



NLP Term Project



Named Entity Recognition
in Biomedical Corpus

Team Name - Vindicators

Here are the main tasks conducted in the study:

1. Detecting **Named Entities** using both context-independent word embeddings and context-dependent word embeddings. Domain specific and non-domain specific embeddings.
2. Used **NCBI-disease, BC5CDR-chem, BC5CDR-disease** datasets in the study.
3. **BI-LSTM-CRF** models for the Named Entity Recognition task.
4. Used deep contextual embeddings using the following architectures:
 - a. BERT (bio-BERT)
 - b. ELMO (bio-ELMO)
 - c. FLAIR (bioFLAIR)
5. Comparison between context-independent word embeddings and context-dependent word embeddings.

Example of NER Tagging

1

○	○	○	○	○	○	○	○	○
○	○	B-GENE	○	○	○	○	○	○
○	○	B-GENE	○	○	○	○	○	○
Expression	of	FNCAT	increased	on	serum	treatment	indicating	that

Context : ... the region of the FN gene between positions +69 and -510 bp mediated serum responsiveness.

2

○	○	○	○	○	○	○	B-GENE	I-GENE	○
○	B-GENE	I-GENE	○	○	○	○	B-GENE	I-GENE	○
○	B-GENE	I-GENE	○	○	○	○	B-GENE	I-GENE	○
whereas	dam	mutants	are	locked	off	for	Ag	43	expression

