

# Named Entity Recognition

**Name – Sachin Suthar**

## Introduction:

Natural Language Processing (NLP) has become a crucial field in machine learning, enabling machines to understand and analyze human language. Named Entity Recognition (NER) is a fundamental task in NLP, focusing on identifying and classifying entities such as names of people, organizations, and locations within a given text. This project explores the application of deep learning techniques, specifically Bidirectional Long Short-Term Memory (BiLSTM), for efficient Named Entity Recognition.

## Project Overview:

The project's primary objective is to develop and implement a robust NER system that can accurately identify and classify entities in text data. Leveraging the power of deep learning, the BiLSTM model is employed to capture contextual information and improve the overall accuracy of entity recognition.

## Project Benefits

- **Enhanced Information Extraction:** The NER model facilitates the extraction of valuable information from unstructured text, aiding in knowledge discovery and analysis.
- **Automation of Entity Identification:** Automation of the entity identification process reduces the manual effort required for text annotation, saving time and resources.
- **Improved Information Retrieval:** The accurate identification of entities enhances information retrieval systems by providing more relevant and context-aware results.

## Key Features

- **Bidirectional LSTM Model:** Utilizing Bidirectional Long Short-Term Memory networks to capture context information from both forward and backward directions.
- **Tokenization and Padding:** Preprocessing techniques like tokenization and padding ensure standardized input for the model.

- **Categorical Crossentropy Loss:** Training the model with categorical crossentropy loss for effective entity classification.
- **TimeDistributed Dense Layer:** Applying a dense layer to each time step independently, allowing for fine-grained predictions.

## **Dataset:**

The dataset used in this project is the "ner\_dataset.csv," containing information about words, parts of speech (POS), and entity tags. The dataset is preprocessed to handle missing values and group sentences.

## **Data Preprocessing:**

1. Tokenization: Words are mapped to unique indices for processing.
2. Padding: Input sequences are padded to have consistent lengths.
3. One-Hot Encoding: Tags are one-hot encoded to categorical format.

## **Model Architecture:**

The BiLSTM model is employed for Named Entity Recognition:

1. Embedding Layer: Converts word indices to dense vectors.
2. Bidirectional LSTM: Captures contextual information in both forward and backward directions.
3. TimeDistributed Dense Layer: Applies a dense layer to each time step independently.

## **Model Training:**

The model is trained using the Adam optimizer and categorical crossentropy loss. The training process involves 25 epochs with batch size 1000. The training and validation loss are monitored.

## **Resources Used:**

- **TensorFlow and Keras:** Frameworks used for building and training the deep learning model.
- **Pandas:** Library for data manipulation and preprocessing.

- **Google Colab:** Cloud-based platform utilized for collaborative coding and GPU support.

## Results:

The training results are stored in a DataFrame for analysis. The BiLSTM model's performance is evaluated based on accuracy and loss metrics.

Some screenshots of the output:

```
import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')
text = nlp('Hi, My name is Sachin Suthar \n')
displacy.render(text, style = 'ent', jupyter=True)
```

Hi, My name is Sachin Suthar PERSON

```
import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')
text = nlp('Hi, My name is Sachin Suthar \n I am pursuing my BTech from Vit Bhopal University \n')
displacy.render(text, style = 'ent', jupyter=True)
```

Hi, My name is Sachin Suthar I am WORK\_OF\_ART pursuing my BTech ORG from Vit Bhopal University ORG

```
import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')
text = nlp('Narendra Modi is the Prime Minister of India ')
displacy.render(text, style = 'ent', jupyter=True)
```

Narendra Modi PERSON is the Prime Minister of India GPE

```
import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')
text = nlp('It is my first NLP project and I have done this project under the guidance of Dr. Anirban Bhowmick sir \n')
displacy.render(text, style = 'ent', jupyter=True)
```

It is my **first** **ORDINAL** **NLP** **ORG** project and I have done this project under the guidance of Dr. **Anirban Bhowmick** **PERSON** sir