QUESTION:

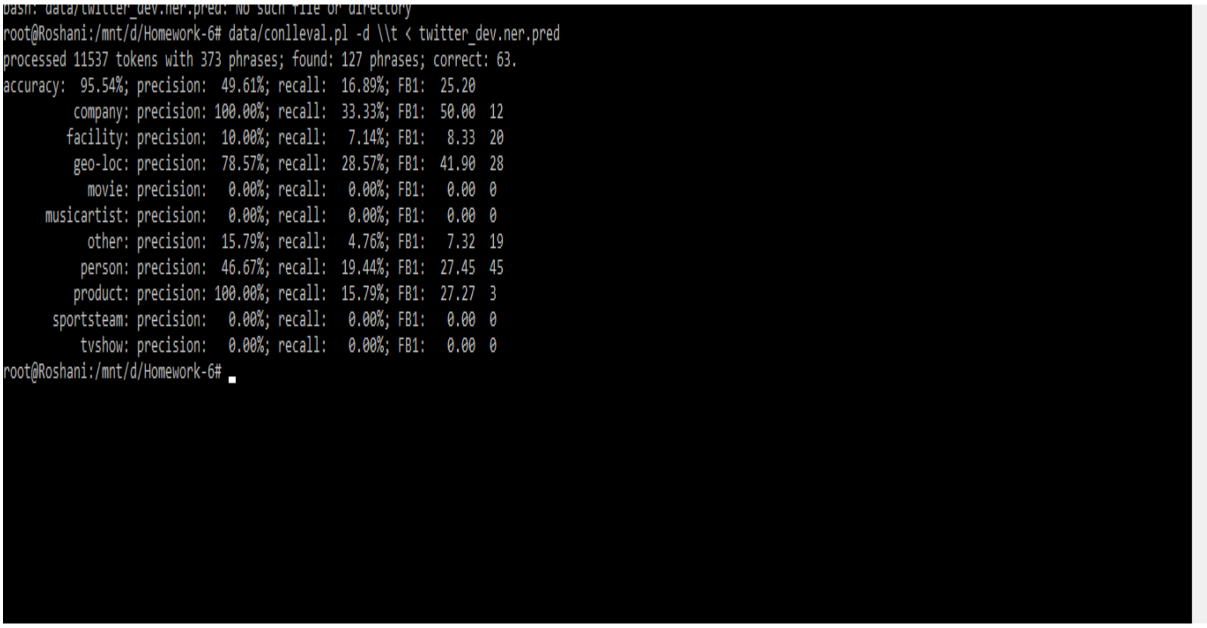
Please write up the following: in each case, which methods give the highest performance, and by how much? Briefly explain why. Use any graphs, tables, and figures to aid your analysis, including ones generated by conlleval.pl.

ANSWER:

With Basic Features only (only conlleval)

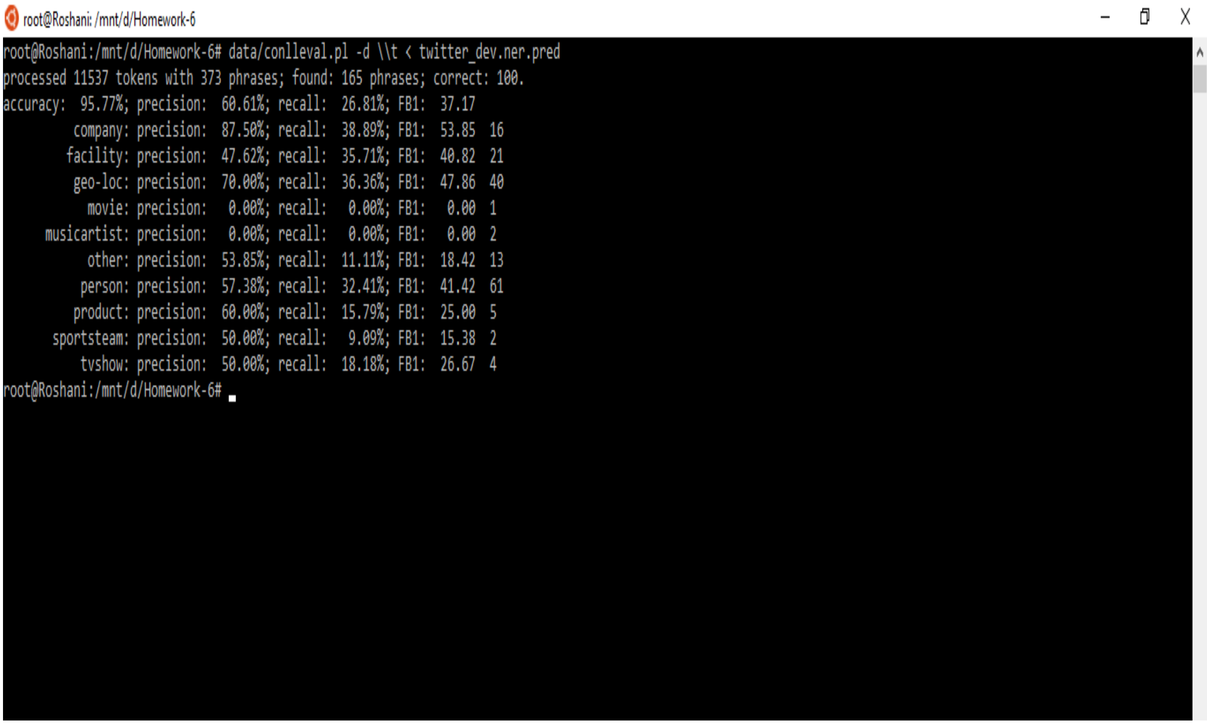
Twitter_dev.ner:

Log-reg (without features)



