**D.K.T.E. Society’s Textile and Engineering Institute, Ichalkaranji.**

**(An Autonomous Institute, Affiliated to Shivaji University, Kolhapur)**

**Accredited with ‘A+’ Grade by NAAC**

Department of Computer Science & Engineering 2023-2024



THE MINI PROJECT REPORT ON

**QuickSumm - Text Summarizer**

Under the guidance of: Prof. S. D. Rane

**DEVELOPED BY:**

|  |  |
| --- | --- |
| 1. Ms. Rutuja Chougule | 21UAI010 |
| 2. Ms. Prajakta Darekar | 21UAI011 |
| 3. Ms. Pooja Hanamane | 21UAI019 |
| 4. Ms. Arya Jadagale | 21UAI020 |

**D.K.T.E. Society’s Textile and Engineering Institute, Ichalkaranji.**

**(An Autonomous Institute, Affiliated to Shivaji University, Kolhapur)**

Department of Computer Science & Engineering

## CERTIFICATE

This is to certify that,

|  |  |
| --- | --- |
| **Name of the students** | **PRN No.** |
| 1. Ms. Rutuja Chougule | 21UAI010 |
| 2. Ms. Prajakta Darekar | 21UAI011 |
| 3. Ms. Pooja Hanamane | 21UAI019 |
| 4. Ms. Arya Jadagale | 21UAI020 |

Have successfully completed the mini project titled,

**QuickSumm - Text Summarizer**

In T.Y.B.Tech .CSE(AIML) academics .This is the record of their work carried out during academic year 2023-2024

Date: 14 May 2024 Place: Ichalkaranji

Prof.S.D.Rane Prof.Dr.S.K.Shirgave

[PROJECT GUIDE] [HOD]

**DECLARATION**

We the undersigned students of T.Y.C.S.E. declare that the Project work report entitled “QuickSumm - Text Summarizer” written and submitted under the guidance of Prof.S.D.Rane is our original work. The empirical findings in this report are based on the data collected by us. The matter assimilated in this report is not reproduction from any readymade report.

Date:14 May 2024 Place: Ichalkaranji

|  |  |
| --- | --- |
| **Name of the students** | **PRN No.** |
| 1. Ms. Rutuja Chougule | 21UAI010 |
| 2. Ms. Prajakta Darekar | 21UAI011 |
| 3. Ms. Pooja Hanamane | 21UAI019 |
| 4. Ms. Arya Jadagale | 21UAI020 |

**INDEX**

**CONTENTS PAGE**

1. Abstract 5
2. Problem Statement 6
3. Problem Description 6
4. Purpose 7
5. Scope of the Project 8
6. Requirement Analysis

* Functional Requirements 9
* Non-Functional Requirements 10

1. System Requirements

* Software Requirements 11
* Hardware Requirements 11

1. System Design 12

* Use case Diagram 13
* Sequence Diagram 14
* Data flow 15

1. Testing 16
2. Snap shots 17
3. Conclusion 19
4. References 20

**ABSTRACT**

The NER-based Text Summarizer is an advanced natural language processing (NLP) system designed to automatically identify and extract named entities from unstructured text data, including individuals, organizations, locations, dates, and more. Leveraging state-of-the-art Named Entity Recognition (NER) technology, the system processes input text to identify relevant entities and synthesizes this information into concise and coherent summaries. By integrating NER functionality with text summarization algorithms, the system offers a seamless solution for efficiently distilling key information from large volumes of textual content. The abstract nature of the system lies in its ability to automate the extraction and summarization process, empowering users to quickly grasp the essential content of input text without extensive manual effort.

**PROBLEM STATEMENT**

Developing a Named Entity Recognition (NER) model with an integrated text summarizer to efficiently extract and summarize key information articles.

**PROBLEM DESCRIPTION**

This project tackles the challenge of efficiently extracting essential information from vast unstructured text data. By combining Named Entity Recognition (NER) technology with text summarization algorithms, it aims to automatically identify and extract key entities like people, organizations, and dates from diverse text sources. Through harnessing natural language processing (NLP) techniques, the system synthesizes these entities into concise summaries, alleviating the need for extensive manual efforts and facilitating quick comprehension of input text across various domains, from news articles to social media posts.

