

COMS 4705 - Natural Language Processing

Homework 2 (Programmatic)

Aayush Maini - am4810

Question 4

USAGE:	<code>python parser.py q4 parse_train.dat parse_train.RARE.dat</code> <code>./parser.py q4 parse_train.dat parse_train.RARE.dat</code> <code>./q4.py parse_train.dat parse_train.RARE.dat</code>
---------------	--

The counts file is not generated for q4 because it doesn't use the counts file at all. The counts file is generated as a temporary file for q5 and q6.

Question 5

USAGE:	<code>python parser.py q5 parse_train.RARE.dat parse_dev.dat q5_prediction_file</code> <code>./parser.py q5 parse_train.RARE.dat parse_dev.dat q5_prediction_file</code> <code>./q5.py parse_train.RARE.dat parse_dev.dat q5_prediction_file</code>
---------------	---

Execution time of q5 (unoptimized): 1 minute 42 seconds

```
14:56 $ time ./parser.py q5 parse_train.RARE.dat parse_dev.dat q5_prediction_file
real    1m42.203s
user    1m42.128s
sys     0m0.052s
```

Execution time of q5 (optimized): 16 seconds
--

```
14:47 $ time ./parser.py q5 parse_train.RARE.dat parse_dev.dat q5_prediction_file
real    0m16.842s
user    0m16.796s
sys     0m0.004s
```

Evaluation Results of q5				
Type	Total	Precision	Recall	F1 Score
.	370	1.000	1.000	1.000
ADJ	164	0.827	0.555	0.664
ADJP	29	0.333	0.241	0.280
ADJP+ADJ	22	0.542	0.591	0.565
ADP	204	0.955	0.946	0.951
ADV	64	0.694	0.531	0.602
ADVP	30	0.333	0.133	0.190
ADVP+ADV	53	0.756	0.642	0.694
CONJ	53	1.000	1.000	1.000
DET	167	0.988	0.976	0.982
NOUN	671	0.752	0.842	0.795
NP	884	0.632	0.529	0.576
NP+ADJ	2	0.286	1.000	0.444
NP+DET	21	0.783	0.857	0.818
NP+NOUN	131	0.641	0.573	0.605
NP+NUM	13	0.214	0.231	0.222
NP+PRON	50	0.980	0.980	0.980
NP+QP	11	0.667	0.182	0.286
NUM	93	0.984	0.645	0.779
PP	208	0.588	0.625	0.606
PRON	14	1.000	0.929	0.963
PRT	45	0.957	0.978	0.967
PRT+PRT	2	0.400	1.000	0.571
QP	26	0.647	0.423	0.512
S	587	0.626	0.782	0.695
SBAR	25	0.091	0.040	0.056
VERB	283	0.683	0.799	0.736
VP	399	0.559	0.594	0.576
VP+VERB	15	0.250	0.267	0.258
total	4664	0.714	0.714	0.714

Question 6

Execution time of q6 (unoptimized): 8 minutes 50 seconds

```
15:00 $ time ./parser.py q6 parse_train_vert.RARE.dat parse_dev.dat q6_prediction_file
real    8m50.545s
user    8m49.368s
sys     0m0.272s
```

Execution time of q6 (optimized): 28 seconds

```
14:47 $ time ./parser.py q6 parse_train_vert.RARE.dat parse_dev.dat q6_prediction_file
real    0m28.945s
user    0m28.748s
sys     0m0.060s
```

Evaluation results for q6

Type	Total	Precision	Recall	F1 Score
.	370	1.000	1.000	1.000
ADJ	164	0.689	0.622	0.654
ADJP	29	0.324	0.414	0.364
ADJP+ADJ	22	0.591	0.591	0.591
ADP	204	0.960	0.951	0.956
ADV	64	0.759	0.641	0.695
ADVP	30	0.417	0.167	0.238
ADVP+ADV	53	0.700	0.660	0.680
CONJ	53	1.000	1.000	1.000
DET	167	0.988	0.994	0.991
NOUN	671	0.795	0.845	0.819
NP	884	0.617	0.548	0.580
NP+ADJ	2	0.333	0.500	0.400
NP+DET	21	0.944	0.810	0.872
NP+NOUN	131	0.610	0.656	0.632
NP+NUM	13	0.375	0.231	0.286
NP+PRON	50	0.980	0.980	0.980
NP+QP	11	0.750	0.273	0.400
NUM	93	0.914	0.688	0.785
PP	208	0.623	0.635	0.629
PRON	14	1.000	0.929	0.963
PRT	45	1.000	0.933	0.966
PRT+PRT	2	0.286	1.000	0.444
QP	26	0.650	0.500	0.565
S	587	0.704	0.814	0.755
SBAR	25	0.667	0.400	0.500
VERB	283	0.790	0.813	0.801
VP	399	0.663	0.677	0.670
VP+VERB	15	0.294	0.333	0.312
total	4664	0.742	0.742	0.742

USAGE: python parser.py q6 parse_train_vert.dat parse_dev.dat q6_prediction_file
./parser.py q6 parse_train_vert.dat parse_dev.dat q6_prediction_file
./q6.py parse_train_vert.dat parse_dev.dat q6_prediction_file