# Named Entity Recognition (NER)

"PERFECT SQUARES"- JAI SAXENA, PANKAJ BALCHANDANI, DURGESH MISHRA, NIRANJAN SOLANKI

## About the Project

Named entity recognition (NER) sometimes referred to as entity chunking, extraction, or identification is the task of identifying and categorizing key information (entities) in text.

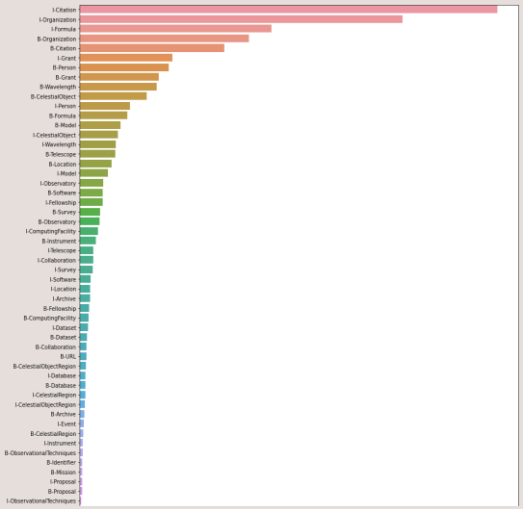
NER is probably the first step towards information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. NER is used in many fields in NLP and used in Biomedical, Human recourses, Content Classification, Customer support, etc.

## Dataset

**WISEP Dataset**

- 1753 training examples
- 63 different types of labels
- 1366 samples for validation

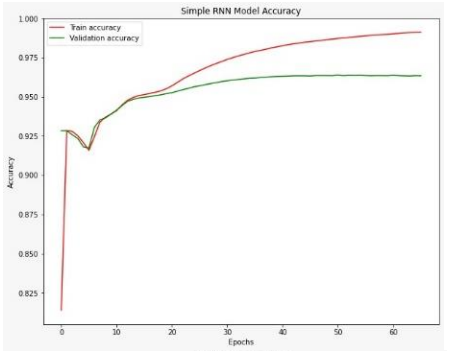
The dataset is created by The Workshop on Information Extraction from Scientific Publications (WIESP) which is a forum to foster discussion and research using Natural Language Processing and Machine Learning.

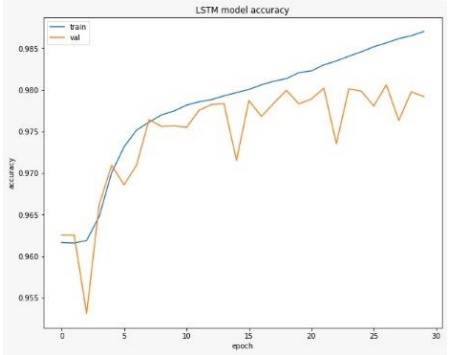


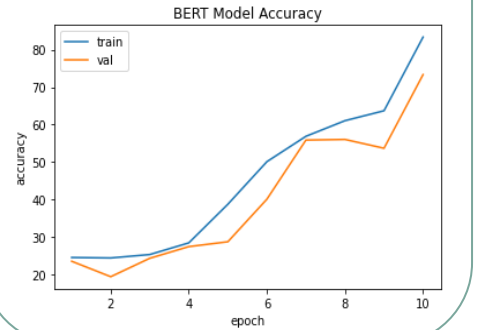
## Models and Results

| Model | Accuracy |
|---|---|
| Simple RNN | 97.60% |
| LSTM | 97.80% |
| BERT | 83% |

### Accuracy plots




### Simple RNN

| | | | |
|---|---|---|---|
| input_1 InputLayer | input: output: | [(None, 800)] | [(None, 800)] |
| embedding Embedding | input: output: | (None, 800) | (None, 800, 30) |
| simple_rnn SimpleRNN | input: output: | (None, 800, 30) | (None, 800, 64) |
| simple_rnn_1 SimpleRNN | input: output: | (None, 800, 64) | (None, 800, 64) |
| dense Dense | input: output: | (None, 800, 64) | (None, 800, 63) |

### LSTM

| | | | |
|---|---|---|---|
| input_2 InputLayer | input: output: | [(None, 800)] | [(None, 800)] |
| embedding Embedding | input: output: | (None, 800) | (None, 800, 100) |
| lstm LSTM | input: output: | (None, 800, 100) | (None, 800, 100) |
| lstm_1 LSTM | input: output: | (None, 800, 100) | (None, 800, 100) |
| dense_1 Dense | input: output: | (None, 800, 100) | (None, 800, 63) |

## Conclusions

- We were able to extract the entities from a sentence and name them with an accuracy as high as 98 %.
- The best model for the NER with the highest accuracy and accurate predictions turned out to be LSTM.

## Future Work

- Use transformer with encoder and decoder to create an NER.
- Use transformer with encoder and decoder to create an NER.
- Also perform Hyper Parameter tuning for all the different Models.
- Work on a larger Dataset and train our models based it.