**PURPOSE**

The purpose of our project is to address this challenge by developing an advanced solution utilizing Named Entity Recognition (NER) technology coupled with a text summarization algorithm. By harnessing the power of natural language processing (NLP) techniques, our goal is to create a robust system capable of automatically identifying and extracting named entities such as people, organizations, locations, dates, and more from unstructured text. Furthermore, we aim to synthesize these identified entities into concise and coherent summaries, enabling users to quickly grasp the core content of the input text without the need for extensive reading.

**SCOPE OF PROJECT**

The project scope encompasses several key areas. Firstly, it involves the design and training of a Named Entity Recognition (NER) model to accurately identify named entities within text data. Following this, the project aims to integrate the NER model with a text summarization algorithm to produce concise summaries that capture essential information. Rigorous evaluation and iterative optimization of model performance will be conducted to ensure accuracy. Scalability is a priority, with efforts directed towards handling large volumes of text data efficiently, and exploring deployment options for accessibility. Additionally, the project includes the development of comprehensive documentation and an intuitive user interface to facilitate seamless interaction with the system.

**REQUIREMENT ANALYSIS**

Software requirement is a functional or non-functional need to be implemented in the system. Functional means providing a particular service to the user.

Software requirement can also be a non-functional, it can be a performance requirement.

The following are the functional and non-functional requirements of

**Text Summarization using NER model**

**Functional Requirements:**

Tokenization:

* + Tokenization involves breaking down the input text into smaller units called tokens, typically words or subwords. The system should include tokenization functionality to preprocess the text before further analysis, ensuring that each token is processed individually.
* Part-of-Speech (POS) Tagging:
  + Part-of-speech tagging is the process of assigning a grammatical category (such as noun, verb, adjective, etc.) to each word in a sentence. The system should include POS tagging functionality to analyze the syntactic structure of the text and extract additional linguistic features for use in the summarization process.
* Abstractive Summarization:
  + The text summarization algorithm should be able to generate abstractive summaries that paraphrase and condense the original text, rather than simply extracting sentences.
* Entity-based Summarization:
  + The system should prioritize information related to identified named entities when generating summaries, ensuring that key entities are adequately represented.

Proper Noun Handling:

* + Proper noun handling involves appropriately recognizing and processing proper nouns, such as names of people, organizations, and locations. Proper handling ensures that important entities are properly identified and represented in the summaries.
* Contextual Understanding:
  + The summarization algorithm should exhibit an understanding of the context in which named entities appear, ensuring that summaries capture the essence of the text.

**Non-Functional Requirements:**

* Performance
* Security
* Availability
* Reliability
* Serviceability
* Maintainability
* Regulatory
* Manageability
* Usability

**System Requirements**

System requirements are the required specifications a device must have in order to use certain [hardware](https://techterms.com/definition/hardware) or [software](https://techterms.com/definition/software).