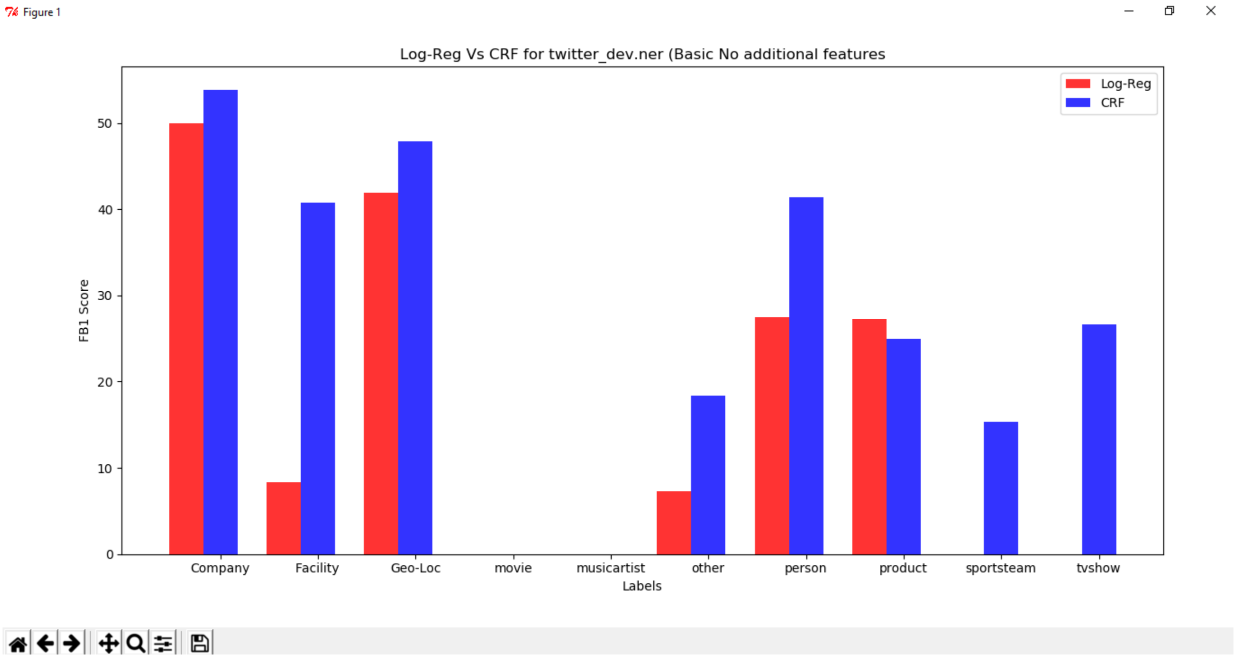
```
bash: data/twitter_dev.ner.pred: No such file or directory
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 127 phrases; correct: 63.
accuracy:  95.54%; precision:  49.61%; recall:  16.89%; FB1:  25.20
          company: precision: 100.00%; recall:  33.33%; FB1:  50.00  12
         facility: precision:  10.00%; recall:   7.14%; FB1:   8.33  20
          geo-loc: precision:  78.57%; recall:  28.57%; FB1:  41.90  28
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
            other: precision:  15.79%; recall:   4.76%; FB1:   7.32  19
           person: precision:  46.67%; recall:  19.44%; FB1:  27.45  45
          product: precision: 100.00%; recall:  15.79%; FB1:  27.27  3
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
root@Roshani:/mnt/d/Homework-6#
```

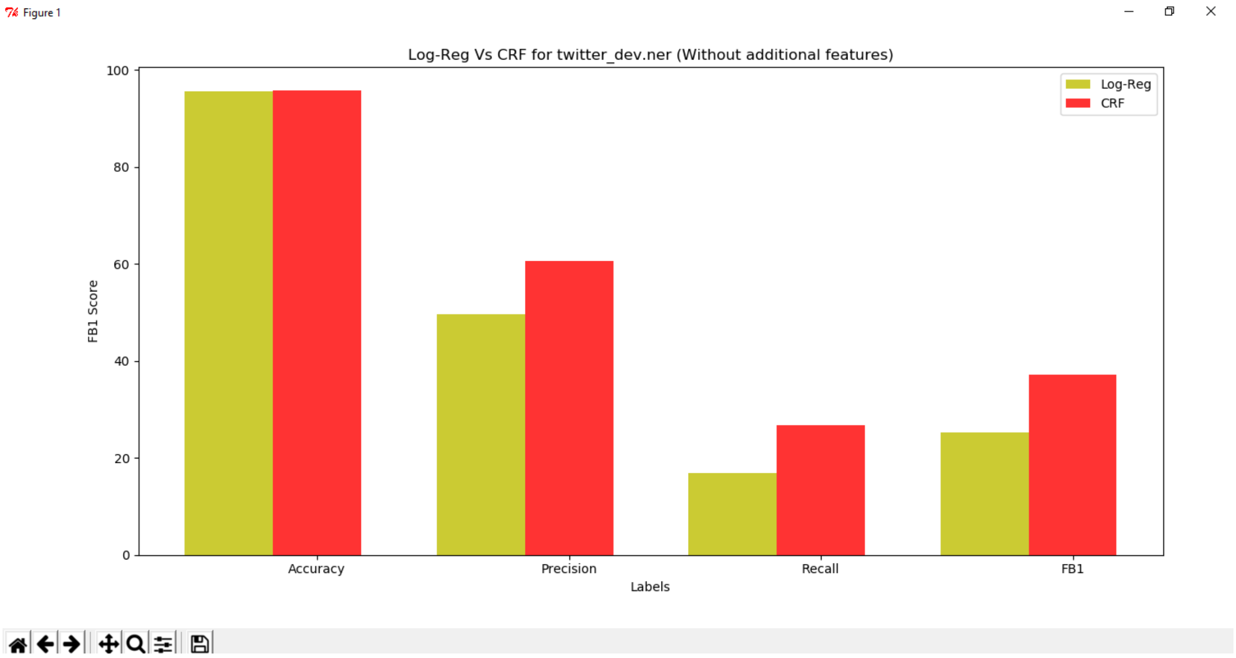
CRF: Conditional Random Field (without features):



```
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 165 phrases; correct: 100.
accuracy:  95.77%; precision:  60.61%; recall:  26.81%; FB1:  37.17
          company: precision:  87.50%; recall:  38.89%; FB1:  53.85  16
         facility: precision:  47.62%; recall:  35.71%; FB1:  40.82  21
          geo-loc: precision:  70.00%; recall:  36.36%; FB1:  47.86  40
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  2
            other: precision:  53.85%; recall:  11.11%; FB1:  18.42  13
           person: precision:  57.38%; recall:  32.41%; FB1:  41.42  61
          product: precision:  60.00%; recall:  15.79%; FB1:  25.00  5
       sportsteam: precision:  50.00%; recall:   9.09%; FB1:  15.38  2
           tvshow: precision:  50.00%; recall:  18.18%; FB1:  26.67  4
root@Roshani:/mnt/d/Homework-6#
```

