Home Work-2
Abhinaya Krishna Mandepudi
Axm163231

Question1: python 3
(?:^|\s)(?!000)\d{3}-?\d{2}-?(?!0000)\d{4}(?:\s|$)

Question2: python 3

To run:

python Q2.py
Enter Testing Sentence. : TestingSentence

TestingSentence :

Question3: python 3

To run:

python Q3.py
 do you want to train Brills ? yes= 1 | no = 0: "enter 0 or 1"
Enter Testing Sentence with its Tags: TestingSentence

FinalRules.txt : Rules fetched are stored in this text file after 1 successful run.
Note : Enter 0 to use the rules created or 1 to create the rules again

TestingSentence Example: The_DT president_NN wants_VBZ to_TO control_VB the_DT board_NN 's_POS abhi_NN

Examples :

~ — jupyter_mac.command — python ◂ -bash ... ~/Desktop/Spring18/NLP/HW2/axm163231_HW2 — -bash

```
cometnet-10-21-27-225:axm163231_HW2 abhinay$ python Q2.py
Enter Testing Sentence.   :was named a director of abhi
Given Sentence :" was named a director of abhi "

Bigram Counts
              Method  (was, named)  (named, a)  (a, director)  (director, of)  \
0       No Smoothing     42.000000   10.000000      24.000000        20.00000
1  Add One Smoothing      1.301678    0.110719       2.351325         0.16354
2        Good Turing     43.000000    8.380952      25.000000        17.50000

   (of, abhi)
0    0.000000
1    0.143337
2    0.000482

Bigram Probabilities
              Method  (was, named)  (named, a)  (a, director)  (director, of)  \
0       No Smoothing      0.240000    0.175439       0.041237        0.454545
1  Add One Smoothing      0.007438    0.001942       0.004040        0.003717
2        Good Turing      0.245714    0.147034       0.042955        0.397727

     (of, abhi)
0  0.000000e+00
1  1.528117e-04
2  5.136491e-07

Total Probabilities
   No Smoothing  Add One Smoothing   Good Turing
0           0.0       3.315346e-14  3.170428e-10


Note: when the bigram counts are high, due to missing buckets probabilities of them can become zero
 as professor suggested not to modify basic Good Turing
cometnet-10-21-27-225:axm163231_HW2 abhinay$
```

~ — jupyter_mac.command — python ◂ -bash ... ~/Desktop/Spring18/NLP/HW2/axm163231_HW2 — -bash

```
cometnet-10-21-27-225:axm163231_HW2 abhinay$ python Q3.py
do you want to train Brills ? yes= 1 | no = 0:  0
using the rules generated earlier, which are stored in FinalRules.txt
Enter Testing Sentence with its Tags:
The_DT president_NN wants_VBZ to_TO control_VB the_DT board_NN 's_POS abhi_NN
Given Sentence :
 The_DT president_NN wants_VBZ to_TO control_VB the_DT board_NN 's_POS abhi_NN
Unigram
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'NN'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Brills
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Bigram
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Manual Tagging
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Total Words in the given Sentence: 9

Unigram Errors for given Sentence = 1  & Error Rate = 11.11 %

Brill's Errors for given Sentence = 0  & Error Rate = 0.0 %

Bigram's Errors for given Sentence = 0  & Error Rate = 0.0 %
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$
```