**Software Requirements:**

**Windows:**10 or newer

**Backend:** Python , flask

**Frontend:** HTML, CSS , JavaScript

## Hardware Requirements:

**Processor (CPU):** Dual-core processor or higher

**RAM:** 4GB or more

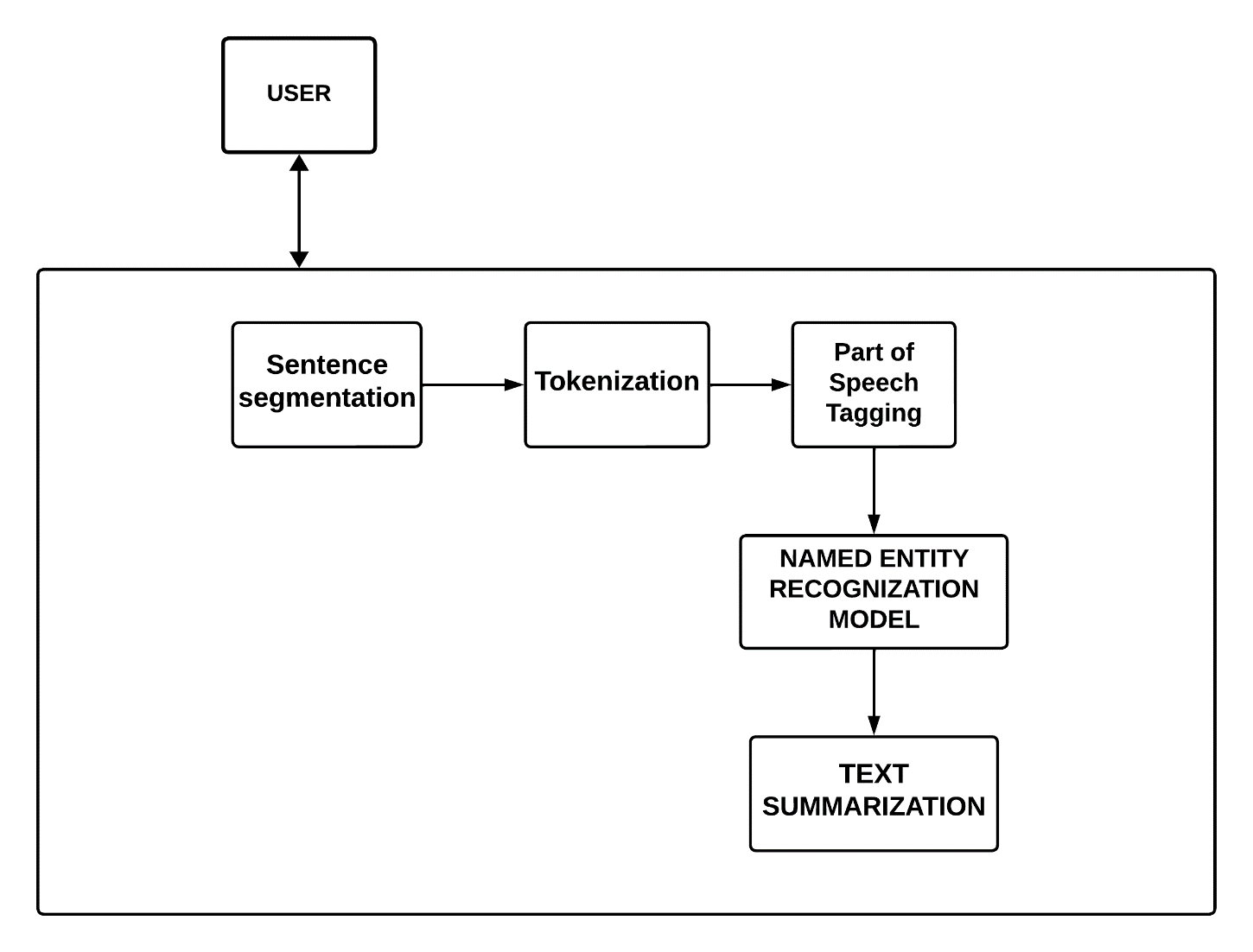
**Storage:** Minimum 20GB of free disk space

**Internet Connection:** Broadband connection for seamless online access and data transfer

**Graphics Card:** Basic graphics capabilities for a smoother user interface experience

**Display:** Monitor with a resolution of 1280x720 pixels or higher

**SYSTEM DESIGN**

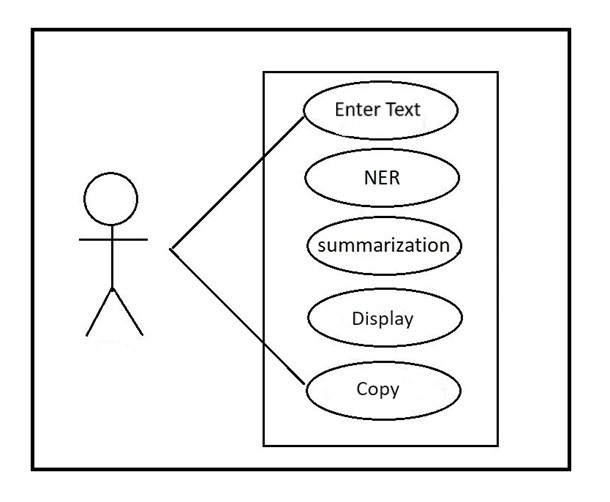


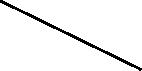
System Architecture

Following are the system designs for the mentioned project:

