLP and detect the language.
- Input file sample or image should not be containing any garbage

3.3.2 Non-Functional Requirements

- The user is Expected to have stable internet connection.
- The User is expected to Provide input file.
- The input file should be of specified format.

3.4 Software Requirements Specification (SRS)

The sole purpose of this project is to correctly identify the utterance in any language available in the training set. In this project this project predicts the natural language of a piece of text. There are several different approaches to language translation Before we can fit a model, we have to transform our dataset. we need to extract features from our list of sentences to create a feature matrix. We do this using character which are sets of n consecutive characters.

3.5 Summary

In this chapter, Analysis of the project is described. In the next Chapter, Design of the project is described.

Chapter 4

Design

4.1 System Architecture

The proposed system architecture is shown in fig. The NLP system takes text sample as the input. There are three processing blocks in the architecture; they are, preprocessing blocks, feature Extraction blocks and machine learning block. This cepstral feature vector serves as the input to the next block which is the Machine Learning Block. The SVM in the training phase creates a model based on the input feature vectors for different languages. This model file is used by the classifier in the testing phase to predict language. The final result is the language Translated for the given test samples.

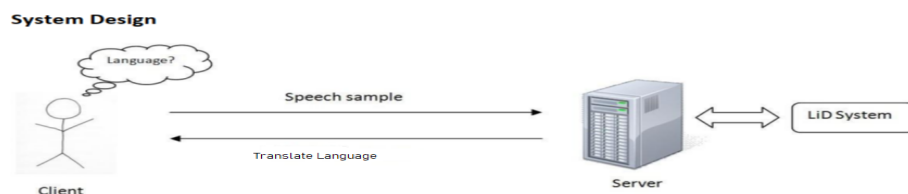


Fig 1. System overview

Figure 4.1: System Architecture

A high level design of the system and it follows client server architecture. The client side comprise of a portal which allows the user to upload an text sample using a browser. The sample is sent over the network to the remote server running an NLP system.