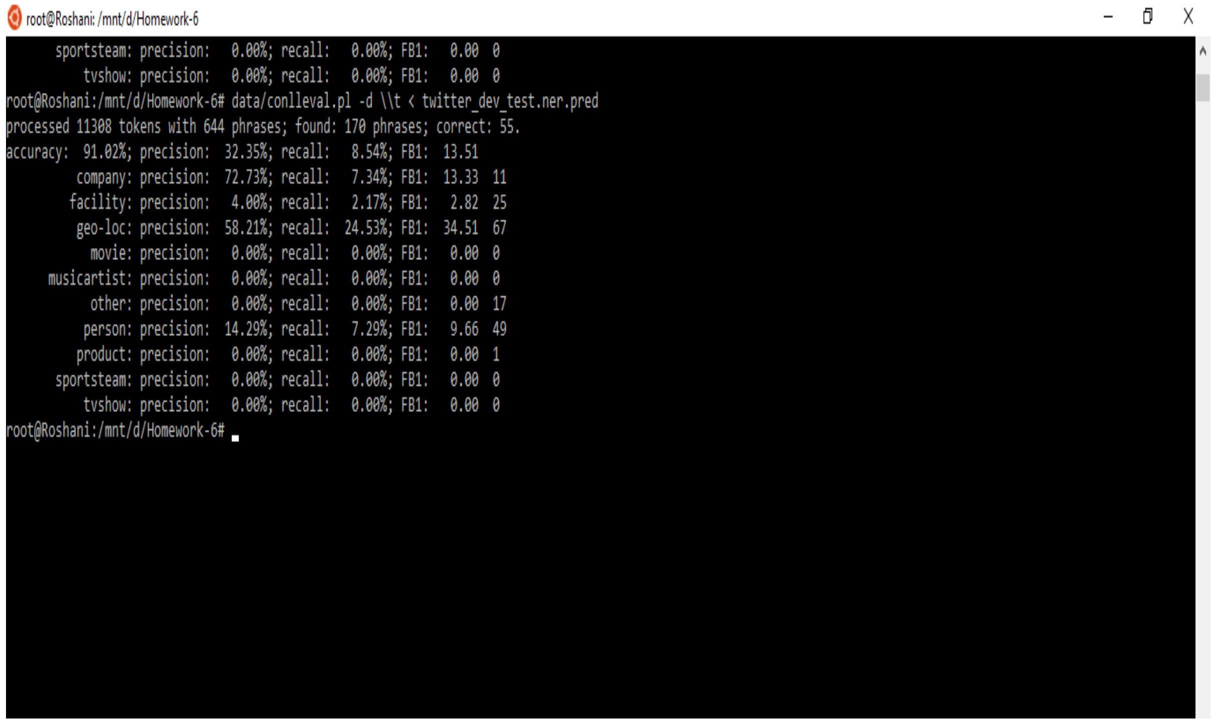
COMPARISION GRAPHS:





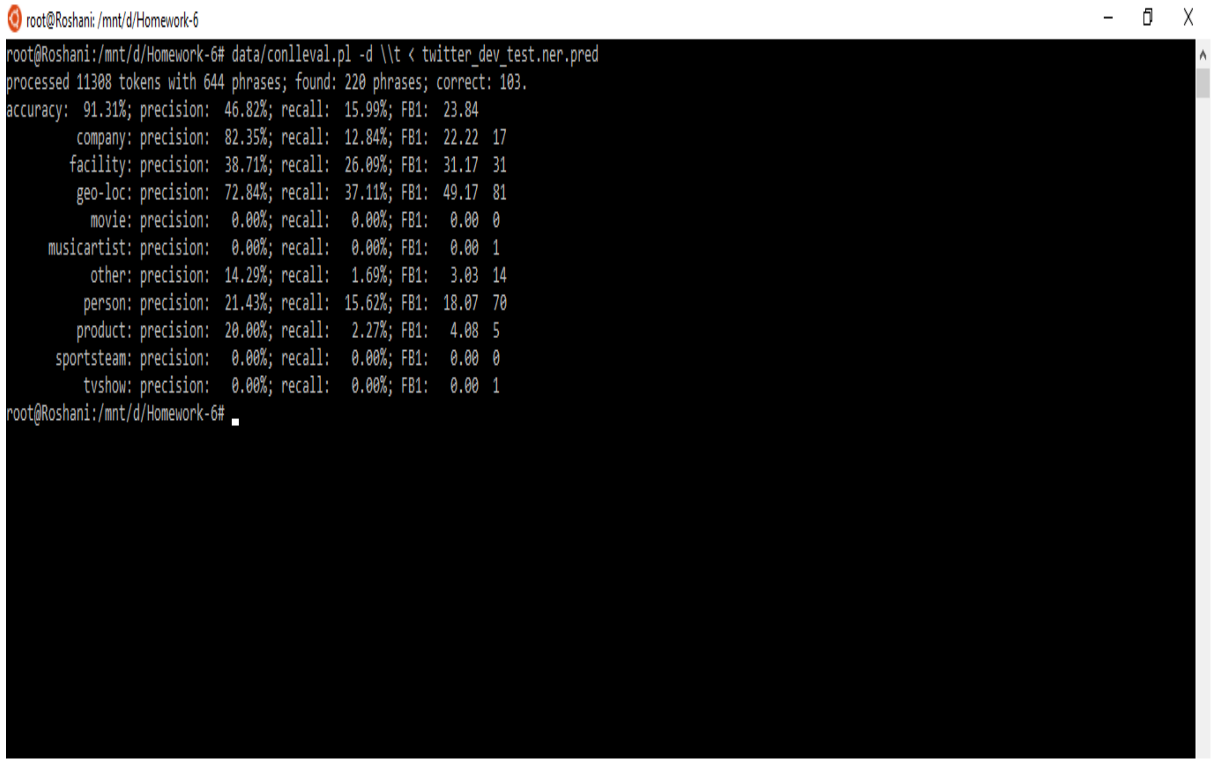
Twitter_dev_test.ner:

Log-reg: (without features)