TABLE OF CONTENTS

01

Bio-ELMO

03

Bio-FLAIR

02

Bio-BERT

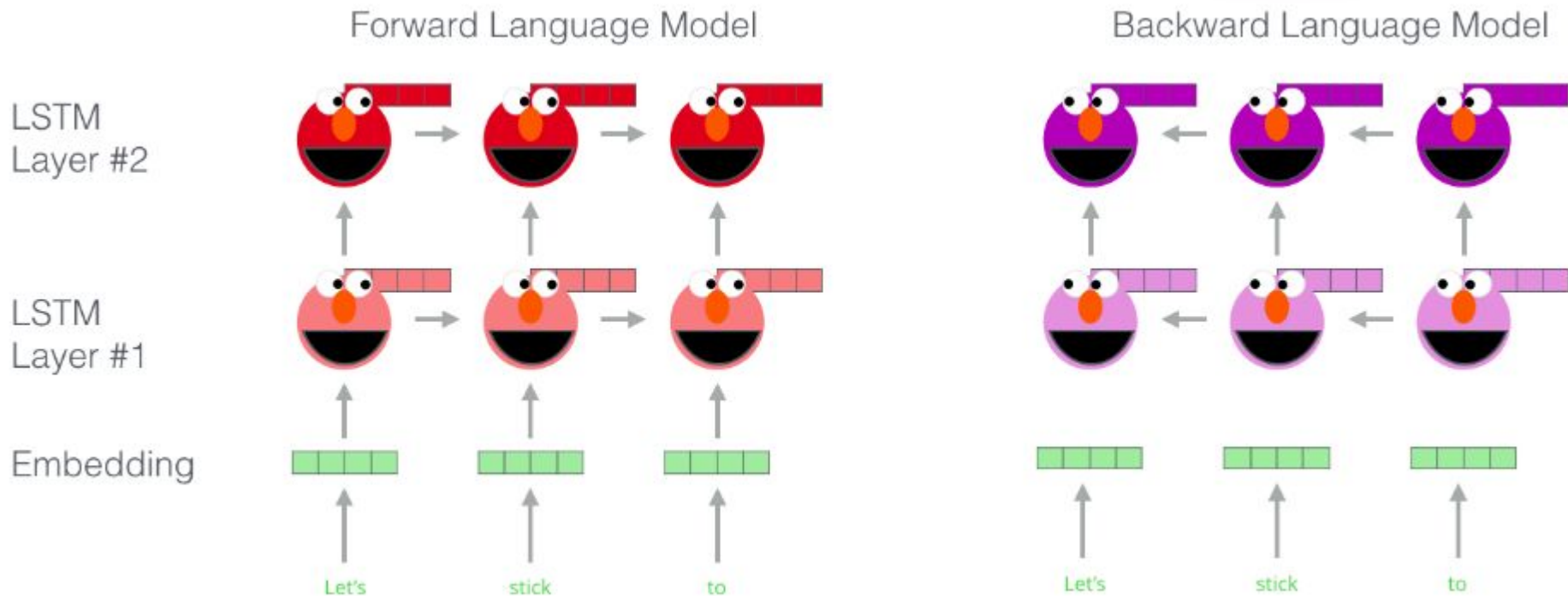
04

Conclusion

01

Bio-ELMO

01



ELMO RESULTS

F1 Score

0.74

0.78

0.70

Precision

0.74

0.83

0.76

Recall

0.73

0.74

0.65

NCBI disease

BC5SDR chem

BC5SDR disease

Bio-ELMO RESULTS

F1 Score

0.87

0.93

0.85

Precision

0.86

0.92

0.86

Recall

0.88

0.94

0.83

NCBI disease

BC5SDR chem

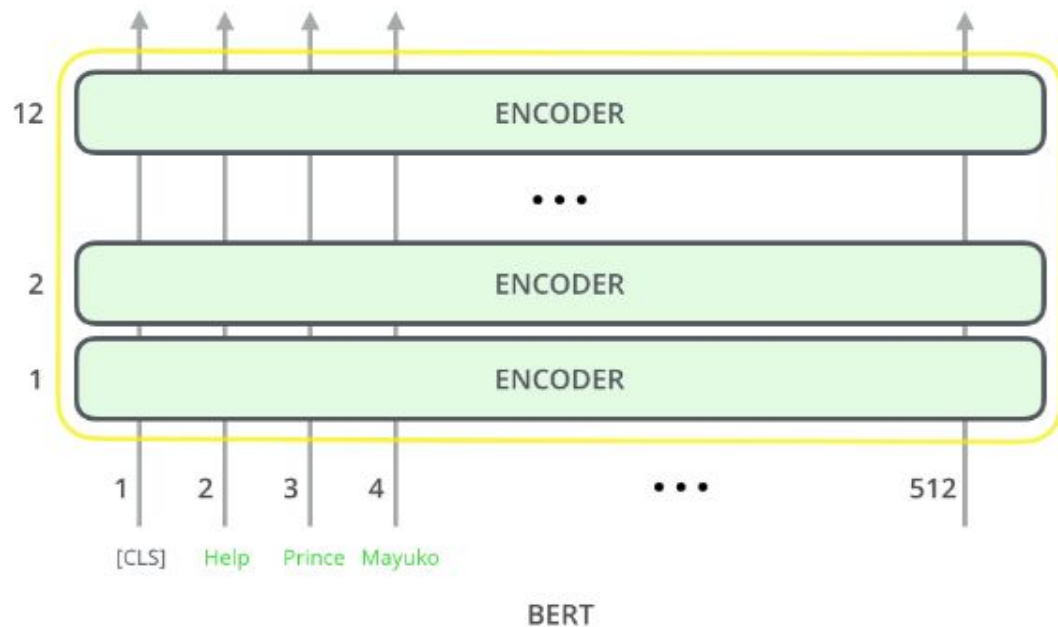
BC5SDR disease

Example of NER Tagging

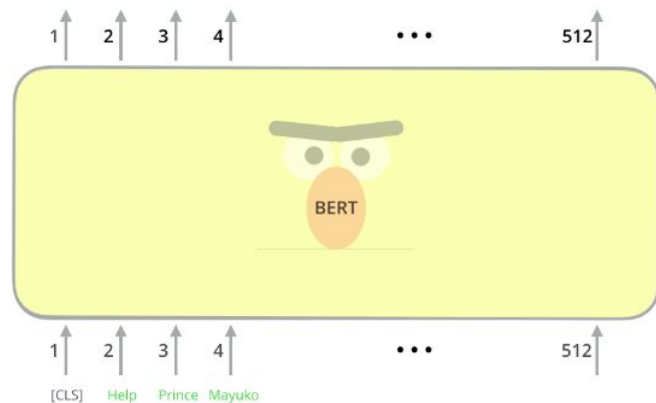
Clustering[O] of[O] missense[O] mutations[O]
in[O] the[O] ataxia[B-Disease] -[I-Disease]
telangiectasia[I-Disease] gene[O] in[O] [O]
sporadic[O] T[B-Disease] -[I-Disease]
cell[I-Disease] leukaemia [I-Disease] .[O]