4.2 Data Flow Diagram

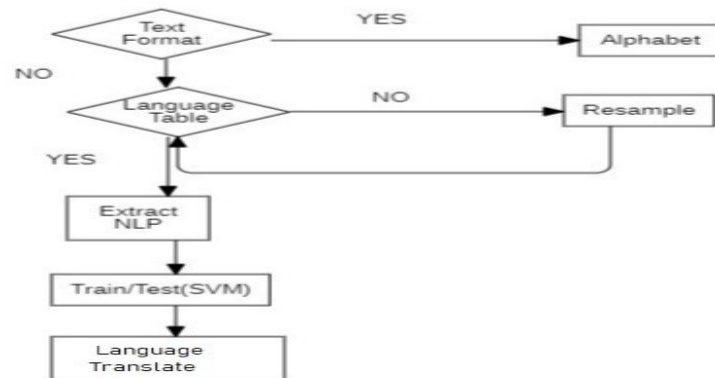


Figure 4.2: Data Flow Diagram

4.3 UML Diagram

4.3.1 Use Case Diagram

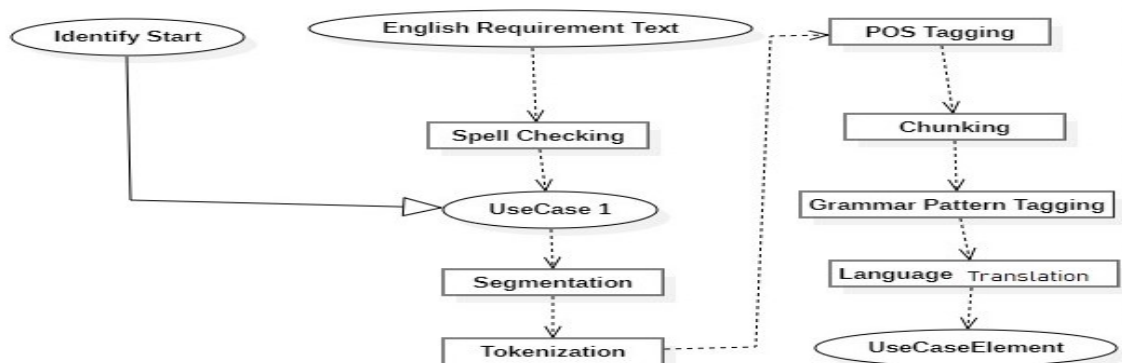


Figure 4.3: Use Case Diagram

4.3.2 Sequence Diagram

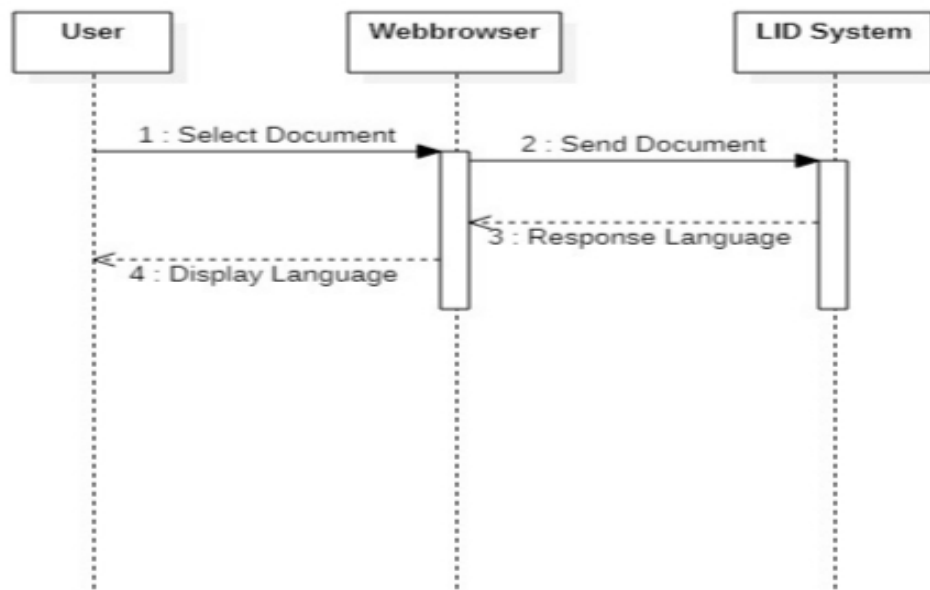


Figure 4.4: Sequence Diagram

4.3.3 Class Diagram

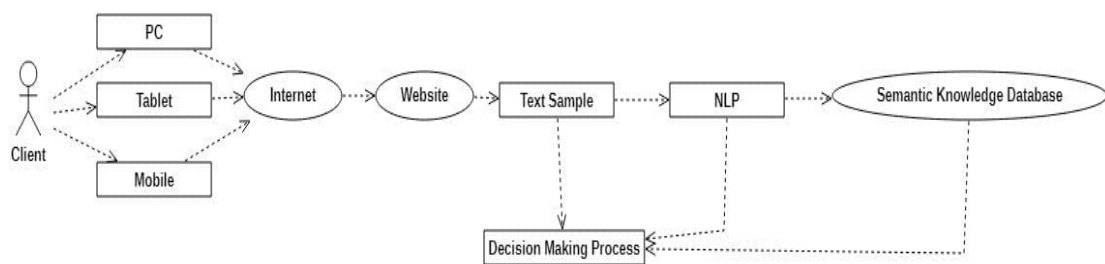


Figure 4.5: Class Diagram

4.3.4 State Diagram

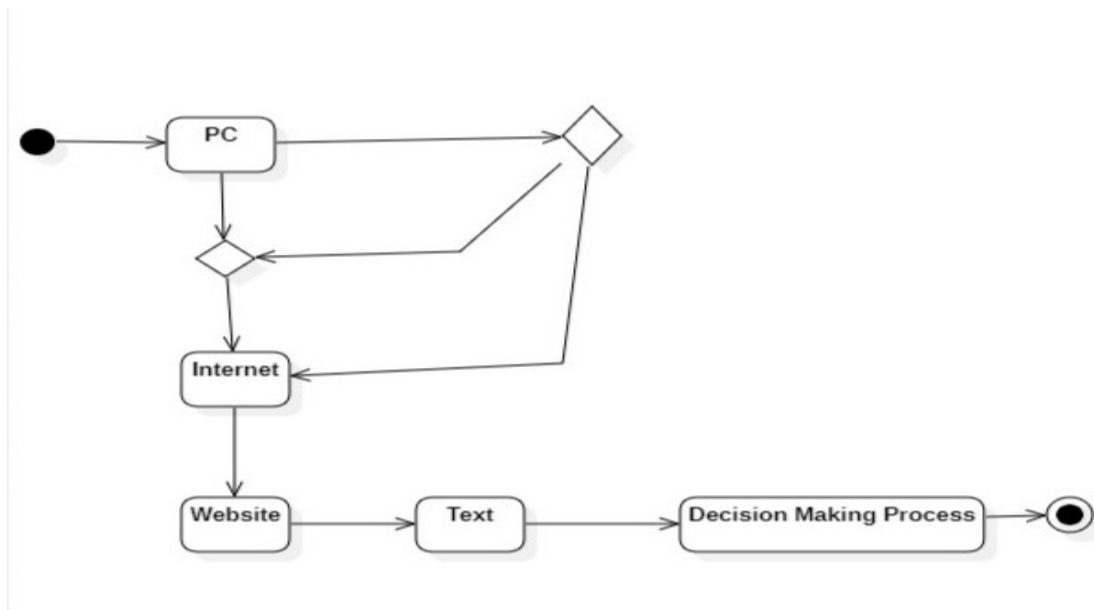


Figure 4.6: State Diagram

4.4 Summary

In this chapter, the Designing of the project is described.

Chapter 5

Implementation and Coding

5.1 Algorithm/Steps

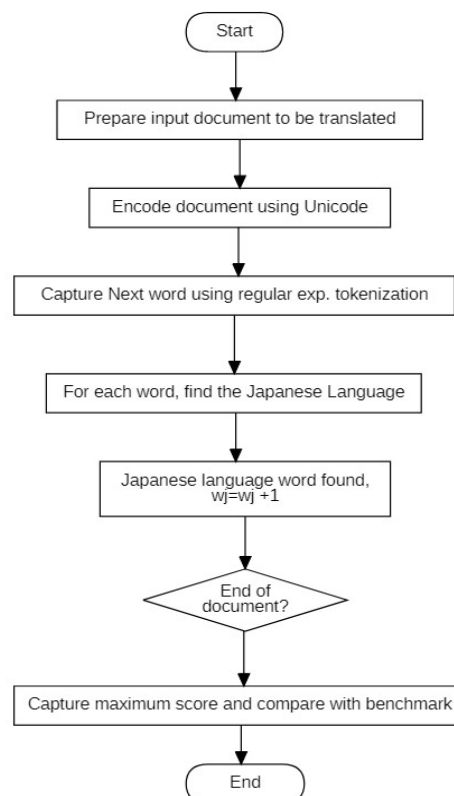


Figure 5.1: Algorithm

5.1.1 Algorithm Used

Input: Set of Spellcheckers, Li; unknown document, D, Benchmark Specification, BM

Output: Language of the input document or document is declared unknown

Begin:

1. Pre-process unknown document and tokenize into words
2. Remove all numeric words and all special characters
3. Convert all words into lowercase and sort words by word length
4. Index word list into set such that each word is searched only once
5. for each word wD
6. if length (w) = n
7. for language in Li
8. if w in lang-word-length (n)
9. lang-word-count = lang-word-count + 1 increment word count
10. end if
11. end for
12. end if
13. end for
14. Compute matrix totals
15. for language in Li
16. if percentage (%) of the highest scoring Spellchecker BM(BenchMark)
17. The document is translated
18. else document is untranslated
19. end for
20. End

5.2 Require Software and Hardware For Development

5.2.1 Software Requirements

1. Visual Studio Code: Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. And same is used for the coding the front-end as well as back-end.
2. Jupyter Notebook: The machine learning module used in the proposed solution is built using Jupyter Notebook code editor.
3. Git and Github: This project uses a version control system to maintain the code and maintain different versions of the project.
4. Formspree: An open-source service that connects the user with the developers in case of registering a complain, appreciating them, and joining the open-source community.
5. Postman: An open-source automated testing software used to test the 'http://' requests of the website.