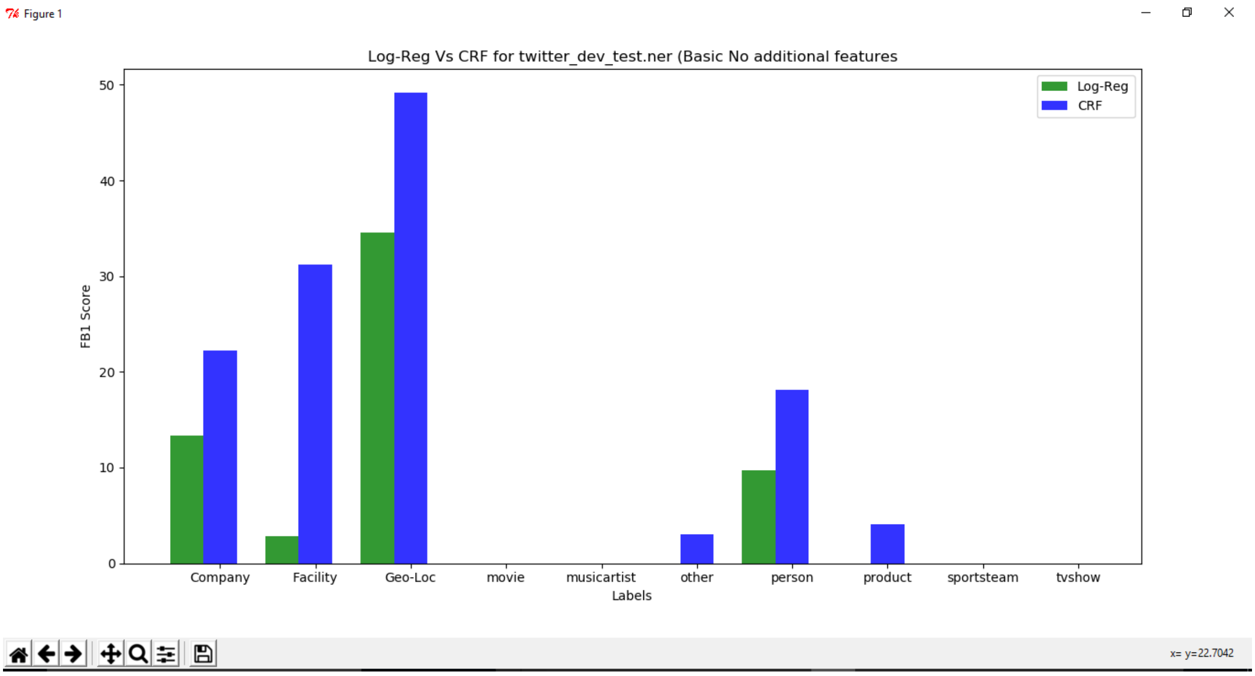
```
      sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
          tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 170 phrases; correct: 55.
accuracy:  91.02%; precision:  32.35%; recall:   8.54%; FB1:  13.51
         company: precision:  72.73%; recall:   7.34%; FB1:  13.33  11
        facility: precision:   4.00%; recall:   2.17%; FB1:   2.82  25
         geo-loc: precision:  58.21%; recall:  24.53%; FB1:  34.51  67
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
     musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           other: precision:   0.00%; recall:   0.00%; FB1:   0.00  17
          person: precision:  14.29%; recall:   7.29%; FB1:   9.66  49
         product: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
      sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
          tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
root@Roshani:/mnt/d/Homework-6#
```

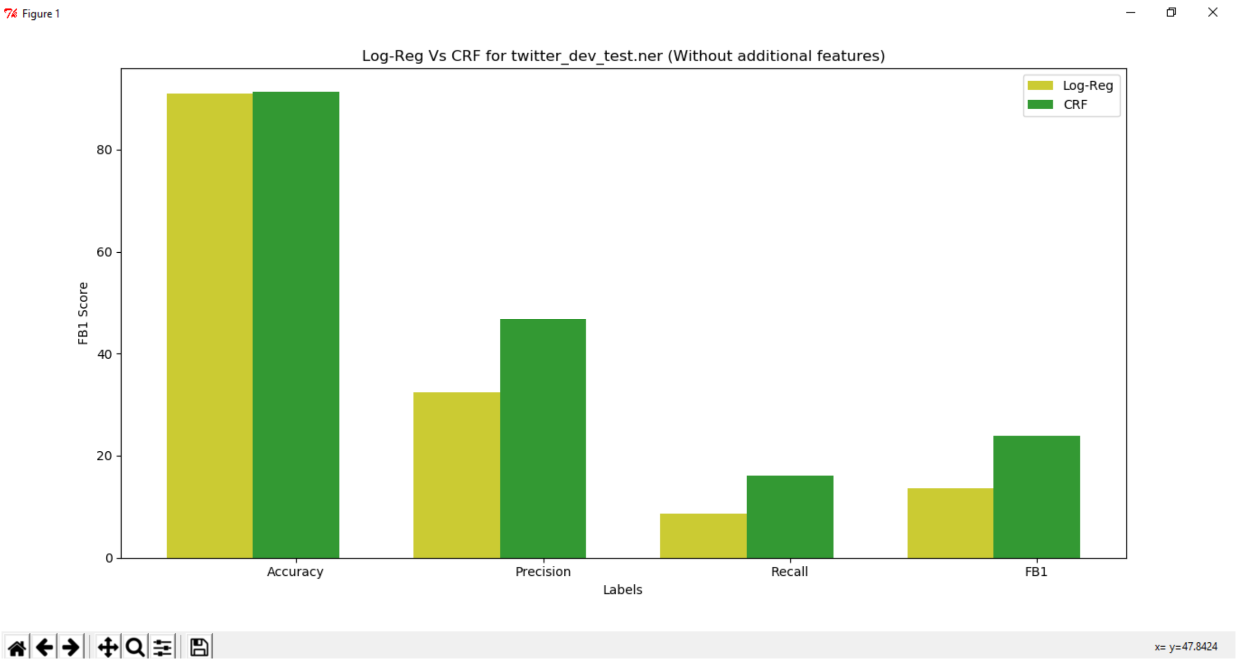
CRF: Conditional Random Field (without features):



```
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 220 phrases; correct: 103.
accuracy:  91.31%; precision:  46.82%; recall:  15.99%; FB1:  23.84
         company: precision:  82.35%; recall:  12.84%; FB1:  22.22  17
        facility: precision:  38.71%; recall:  26.09%; FB1:  31.17  31
         geo-loc: precision:  72.84%; recall:  37.11%; FB1:  49.17  81
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
     musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
           other: precision:  14.29%; recall:   1.69%; FB1:   3.03  14
          person: precision:  21.43%; recall:  15.62%; FB1:  18.07  70
         product: precision:  20.00%; recall:   2.27%; FB1:   4.08  5
      sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
          tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
root@Roshani:/mnt/d/Homework-6#
```

