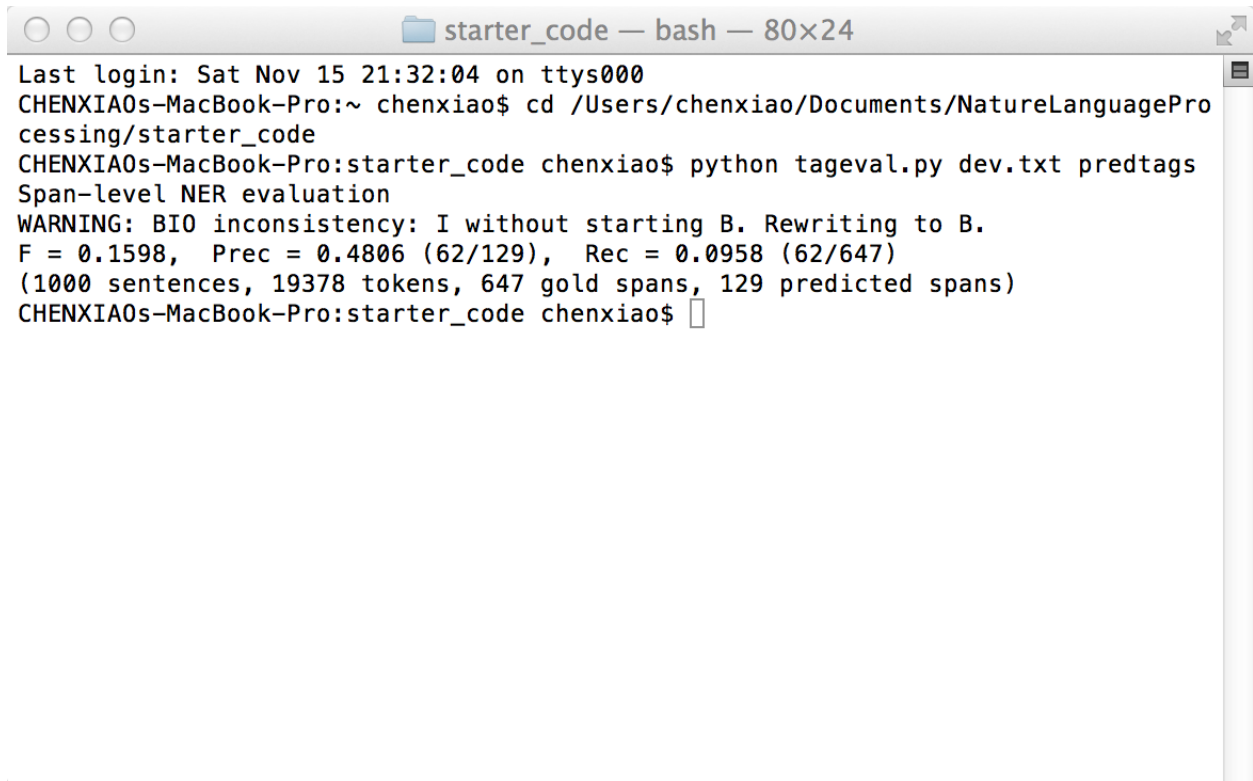


NLP Milestone

1.

A terminal window titled 'starter_code — bash — 80x24' showing the execution of a Python script for Named Entity Recognition (NER) evaluation. The output indicates a span-level evaluation with a low F-score due to BIOES-style inconsistencies.

```
Last login: Sat Nov 15 21:32:04 on ttys000
CHENXIAOs-MacBook-Pro:~ chenxiao$ cd /Users/chenxiao/Documents/NatureLanguagePro
cessing/starter_code
CHENXIAOs-MacBook-Pro:starter_code chenxiao$ python tageval.py dev.txt predtags
Span-level NER evaluation
WARNING: BIO inconsistency: I without starting B. Rewriting to B.
F = 0.1598, Prec = 0.4806 (62/129), Rec = 0.0958 (62/647)
(1000 sentences, 19378 tokens, 647 gold spans, 129 predicted spans)
CHENXIAOs-MacBook-Pro:starter_code chenxiao$
```

We made some simple changes to start code and we have our F score to be 0.16

Precision is 62/129, which means we have made 129 name predictions and 62 of them are correct, Recall is only 62/647, which means there are totally 647 names in gold standard and we only found 62 of them and made the prediction, this number is the main reason that we have such small F score and we then will focus on how to improve this.

2. After taking a look at tag predictions and gold standard, I noticed that we get good predictions for the words or phrase that are totally the same. This is reasonable considering that our codes right now only take the words around one certain word to be its attributes, and we assign them weights regarding the distance. But by doing this way it is hard to predict for unseen word in a new file. So we need to add new attributes to analyze structure of the sentences.

3. In model, we have FILEHEADER, LABELS, ATTRIBUTES, TRANSITIONS, STATE_FEATURES

example of high weights:

word[0]=M --> B: 0.687369
 word[0]=the --> O: 1.024825
 __BOS__ --> I: 0.614590
 word[1]='s --> B: 0.746043

the problem is just I mentioned above, we only take word and relative position into consideration and this makes prediction bad

4.

Two example of False Positive:

0	Buck	B	
0	mountain		I
0	tomorrow		0
0	.	0	
B	#adironacks		0
I	#hiking	0	

0	New	0
0	Comic	0
0	List	0
0	is	0
0	out	0
0	for	0
B	Next	0
I	Week	0

Example of False Negative:

0	coast	0	
0	then	0	
0	a	0	
0	stop	0	
0	in	0	
0	Glenveagh	0	B
0	national	0	I
0	park	I	
0	,	0	
0		0	

For this I was considering to add an attributes to check if the word starts with a capital letter