~ — jupyter_mac.command — python ◂ -bash ... ~/Desktop/Spring18/NLP/HW2/axm163231_HW2 — -bash

```
Total Words in the given Sentence: 9

Unigram Errors for given Sentence = 1  & Error Rate = 11.11 %

Brill's Errors for given Sentence = 0  & Error Rate = 0.0 %

Bigram's Errors for given Sentence = 0  & Error Rate = 0.0 %
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$
cometnet-10-21-27-225:axm163231_HW2 abhinay$ python Q3.py
do you want to train Brills ? yes= 1 | no = 0:  1
Training Selected
Started Fetching Rules
Time required to learn : 32.91 seconds
[(('VBP', 'VB', 'MD'), 53), (('NN', 'VB', 'MD'), 43), (('VBD', 'VBN', 'VBZ'), 40), (('VBN', 'VBD', 'PRP'), 35), (('VB', 'VBP', 'NNS'), 34), (('VBN', 'VBD', 'NNP'), 34), (('POS', 'VB
Z', 'PRP'), 31), (('VBP', 'VB', 'TO'), 29), (('VB', 'VBP', 'PRP'), 28), (('NN', 'VB', 'TO'), 23), (('VBD', 'VBN', 'VB'), 18), (('VB', 'NN', 'DT'), 16), (('VBD', 'VBN', 'VBP'), 15),
 (('VBD', 'VBN', 'VBD'), 15), (('VBZ', 'NNS', 'VBN'), 14), (('VB', 'NN', 'NN'), 13), (('VBN', 'VB', 'TO'), 12), (('VBN', 'VB', 'MD'), 12), (('VB', 'NN', 'IN'), 11), (('POS', 'VBZ', '
EX'), 10), (('VB', 'NN', 'JJ'), 9), (('RP', 'IN', 'NN'), 7), (('POS', 'VBZ', 'DT'), 7), (('VBN', 'VBD', 'WP'), 7), (('VBD', 'VBN', 'VBN'), 7), (('RP', 'RB', 'NN'), 6), (('VB', 'VBP
', 'NNP'), 6), (('VBD', 'VB', 'TO'), 6), (('RP', 'IN', 'RB'), 5), (('RP', 'RB', ','), 5), (('VB', 'VBP', 'IN'), 5), (('RP', 'IN', 'NNS'), 4), (('RP', 'RB', 'IN'), 4), (('POS', 'VBZ',
 'IN'), 4), (('EX', 'RB', 'NNS'), 4), (('EX', 'RB', 'NN'), 4), (('VB', 'VBN', 'VBP'), 4), (('VB', 'NN', 'PRP$'), 4), (('RB', 'IN', 'CD'), 3), (('JJR', 'RBR', 'NN'), 3), (('VBZ', 'NN
S', 'JJR'), 3), (('VB', 'VBP', 'NNPS'), 3), (('VB', 'VBD', 'NNP'), 3), (('VB', 'NN', 'CD'), 3), (('JJS', 'NN', 'NN'), 3), (('IN', 'DT', 'TO'), 2), (('NNP', 'JJ', 'WP$'), 2), (('JJR'
, 'RBR', 'VBG'), 2), (('JJR', 'RBR', 'VBZ'), 2), (('NNS', 'VBZ', 'PRP'), 2), (('RBS', 'JJS', '.'), 2), (('RBR', 'JJR', ':'), 2), (('RP', 'IN', 'CC'), 2), (('POS', 'VBZ', 'WP'), 2),
 (('VBZ', 'NNS', 'PRP$'), 2), (('EX', 'RB', 'VBN'), 2), (('NNPS', 'NNS', '.'), 2), (('VB', 'VBP', 'WP'), 2), (('VB', 'JJ', 'DT'), 2), (('VBD', 'VB', 'MD'), 2), (('VBD', 'JJ', 'PRP$')
, 2), (('IN', 'VB', 'MD'), 1), (('NNP', 'NN', 'WP$'), 1), (('JJR', 'RB', 'MD'), 1), (('JJR', 'RBR', 'MD'), 1), (('VBG', 'NN', '-RRB-'), 1), (('DT', 'RB', 'WP'), 1), (('RBR', 'RB', '
PRP'), 1), (('RBR', 'RB', 'MD'), 1), (('RBR', 'JJ', ','), 1), (('RP', 'RB', 'NNS'), 1), (('RP', 'RB', 'DT'), 1), (('RP', 'RB', 'RP'), 1), (('VBP', 'VB', 'VBZ'), 1), (('VBP', 'JJ', '
PRP$'), 1), (('VBZ', 'NNS', 'VBZ'), 1), (('VBZ', 'NNS', 'VBD'), 1), (('EX', 'RB', 'VBG'), 1), (('EX', 'RB', 'VB'), 1), (('NNPS', 'NNS', 'VBG'), 1), (('VB', 'VBP', 'DT'), 1), (('VB',
 'VBP', 'VBP'), 1), (('VB', 'VBP', 'JJ'), 1), (('VB', 'JJ', 'JJ'), 1), (('VB', 'NN', 'NNP'), 1), (('VB', 'NN', 'JJS'), 1), (('JJS', 'RBS', 'WP'), 1), (('JJS', 'RBS', 'MD'), 1), ((('V
BN', 'JJ', 'PRP$'), 1), (('VBN', 'VBD', 'WDT'), 1), (('VBN', 'VBD', 'NNPS'), 1), (('VBD', 'VBN', 'JJR'), 1), (('VBD', 'VBN', 'POS'), 1)]
Enter Testing Sentence with its Tags:
The_DT president_NN wants_VBZ to_TO control_VB the_DT board_NN 's_POS abhi_NN
Given Sentence :
 The_DT president_NN wants_VBZ to_TO control_VB the_DT board_NN 's_POS abhi_NN
Unigram
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'NN'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Brills
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Bigram
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Manual Tagging
[['The', 'DT'], ['president', 'NN'], ['wants', 'VBZ'], ['to', 'TO'], ['control', 'VB'], ['the', 'DT'], ['board', 'NN'], ["'s", 'POS'], ['abhi', 'NN']]

Total Words in the given Sentence: 9

Unigram Errors for given Sentence = 1  & Error Rate = 11.11 %

Brill's Errors for given Sentence = 0  & Error Rate = 0.0 %

Bigram's Errors for given Sentence = 0  & Error Rate = 0.0 %
cometnet-10-21-27-225:axm163231_HW2 abhinay$
```