Report on HW 4 CRF assignment (Kaladhar Reddy Mummadi: 7761016469)

Implementation of Extraction.py File

> - First step I did was cleaning the train-ucla.txt to get chunks and their corresponding sentences in each line.
> - Then using spacy extracted properties of each word like `pos` (Parts of Speech), `lemma`.. details below.
> - Using crfutils.py file created featuers and labels in crfsuite format.
> - Installed python-crfsuite to perform a crf tagging using these features.
> - Trained four different types of models with different set of features, following are the observations and explanations and further improvements.

Feature List:

- **w** : current word
- **pos** : Parts Of Speech of each Word
  - Example : Letter : Noun
- **lemma** : Lemma of a word
  - Example : buying : buy
- **stop** (True/False) : Is stop word or not
  - Example : is : True
- **alpha** (True/False): is Alpha
- **shape** (Xxxx,xxx): shape of the word (This feature can be helpful in Course numbers)
- **tag** : if a word in cardnial Number
- **dep**: word dependency

> Apart from this i've also use next 1-3 words and previous 1-3 words which builds the context of a sentence.

Observations:

> Following values are F1-Scores on test-ucla.txt

Chunk Label	Model-1-Test	Model-2-Test	Model-3_Test	Model-4_Test
format	0.9523809523809523	0.99009900990099	1.0	1.0
others	0.35714285714285715	0.3181818181818182	0.3218390804597701	0.3218390804597701
description	0.10852713178294573	0.21238938053097345	0.23214285714285715	0.3218390804597701
grading	0.86	0.8775510204081631	0.8775510204081631	0.8775510204081631
requisite	0.08080808080808081	0.2916666666666667	0.3092783505154639	0.3092783505154639

- Model-1

```
  1   Used only current word as the feature, which includes 1-3 prev/next
words.
  2   As we can see it performed poorly on on `requisite` chunk
```

- Model-2

```
1   Used `word`, `POS`, `lemma` as features
2   There are improvement for `description`
```

- Model-3

```
1   Used `word`, `pos`, `lemma`, `stop`,`alpha` as features
2   we see slight improvements bbut not many
```

- Model-4

```
1   Used `w`, `pos`, `lemma`, `stop`,`alpha`, `shape`, `tag`, `dep`
2   This is the highest F-1 Score i've got and i've also used 4 prev/next
words
```

Following are the Detailed Values

Model-1 : Features Used w

Chunk Label	Scores	On Test Set	On Train Set
format	True Positives	50	158
	False Positives	5	0
	False Negatives	0	0
	Precision	0.9090909090909091	1.0
	Recall	1.0	1.0
	F1-Score	0.9523809523809523	1.0
others	True Positives	15	104
	False Positives	39	54
	False Negatives	15	48
	Precision	0.2777777777777778	0.6582278481012658
	Recall	0.5	0.6842105263157895
	F1-Score	0.35714285714285715	0.6709677419354839
description	True Positives	7	66
	False Positives	76	104
	False Negatives	39	89

Chunk Label	Scores	On Test Set	On Train Set
	Precision	0.08433734939759036	0.38823529411764707
	Recall	0.15217391304347827	0.4258064516129032
	F1-Score	0.10852713178294573	0.4061538461538462
grading	True Positives	43	136
	False Positives	8	22
	False Negatives	6	22
	Precision	0.8431372549019608	0.8607594936708861
	Recall	0.8775510204081632	0.8607594936708861
	F1-Score	0.86	0.8607594936708861
requisite	True Positives	4	105
	False Positives	60	19
	False Negatives	31	3
	Precision	0.0625	0.8467741935483871
	Recall	0.11428571428571428	0.9722222222222222
	F1-Score	0.08080808080808081	0.9051724137931034