02

Bio-BERT



02



BERT RESULTS

F1 Score

0.72

0.77

0.84

Precision

0.70

0.82

0.85

Recall

0.75

0.73

0.84

NCBI disease

BC5SDR chem

BC5SDR disease

Bio-BERT RESULTS

F1 Score

0.87

0.92

0.84

Precision

0.85

0.91

0.85

Recall

0.90

0.93

0.84

NCBI disease

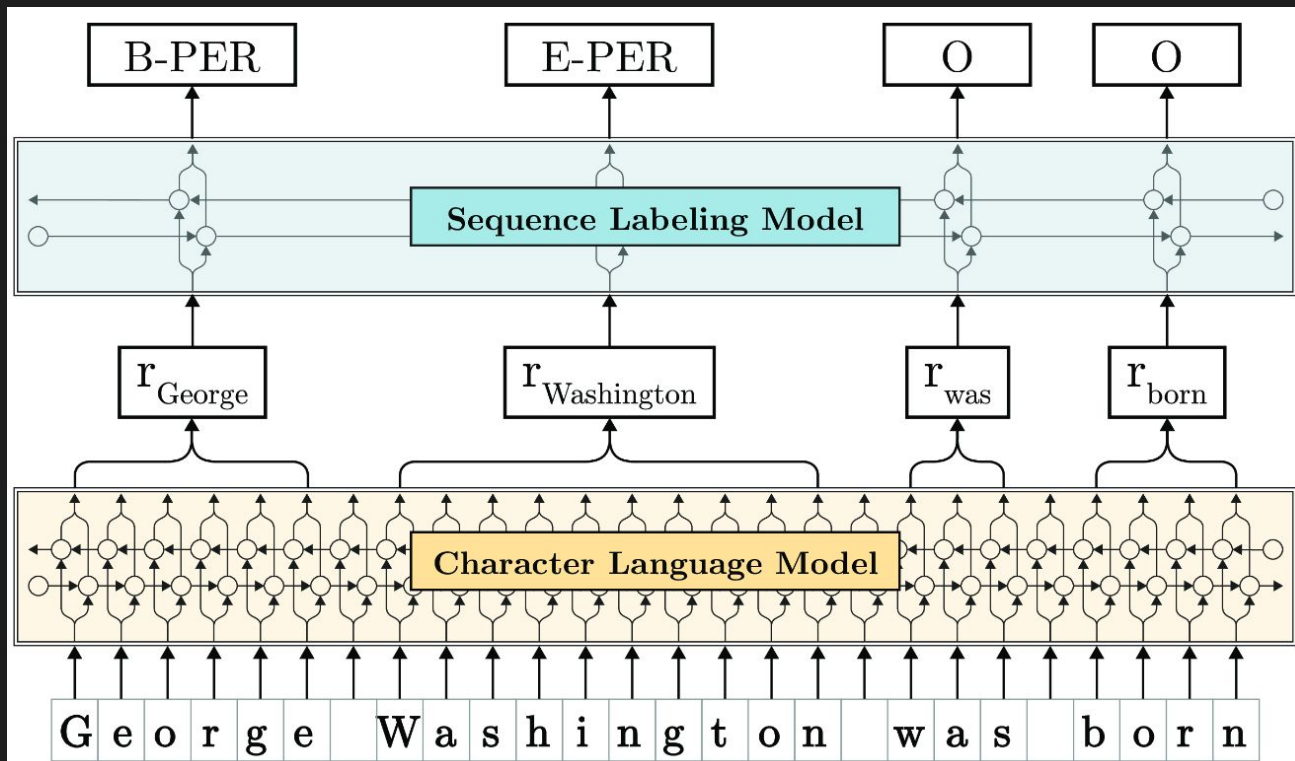
BC5SDR chem

BC5SDR disease

03

Bio-FLAIR

03



FLAIR RESULTS

F1 Score

0.85

0.83

0.72

Precision

0.85

0.82

0.81

Recall

0.84

0.85

0.64

NCBI disease

BC5SDR chem

BC5SDR disease

Bio-FLAIR RESULTS

F1 Score

0.86

0.90

0.82

Precision

0.85

0.89

0.84

Recall

0.88

0.91

0.80

NCBI disease

BC5SDR chem

BC5SDR disease



THANKS!

Do you have any
questions?

Code link : <https://github.com/iliox/NLP-term-project>