5.2.2 Hardware Requirements

Server side:

- Intel core due 3, 5, 7 processor
- Window server environment with python Requirement
- 2 gigabyte of random access memory

5.3 Summary

In this chapter, Implementation Details and Implementation Environment is discussed. In the next, chapter Testing is presented.

Chapter 6

Testing

6.1 Manual Testing

Manual testing is the process of manually testing software for defects. It requires a tester to play the role of an end user whereby they use most of the application's features to ensure correct behavior. For the testing of the proposed solution below mentioned is the software program to run the test cases designed to check the correct working of the software.

```
{
  "host": {
    "connectionReferenceName": "shared_translatorv2",
    "operationId": "Translate"
  },
  "parameters": {
    "to": "en",
    "body/Text": "ウィリー王子の護衛依頼を完遂し、一つ肩の荷が下りた涼。王子の水魔法の師匠ポジションまでゲットした彼は、セーラと合流し、ほくほくとスイーツの食べ歩きに興じていた。エルフの自治庁と一緒に訪問したりと距離を縮めていく二人だったが、別行動の翌日、神殿地下から魔物の大群が出現！ 更には、原因究明に努める涼とアベルの前で、空から降ってきた\"島\"が王城に突き刺さる事態に。異常事態が次々と発生する中、仲間を守ろうと奮闘するセーラにも強敵・アークデビルが迫りー「僕の大事な人を傷つけてタダで済むとでも？」 王国へと牙をむいた敵国の策謀に、涼が超高压水流【ウォータージェット】で迎え撃つ！ 大人気シリーズ・最強水魔法使いの気ままな冒険譚・第4弾！",
    "from": "ja"
  }
}
```

Figure 6.1: Manual Un-Translated text

```

"body": "After completing Prince Willie's escort request, Ryo has been
unloaded by one shoulder. After getting the prince's water magic master position,
he joined Sarah and enjoyed eating hokuhoku and sweets. They were two people who
shortened the distance by visiting the self-
governing body of the elves together, but the day after another action, a horde of
demons appeared from the temple basement!
In addition, in front of Ryo and Abel trying to investigate the cause, the \"
island\" that came down from the sky pierced the Wang Castle. While abnormal
situations occur one after another, a strong enemy, Arc Devil, who struggles to --
protect her friends, also approaches - \"Even if you hurt my important person and
you can do
it for free?\" Ryo intercepts the enemy's conspiracy that peeled off his fangs to
the kingdom with an ultra-
high pressure water flow [water jet]! The 4th casual adventure story of the
popular series, the strongest water wizard!"