Model-2

Features used `w`, `pos`, `lemma`

Chunk Label	Scores	On Test Set	On Train Set
format	True Positives	50	158
	False Positives	1	0
	False Negatives	0	0
	Precision	0.9803921568627451	1.0
	Recall	1.0	1.0
	F1-Score	0.99009900990099	1.0
others	True Positives	14	102
	False Positives	44	56
	False Negatives	16	50
	Precision	0.2413793103448276	0.6455696202531646
	Recall	0.4666666666666667	0.6710526315789473
	F1-Score	0.3181818181818182	0.6580645161290323
description	True Positives	12	66

Chunk Label	Scores	On Test Set	On Train Set
	False Positives	55	104
	False Negatives	34	89
	Precision	0.1791044776119403	0.38823529411764707
	Recall	0.2608695652173913	0.4258064516129032
	F1-Score	0.21238938053097345	0.4061538461538462
grading	True Positives	43	136
	False Positives	6	22
	False Negatives	6	22
	Precision	0.8775510204081632	0.8607594936708861
	Recall	0.8775510204081632	0.8607594936708861
	F1-Score	0.8775510204081631	0.8607594936708861
requisite	True Positives	14	105
	False Positives	47	19
	False Negatives	21	3
	Precision	0.22950819672131148	0.8467741935483871
	Recall	0.4	0.9722222222222222
	F1-Score	0.2916666666666667	0.9051724137931034

Features Used `w`, `pos`, `lemma`, `stop`,`alpha`

Chunk Label	Scores	On Test Set	On Train Set
format	True Positives	50	158
	False Positives	0	0
	False Negatives	0	0
	Precision	1.0	1.0
	Recall	1.0	1.0
	F1-Score	1.0	1.0
others	True Positives	14	102
	False Positives	43	54
	False Negatives	16	50
	Precision	0.24561403508771928	0.6538461538461539
	Recall	0.4666666666666667	0.6710526315789473
	F1-Score	0.3218390804597701	0.6623376623376623

Chunk Label	Scores	On Test Set	On Train Set
description	True Positives	13	66
	False Positives	53	102
	False Negatives	33	89
	Precision	0.19696969696969696	0.39285714285714285
	Recall	0.2826086956521739	0.4258064516129032
	F1-Score	0.23214285714285715	0.4086687306501548
grading	True Positives	43	136
	False Positives	6	22
	False Negatives	6	22
	Precision	0.8775510204081632	0.8607594936708861
	Recall	0.8775510204081632	0.8607594936708861
	F1-Score	0.8775510204081631	0.8607594936708861
requisite	True Positives	15	105
	False Positives	47	19
	False Negatives	20	3
	Precision	0.24193548387096775	0.8467741935483871
	Recall	0.42857142857142855	0.9722222222222222
	F1-Score	0.3092783505154639	0.9051724137931034

Features Used `w`, `pos`, `lemma`, `stop`,`alpha`, `shape`, `tag`, `dep`

Chunk Label	Scores	On Test Set	On Train Set
format	True Positives	50	158
	False Positives	0	0
	False Negatives	0	0
	Precision	1.0	1.0
	Recall	1.0	1.0
	F1-Score	1.0	1.0
others	True Positives	14	102
	False Positives	43	54
	False Negatives	16	50
	Precision	0.24561403508771928	0.6538461538461539
	Recall	0.4666666666666667	0.6710526315789473

Chunk Label	Scores	On Test Set	On Train Set
	F1-Score	0.3218390804597701	0.6623376623376623
description	True Positives	13	66
	False Positives	53	102
	False Negatives	33	89
	Precision	0.19696969696969696	0.39285714285714285
	Recall	0.2826086956521739	0.4258064516129032
	F1-Score	0.23214285714285715	0.4086687306501548
grading	True Positives	43	136
	False Positives	6	22
	False Negatives	6	22
	Precision	0.8775510204081632	0.8607594936708861
	Recall	0.8775510204081632	0.8607594936708861
	F1-Score	0.8775510204081631	0.8607594936708861
requisite	True Positives	15	105
	False Positives	47	19
	False Negatives	20	3
	Precision	0.24193548387096775	0.8467741935483871
	Recall	0.42857142857142855	0.9722222222222222
	F1-Score	0.3092783505154639	0.9051724137931034