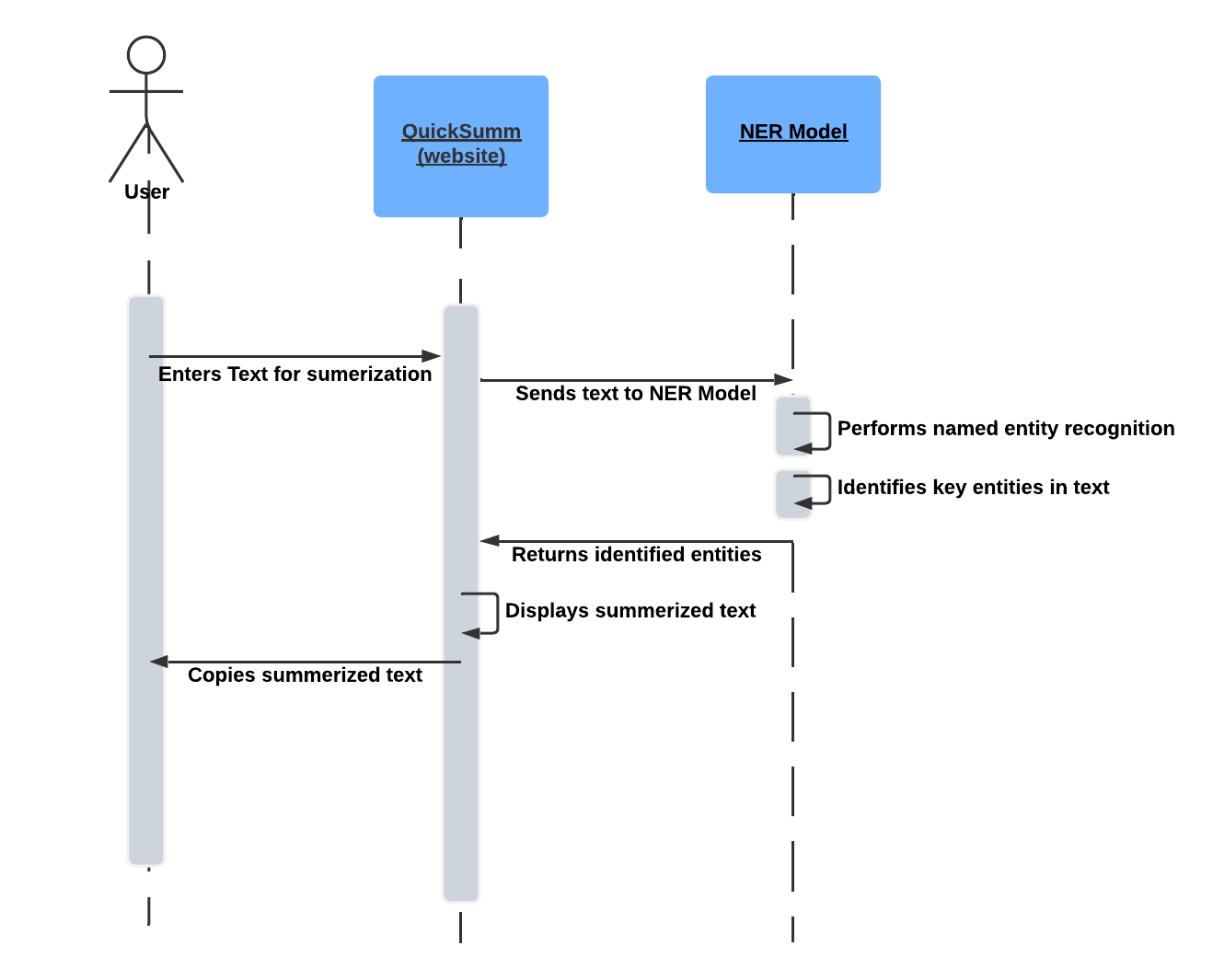
* Use case Diagram
* Sequence Diagram
* Data Flow Diagram