COMPARISION GRAPHS:

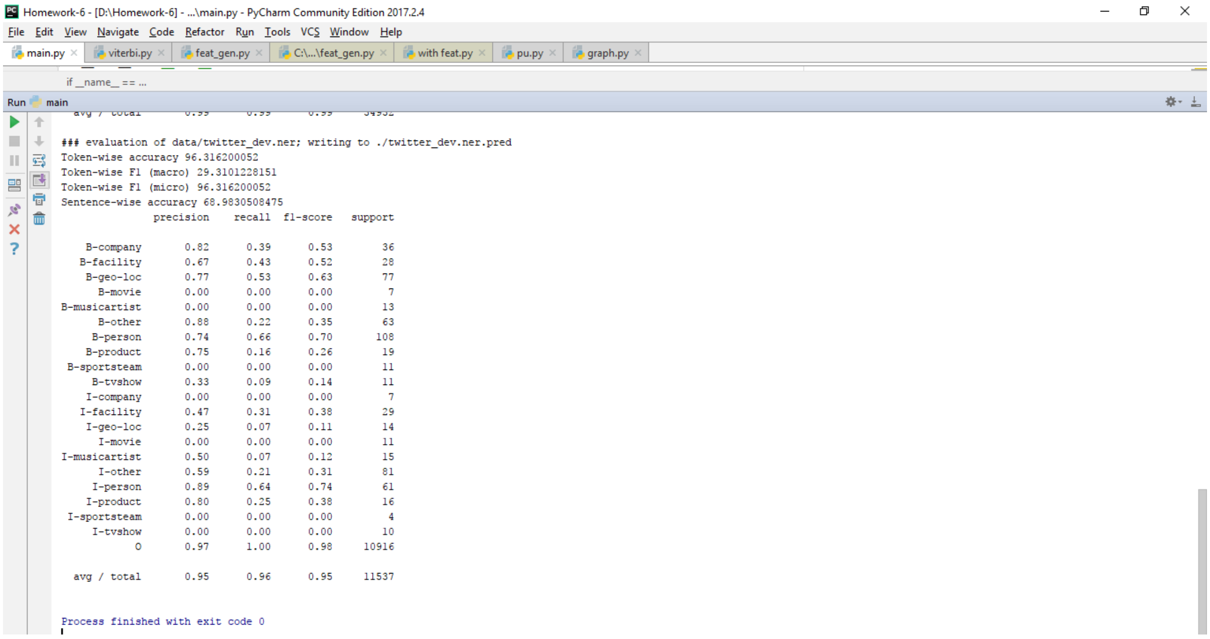




With Additional features: (both main.py and conlleval.pl and output with graph comparisons of log-reg and crf)

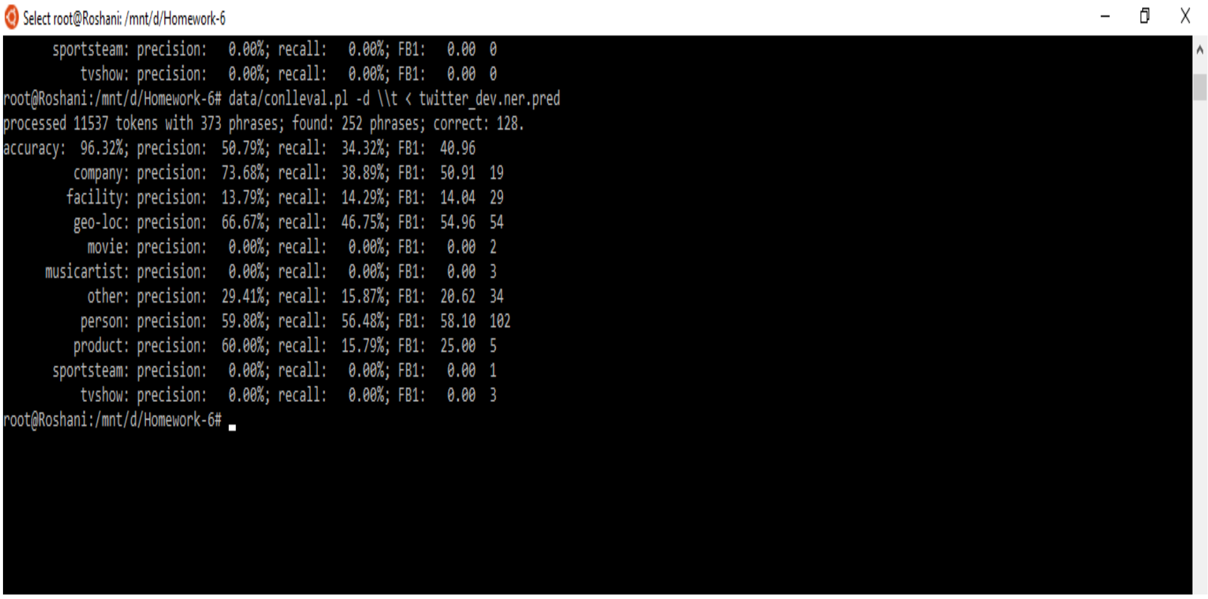
Twitter_dev.ner:

Log-reg (With features):



