

Feature-sets that has been tried:

version1: pos-tag

48943 out of 51578 tags correct

accuracy: 94.89

5917 groups in key

6458 groups in response

4118 correct groups

precision: 63.77

recall: 69.60

F1: 66.55

version2: pos-tag, current-token

48947 out of 51578 tags correct

accuracy: 94.90

5917 groups in key

6464 groups in response

4128 correct groups

precision: 63.86

recall: 69.77

F1: 66.68

Version3: pos-tag, current-token, previous token

48873 out of 51578 tags correct

accuracy: 94.76

5917 groups in key

6566 groups in response

4064 correct groups

precision: 61.89

recall: 68.68

F1: 65.11

Version4: pos-tag, current-token, next-token

45294 out of 51578 tags correct

accuracy: 87.82

5917 groups in key

6340 groups in response

3548 correct groups

precision: 55.96

recall: 59.96

F1: 57.89

Version5: pos-tag, previous token, current-token, next-token

49662 out of 51578 tags correct

accuracy: 96.29

5917 groups in key

6111 groups in response

4565 correct groups

precision: 74.70

recall: 77.15

F1: 75.91

Version 6: pos-tag, previous token, current-token, next-token, tag-type

49426 out of 51578 tags correct

accuracy: 95.83

5917 groups in key

5623 groups in response

4347 correct groups

precision: 77.31

recall: 73.47

F1: 75.34

Version 7: pos-tag, previous token, current-token, next-token, bio-chunk

accuracy: 96.22

5917 groups in key

6114 groups in response

4552 correct groups

precision: 74.45

recall: 76.93

F1: 75.67

Version 8: pos-tag, previous token, current-token, next-token, chunk

49628 out of 51578 tags correct

accuracy: 96.22

5917 groups in key

6117 groups in response

4553 correct groups

precision: 74.43
recall: 76.95
F1: 75.67

Version 9: pos, previous token, current-token, next-token, previous-tag, next-tag

48822 out of 51578 tags correct
accuracy: 94.66
5917 groups in key
5526 groups in response
4206 correct groups
precision: 76.11
recall: 71.08
F1: 73.51

Version 10: pos, previous token, current-token, next-token, previous-tag, chunk

49601 out of 51578 tags correct
accuracy: 96.17
5917 groups in key
5633 groups in response
4538 correct groups
precision: 80.56
recall: 76.69
F1: 78.58

Version 11: pos, previous token, current-token, previous-tag, chunk

49236 out of 51578 tags correct
accuracy: 95.46
5917 groups in key
5747 groups in response
4357 correct groups
precision: 75.81
recall: 73.64
F1: 74.71

Version 12: pos, previous token, current-token, next-token, previous-tag, tag-type

dev.name dev-output.txt
48558 out of 51578 tags correct

accuracy: 94.14
5917 groups in key
4383 groups in response
3776 correct groups
precision: 86.15
recall: 63.82
F1: 73.32

Version 13: pos, previous token, current-token, next-token, previous-tag

49622 out of 51578 tags correct

accuracy: 96.21
5917 groups in key
5655 groups in response
4568 correct groups
precision: 80.78
recall: 77.20
F1: 78.95

Instructions to run the code:

- All files must be in the same directory.
- Cd to that directory
- Run the following commands sequentially:
 - javac -cp .:maxent-3.0.0.jar:trove.jar MEtrain.java
 - javac -cp .:maxent-3.0.0.jar:trove.jar MEdtag.java
 - Python3 v10-p1.py
 - java -cp .:maxent-3.0.0.jar:trove.jar MEtrain train-input.txt train-output.txt
 - Python3 v10-p2.py
 - java -cp .:maxent-3.0.0.jar:trove.jar MEdtag dev-input.txt train-output.txt dev-output.txt
 - python3 score.name.py dev.name dev-output.txt
 - java -cp .:maxent-3.0.0.jar:trove.jar MEdtag test.pos-chunk train-output.txt test.name