**USE CASE DIAGRAM:**

****

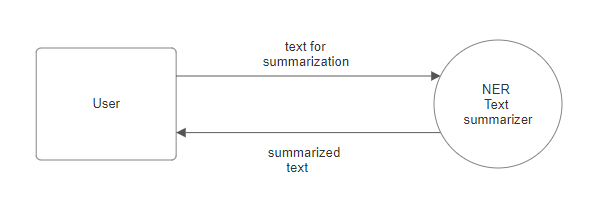


**SEQUENCE DIAGRAM:**

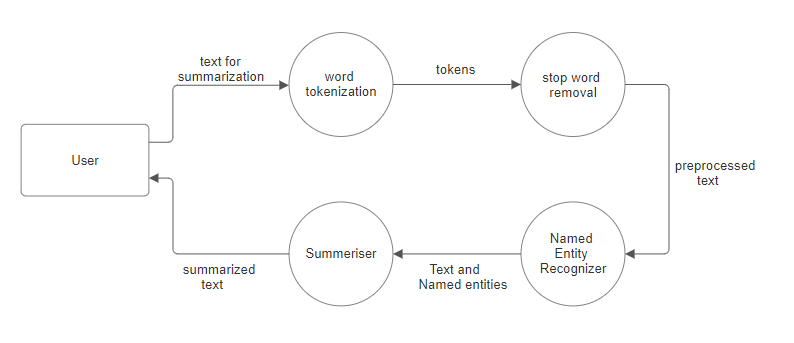


**Data Flow Diagram**

**Level 0 :**



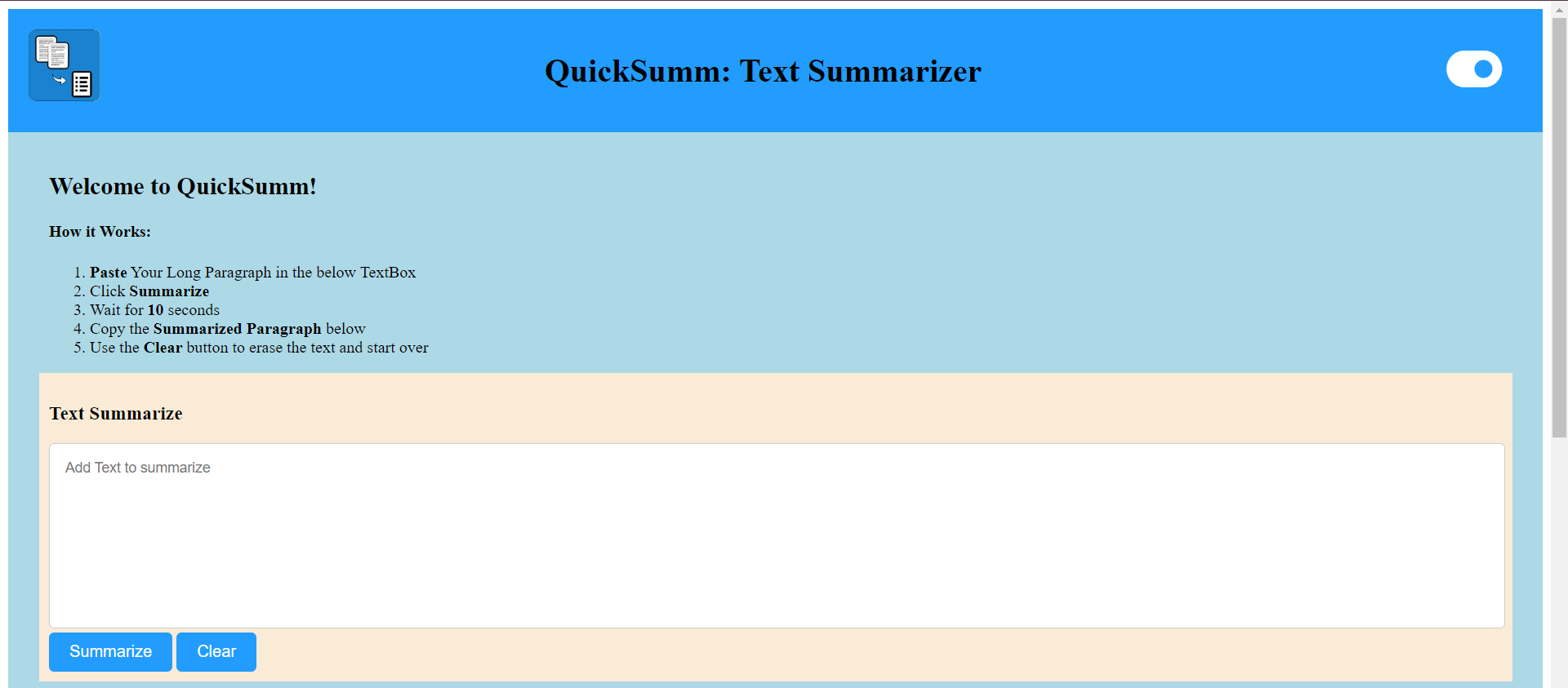
**Level 1 :**

****

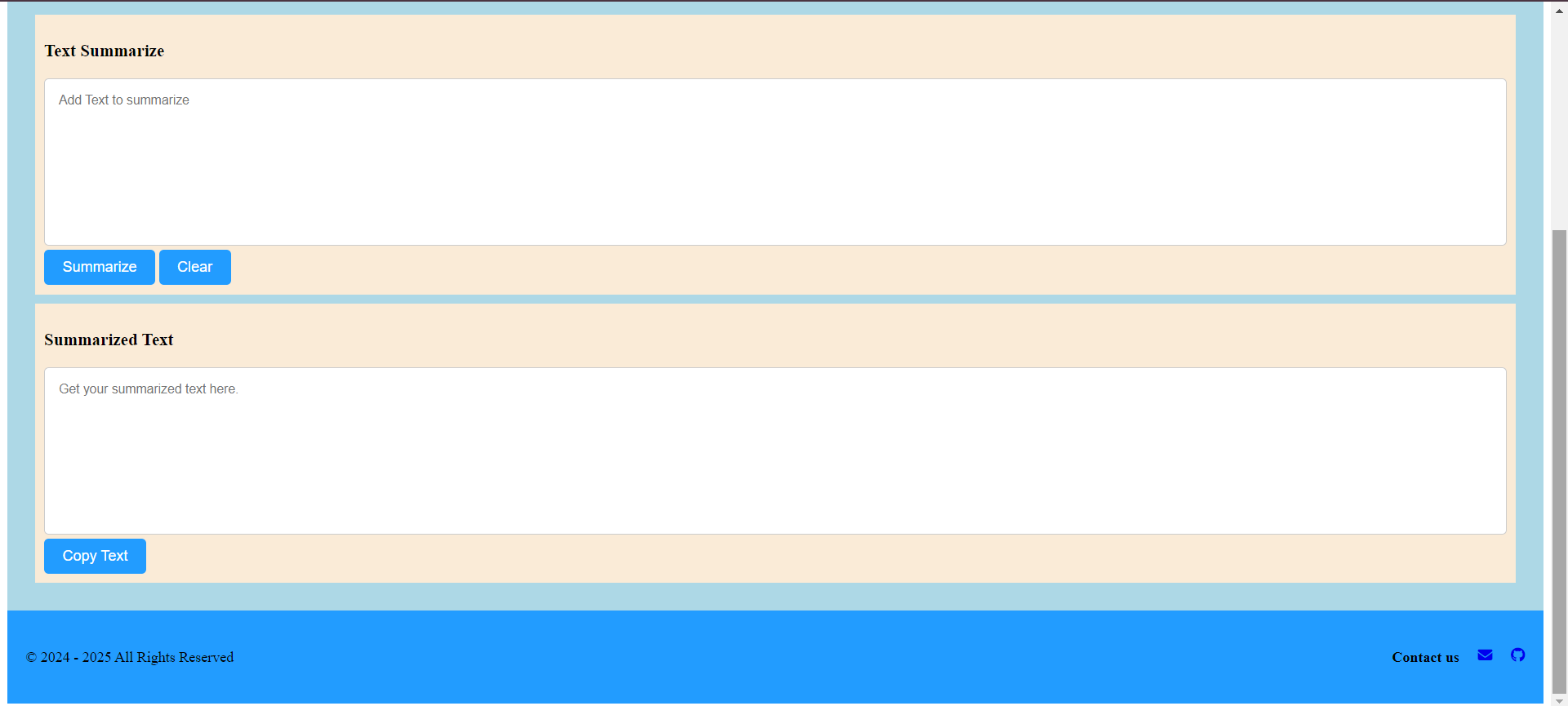
**Testing**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Case ID | Test Cases | Expected Output | Actual Output | Result  Pass/Fail |
| 1. | Summarize a long text | Consize summary of the input text | [Actual summary] | Pass |
| 2. | Summarize text with technical terms | Summary understandable to non-experts | "The nucleus is the central part of an atom." → "The core of an atom." | Pass |
| 3. | Summarize text with multiple paragraphs | Summary covering all main points | [Actual summary] | Pass |
| 4. | Summarize a full research paper | Comprehensive summary covering introduction, methodology, results, and conclusion | [Actual summary] | Pass |
| 5. | Summarize an article | Brief summary capturing main points and arguments | "The article discusses the impacts of climate change on agriculture." → "Climate change's effects on agriculture." | Pass |