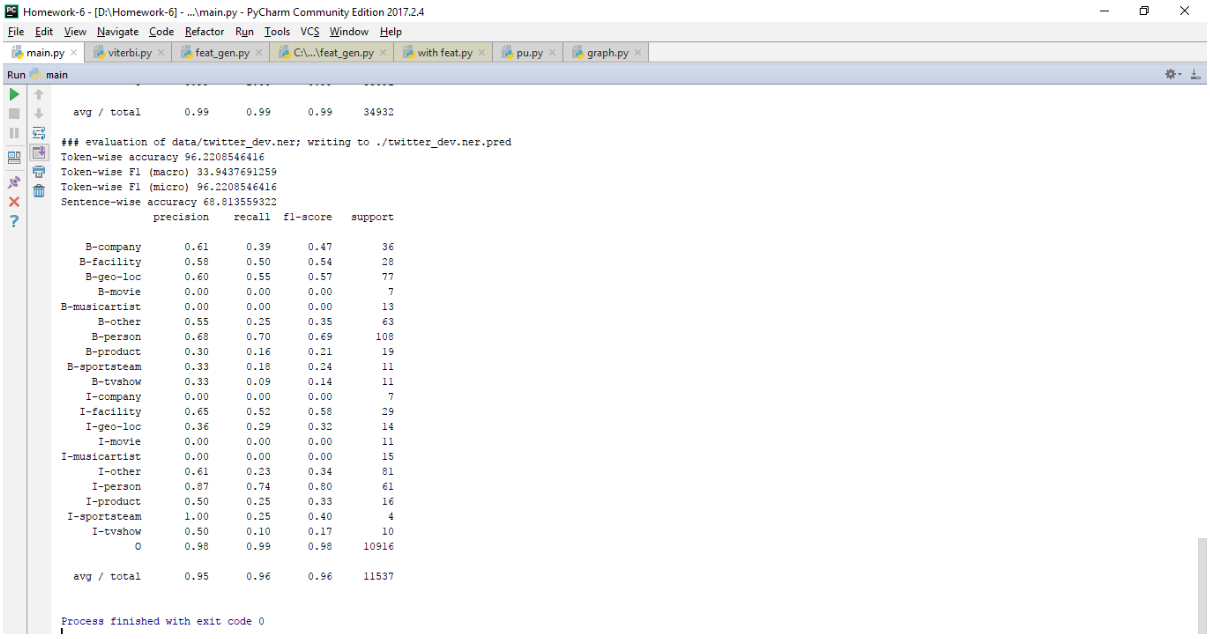
Homework-6 - [D:\Homework-6] - ...\main.py - PyCharm Community Edition 2017.2.4
File Edit View Navigate Code Refactor Run Tools VCS Window Help
main.py
viterbi.py
feat_gen.py
C:\...\feat_gen.py
with feat.py
pu.py
graph.py
if __name__ == ...
Run main

```
   avg / total       0.99      0.99      0.99     34932

### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 96.316200052
Token-wise F1 (macro) 29.3101228151
Token-wise F1 (micro) 96.316200052
Sentence-wise accuracy 68.9830508475
               precision    recall  f1-score   support

    B-company       0.82      0.39      0.53        36
   B-facility       0.67      0.43      0.52        28
    B-geo-loc       0.77      0.53      0.63        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       0.88      0.22      0.35        63
     B-person       0.74      0.66      0.70       108
    B-product       0.75      0.16      0.26        19
 B-sportsteam       0.00      0.00      0.00        11
     B-tvshow       0.33      0.09      0.14        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.47      0.31      0.38        29
    I-geo-loc       0.25      0.07      0.11        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.50      0.07      0.12        15
      I-other       0.59      0.21      0.31        81
     I-person       0.89      0.64      0.74        61
    I-product       0.80      0.25      0.38        16
 I-sportsteam       0.00      0.00      0.00         4
     I-tvshow       0.00      0.00      0.00        10
            O       0.97      1.00      0.98     10916

  avg / total       0.95      0.96      0.95     11537


Process finished with exit code 0
```

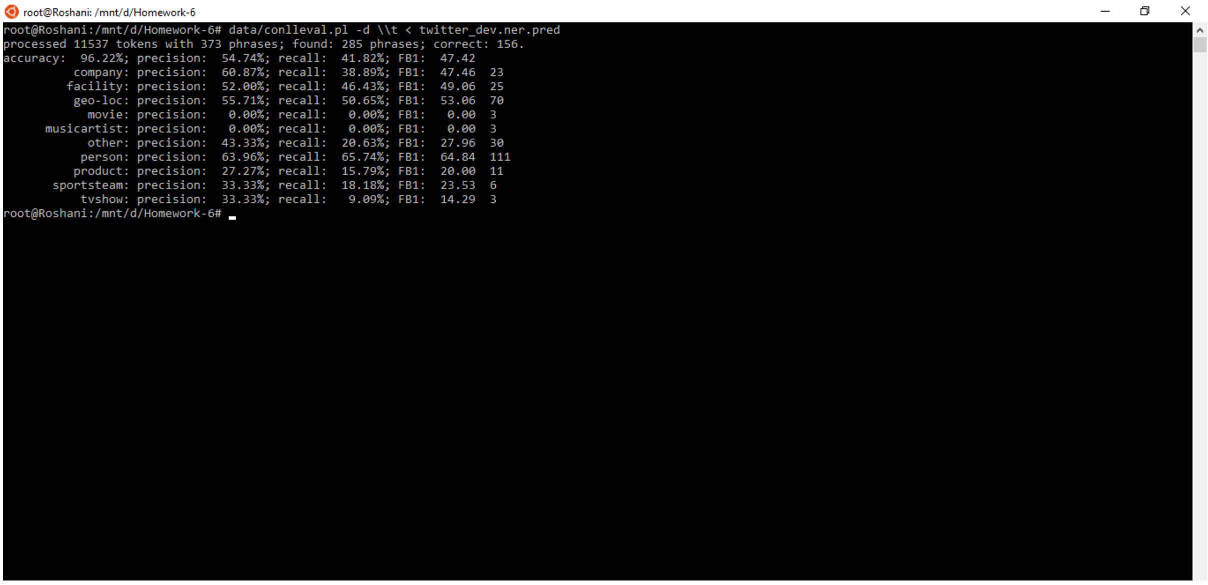


```
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 252 phrases; correct: 128.
accuracy:  96.32%; precision:  50.79%; recall:  34.32%; FB1:  40.96
          company: precision:  73.68%; recall:  38.89%; FB1:  50.91  19
         facility: precision:  13.79%; recall:  14.29%; FB1:  14.04  29
          geo-loc: precision:  66.67%; recall:  46.75%; FB1:  54.96  54
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  2
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
            other: precision:  29.41%; recall:  15.87%; FB1:  20.62  34
           person: precision:  59.80%; recall:  56.48%; FB1:  58.10  102
          product: precision:  60.00%; recall:  15.79%; FB1:  25.00  5
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
root@Roshani:/mnt/d/Homework-6#
```

