Copy Cat Bot

S. R. S. Report

Rishabh Sengar

(161500451)

Prashant Raghav

(161500401)

Priyank Kaushik

(161500414)

Ram Kamra

(161500437)

**Table of Contents**

**1. Introduction 1**

1.1 Purpose 1

1.2 Scope 1

1.3 Definitions, Acronyms, and Abbreviations 2

1.4 References 2

**2. General Description 3**

2.1 Product Perspective 3

2.2 Tool Used 4

2.3 General Constraints 4

2.4 Assumptions and Dependencies 4

**3. Specific Requirements 5**

3.1 Hardware and Software Interfaces 5

3.2 Use Cases 6

3.3 Classes / Objects 6

**4. Analysis Models 7**

4.1 Sequence Diagrams 7

**A. Appendices 8**

**1. Introduction**

* 1. **Purpose**

The Project’s purpose is to ‘generate the text', i.e., to generates the text by sampling the next character based on the probability distribution of the last character of the current sequence.

* 1. **Scope**

Machine writing entire new chapters of popular novels like Game of Thrones and Harry Potter.

Dialogue Generation and Script writing.

Essence of Great Leaders speeches can be copied with the help of text generation

**1.3 Definitions, Acronyms, and Abbreviations**

**Deep Learning:** A machine learning method that stimulates the neural network inhuman brain.  
**Python:**​​ One of the most commonly used programming languages.  
**IEEE:***​​* Institute of Electrical and Electronics Engineers  
**NLP*:***​​ Natural Language Processing  
**Recurrent Neural Network (RNN)**  : They are a type of Neural Network where the output from previous step are fed as input to the current step.

**Long short-term memory (LSTM**) units (or blocks) are a building unit for layers of a recurrent neural network (RNN). A RNN composed of LSTM units is often called an LSTM network. A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate.

**Web Scrapping :** Itis a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

**1.4 References**

Wikipedia – [www.wikipedia.com](http://www.wikipedia.com/)

Software Engineering, Seventh Edition, Ian Sommerville.

Dash – [www.plotly.com/dash](http://www.plotly.com/dash)

Datacamp – [www.datacamp.com](http://www.datacamp.com)

Researchgate – [www.researchgate.com](http://www.researchgate.com)

Kdnuggets – [www.kdnuggets.com](http://www.kdnuggets.com)

Analyticsvidhya – [www.analyticsvidhya.com](http://www.analyticsvidhya.com)

Kaggle – [www.kaggle.com](http://www.kaggle.com/)

Github – [www.github.com](http://www.github.com/)

Keras – [www.keras.io](http://www.keras.io)

Medium – [www.medium.com](http://www.medium.com)

**2. General Description**

The following section describes the various requirements and modules of the Project.

**2.1 Product Perspective**

**Drawbacks :**

Requires a large amount of data.

Extremely computationally expensive to train.

Determining the topology/flavor/training method/hyperparameters for deep learning is a

black art with no theory to guide you.

**Our Plan :**

To generate similar speech.

**2.2 Tool Used:**

**Anaconda :** Anaconda is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications, that aims to simplify package management and deployment.

**Python :** Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

**Spyder :** Itis a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It offers a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

**2.3 General Constraints:**

This system is dataset dependent.

**2.5 Assumption and Dependencies**

Scrapped data is from limited sources.

Operating system should be Windows or Linux.

Hyperparameters are predefined.

**3. Specific Requirements**

**3.1 Hardware and Software Interfaces**

**Minimum**

|  |  |
| --- | --- |
| **HARDWARE** | **SOFTWARE** |
| 1.2 GHz processor | Windows/Linux |
| 8Gb RAM | Python 3, Git, Anaconda |
| 10 GB HDD | Keras, Tensotflow |

**Recommended**

|  |  |
| --- | --- |
| **HARDWARE** | **SOFTWARE** |
| 2GHz Processor and above | Windows/Linux |
| NVIDIA GPU(940mx and above) | Anaconda |
| 16Gb RAM and above | Python 3 |
| 50 GB HDD and above | Tensorflow , Keras |

**3.2 USE CASE Diagram**

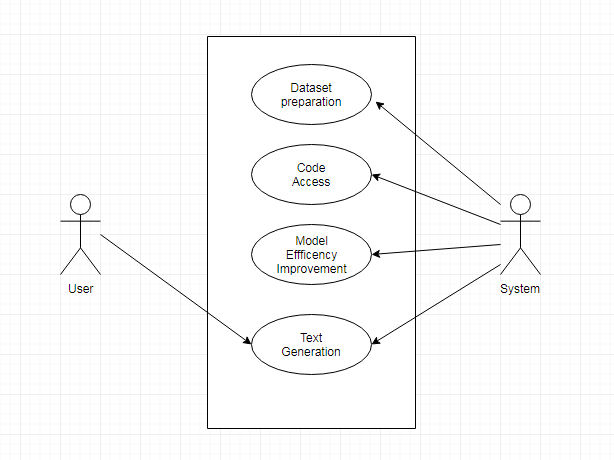
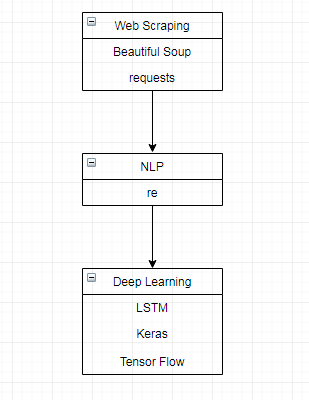


Fig1: Use case diagram

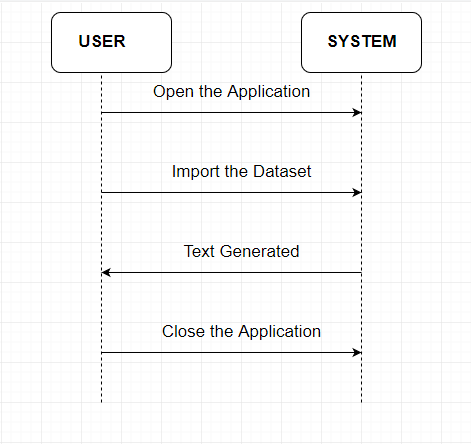
|  |  |
| --- | --- |
| **USE CASE** | **DESCRIPTION** |
| Update dataset | Admin can update dataset. |
| Code access | Admin can update/modify the code as per the requirements. |
| Model Efficiency Improvement | Admin can increase the efficiency by updating the models |
| Text Generation | User gets the generated text. |

**3.3Class / Object Diagram**



|  |  |
| --- | --- |
| **Class** | **Description** |
| Web Scrapping | Data extraction from different sources to prepare dataset |
| NLP(Natural Language Processing) | Manipulating text data. |
| Deep learning | A machine learning method that stimulates the neural network in human brain. |
| LSTM(Long Short Term Memory) | A building unit for layers of a recurrent neural network (RNN). |

**3.4 Sequence Diagram**



**Appendices**

**A**

*Acronym*

*Abbreviation*

*API*

*Anaconda*

**C**

*Class Diagram*

*Constraints*

**D**

*Datase*

*Deep Learning*

**I**

*IEEE*

**K**

*Keras*

**L**

*Linux*

**M**

*Machine learning*

**N**

*NLP (Natural Language Processing)*

**O**

*os*

**P**

*Python*

*Pandas*

**R**

*Random*

**S**

*Sequence Diagram*

*string*

**T**

*Tensorflow*

**U**

*Use Case*

**W**

*Windows*

*Web Scrapping*

*Warnings*