**Snapshot**

****

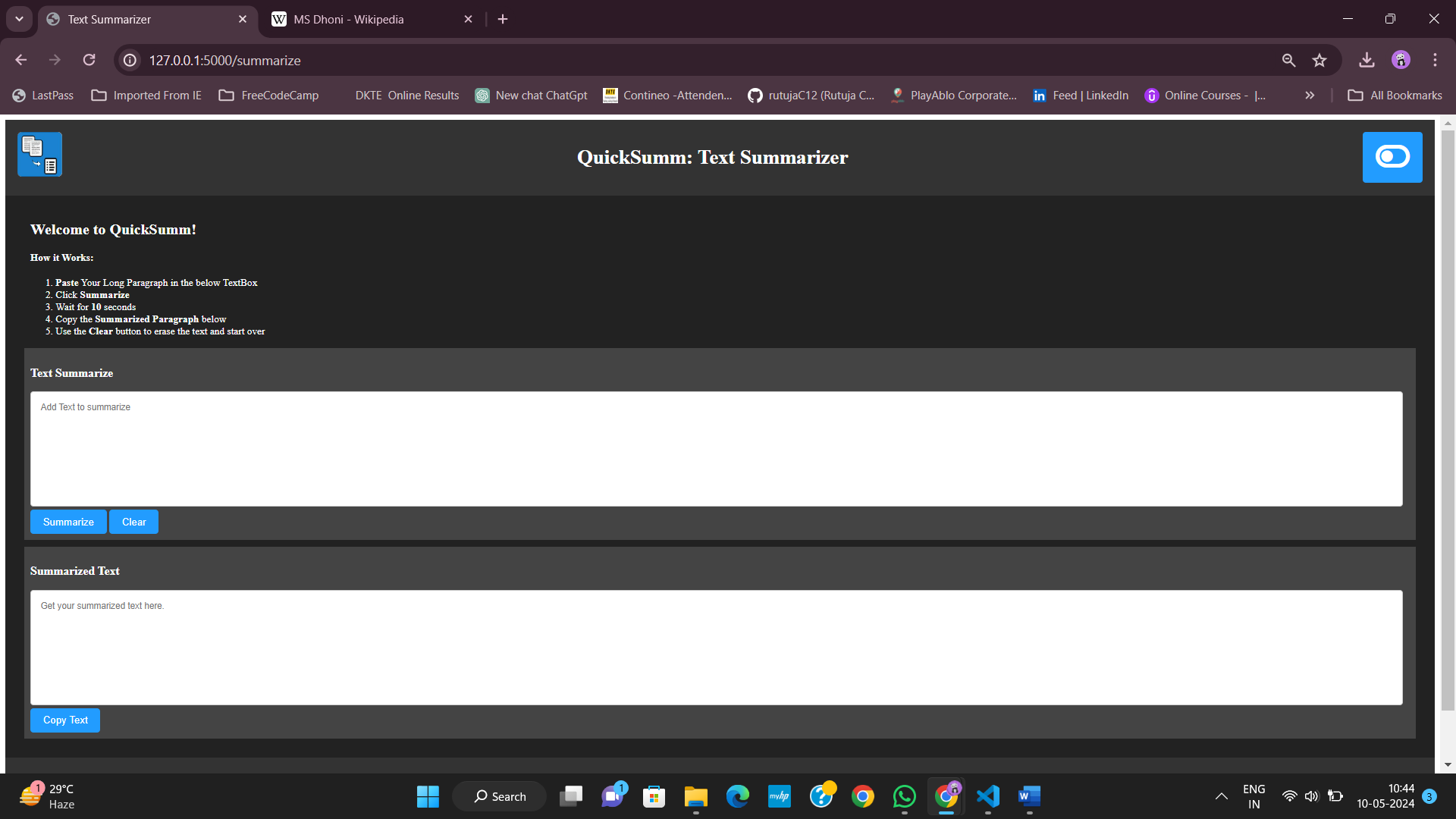
**Fig1: Input Field**

****

**Fig2: Output Field**

****

**Fig3: Sample test case**

****

**Fig4: Dark mode functionality**

**CONCLUSION**

Our text summarizer project with Named Entity Recognition (NER) is great at making short summaries from all sorts of texts. NER helps us pick out important names and places, making the summaries more useful. The website we made is easy to use, so anyone can quickly get helpful summaries. This project shows how important NER is for making sense of lots of text. We can keep improving it to make it even better at giving useful information, helping people understand big chunks of text faster and easier.

**REFERENCES**

[**www.w3school.com**](http://www.w3school.com)

**www.youtube.com**