CRF: Conditional Random Field (with features):

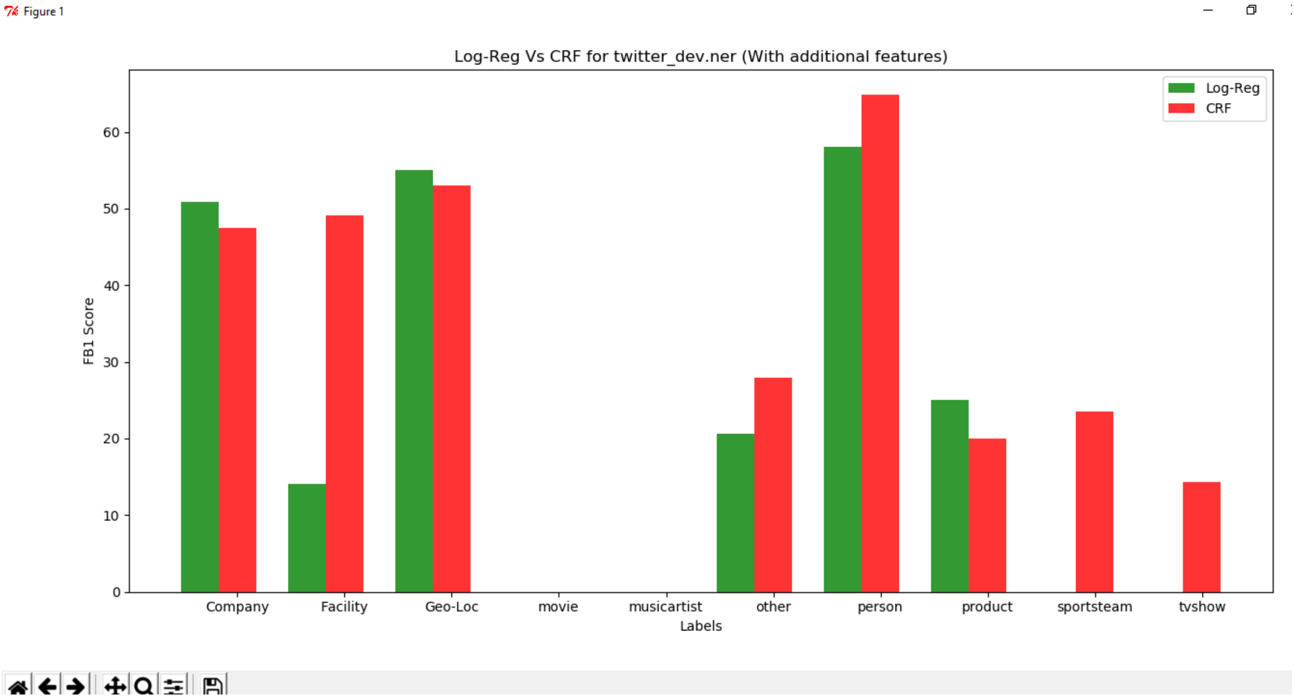

Homework-6 - [D:\Homework-6] - ...\main.py - PyCharm Community Edition 2017.2.4
File Edit View Navigate Code Refactor Run Tools VCS Window Help
main.py
viterbi.py
feat_gen.py
C:\...\feat_gen.py
with feat.py
pu.py
graph.py
Run main

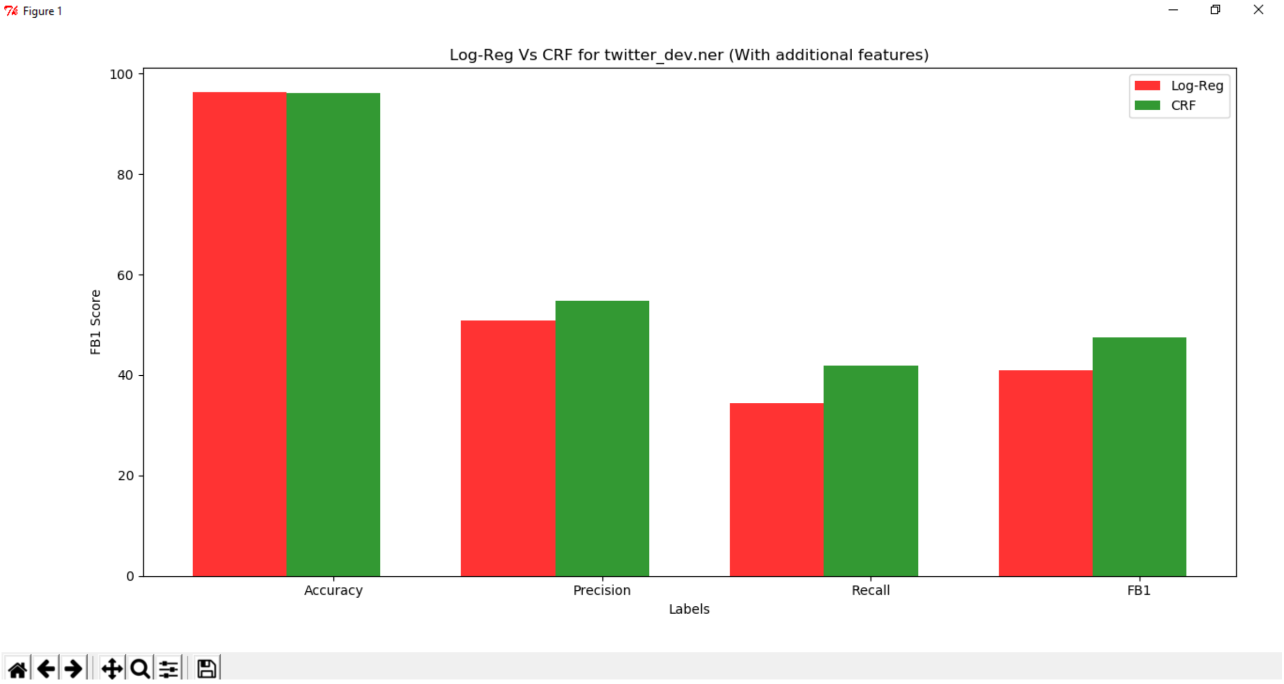
```
  avg / total       0.99      0.99      0.99     34932

### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 96.2208546416
Token-wise F1 (macro) 33.9437691259
Token-wise F1 (micro) 96.2208546416
Sentence-wise accuracy 68.813559322
              precision    recall  f1-score   support

    B-company       0.61      0.39      0.47        36
   B-facility       0.58      0.50      0.54        28
    B-geo-loc       0.60      0.55      0.57        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       0.55      0.25      0.35        63
     B-person       0.68      0.70      0.69       108
    B-product       0.30      0.16      0.21        19
  B-sportsteam      0.33      0.18      0.24        11
     B-tvshow       0.33      0.09      0.14        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.65      0.52      0.58        29
    I-geo-loc       0.36      0.29      0.32        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.61      0.23      0.34        81
     I-person       0.87      0.74      0.80        61
    I-product       0.50      0.25      0.33        16
  I-sportsteam      1.00      0.25      0.40         4
     I-tvshow       0.50      0.10      0.17        10
            O       0.98      0.99      0.98     10916

  avg / total       0.95      0.96      0.96     11537


Process finished with exit code 0
```



```
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 285 phrases; correct: 156.
accuracy:  96.22%; precision:  54.74%; recall:  41.82%; FB1:  47.42
          company: precision:  60.87%; recall:  38.89%; FB1:  47.46  23
         facility: precision:  52.00%; recall:  46.43%; FB1:  49.06  25
          geo-loc: precision:  55.71%; recall:  50.65%; FB1:  53.06  70
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
            other: precision:  43.33%; recall:  20.63%; FB1:  27.96  30
           person: precision:  63.96%; recall:  65.74%; FB1:  64.84  111
          product: precision:  27.27%; recall:  15.79%; FB1:  20.00  11
       sportsteam: precision:  33.33%; recall:  18.18%; FB1:  23.53  6
           tvshow: precision:  33.33%; recall:   9.09%; FB1:  14.29  3
root@Roshani:/mnt/d/Homework-6# 
```