```

Figure 6.2: Manual Translated text

6.2 Automated Testing

This section discuss about the automated testing associated with the solution. In order to automate the testing of the website the following code is written.

```

{
  "host": {
    "connectionReferenceName": "shared_translatorv2",
    "operationId": "Translate"
  },
  "parameters": {
    "to": "en",
    "body/Text": "火属性の魔法を使う者は他の属性より魔力が多く、攻撃力が高い。\n\n
そのせいか、昔から好戦的で気性が荒いと言われている。\n\n ただ、攻撃力に長けているため、
レイブロン公爵家一族は王家の剣として首都に近い場所に領地をもらい、王族たちを傍で守ってきた。
一方、華やかな首都暮らしに慣れているだけに、屋敷に足を踏み入れた瞬間から豪華な装
飾品や美術品が並び、目が眩みそうになった。\n\n ヘルミーナは久しぶりに社交界デビューを思
い出した。\n\n どのパーティーに参加しても場違いな自分を感じてきたが、今日はとくに緊張し
ていた。\n\n やはり招待客の大半が火の魔法を使う貴族だからか。\n\n 額にじわりと汗が滲ん
でくる。\n\n\n エーリッヒのエスコートで侯爵家の会場ホールに入ると、皆の視線が二人に集
中した。\n\n 堂々と歩きたくても突き刺さってくる視線に、ヘルミーナは転ばないようにするの
が精一杯だった。\n\n 招待客は次から次にやって来るのに、いつも以上に見られている気がする
。\n\n やはり怖くなって俯いてしまうと、エーリッヒはそんなヘルミーナに溜め息をついた。
\n\n\n「しっかりしてくれ。僕たちは一族の代表として参加しているんだ」\n\n「.....ご、ごめん
なさい」",
    "from": "ja"
  }
}

```

Figure 6.3: Auto Un-Translated text

```

{
  "body": {
    "name": "translated text",
    "value": "Those who use fire spells have more magical power and higher attack power than other attributes.\n\nPerhaps because of that, it has long been said to be belligerent and rough-tempered.\n\nHowever, due to their strong attack, the Dukes of Raybron's family received territories near the capital as the swords of the royal family and protected the royals by their side.\n\nOn the other hand, because he was used to living in the gorgeous capital, from the moment he set foot in the mansion, gorgeous ornaments and art works lined up, and I almost dazzled.\n\nHermina recalled her social debut after a long absence.\n\nI've felt that I'm out of place no matter which party I've attended, but I was particularly nervous today.\n\nAfter all, the majority of the invited guests are nobles who use the magic of fire.\n\nSweat oozes gently on the forehead.\n\nWhen entering the hall of the Marquis family by the escort of Erich, everyone's glance concentrated on two people.\n\nEven if she wanted to walk in a dignified manner, Hermina did her best not to fall down in the piercing gaze.\n\nI feel that the invited guests are watching more than usual, even though they come one after another.\n\nAfter all, Erich sighed at such Hermina when becoming scared and it took a bird's eye.\n\n'Hold on, we're on behalf of the clan.'\n\n「..... I'm sorry.»
  }
}

```

Figure 6.4: Auto Translated text

6.3 Test Cases Identification and Execution

Test Case	Input	Expected Output	Actual Output	Status
Translate Text	Untranslated Text	Successful Translation	Translation	Pass

Table 6.1: Test Case For Translation

6.4 Summary

In this chapter, Test Methodology in system testing is presented. In the next chapter Result Analysis is presented.

Chapter 7

Result and Discussion

After training the dataset on three different machine learning model the outcome that has been extracted is as follows, the linear regression model performed the best with the score of approximately 91%, followed by the lasso regression with an approximate score of 91%, lastly the decision tree score almost 90% score when trained on a dataset.

As the final decision to choose over these machine learning models the optimal choice had to be the linear regression model and hence, that is the reason of using the same in the proposed solution.

7.1 Result

True Positive 260	False Positive 19
True Negative 20	False Negative 135

Table 7.1: Confusion Matrix

7.2 Summary

By considering various chapters the result is discussed. The different aspect of results are discussed in above mentioned section which will be based for the conclusion. In the next chapter Conclusion is presented.

Chapter 8

Conclusion

- The current system is capable of translating Japanese language to English Successfully.
- The system can distinguish between Japanese and English words.
- The Translation accuracy is approximately 91%.

Bibliography

- [1] Miyako Inoue. Speech without a speaking body: “japanese women’s language” in translation. *Language & Communication*, 23(3-4):315–330, 2003.
- [2] Makoto Nagao. *A framework of a mechanical translation between Japanese and English by analogy principle*. na, 1984.
- [3] Robert Weschler. Uses of japanese in the english classroom: Introducing the functional-translation method. *Kyoritsu Women’s University Department of International Studies Journal*, 12:87–110, 1997.