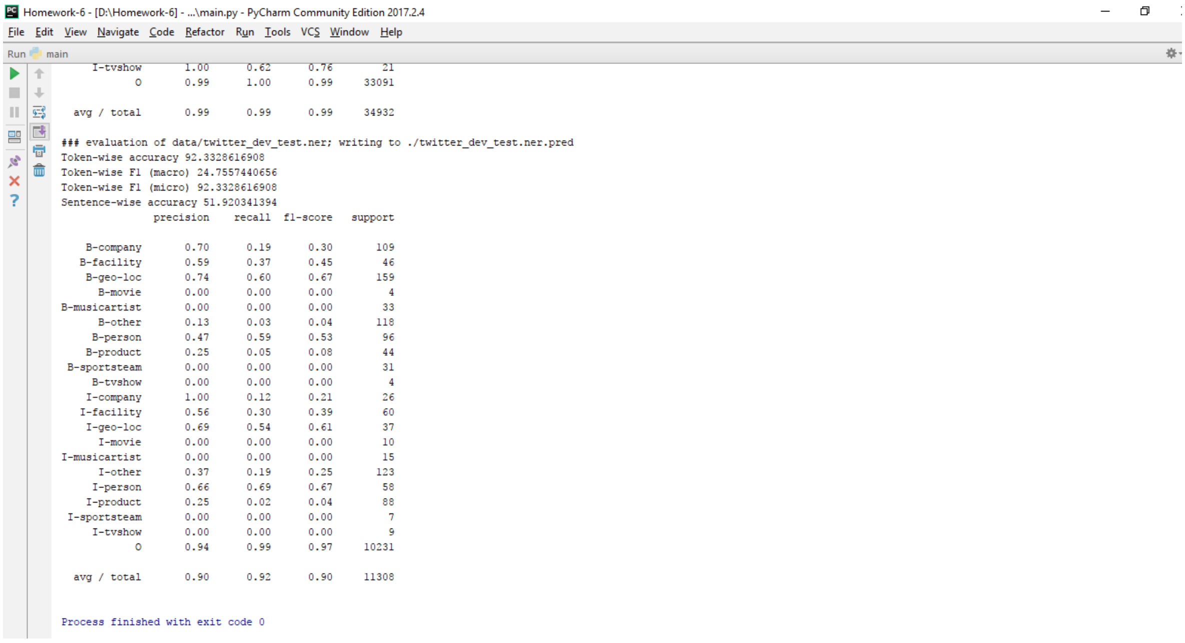
COMPARISION GRAPHS:





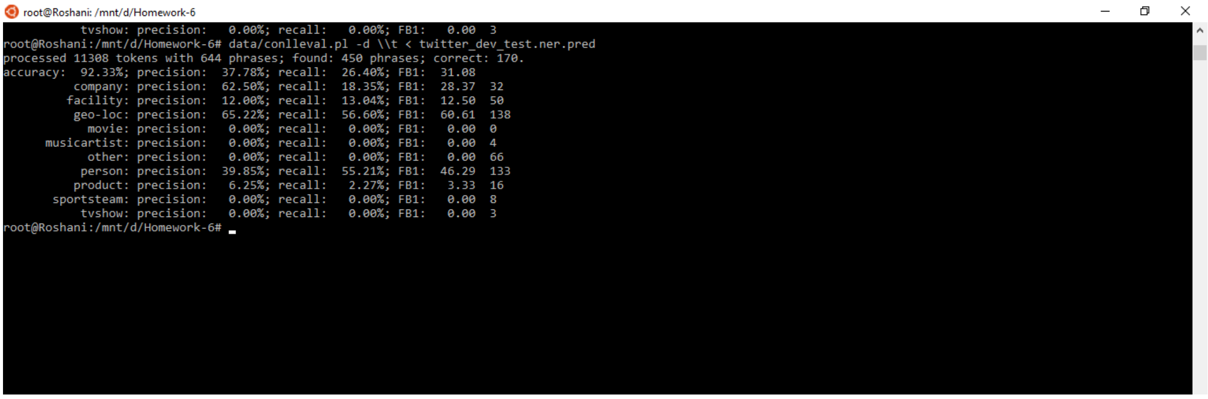
Twitter_dev_test.ner:

Log-reg: (With features)



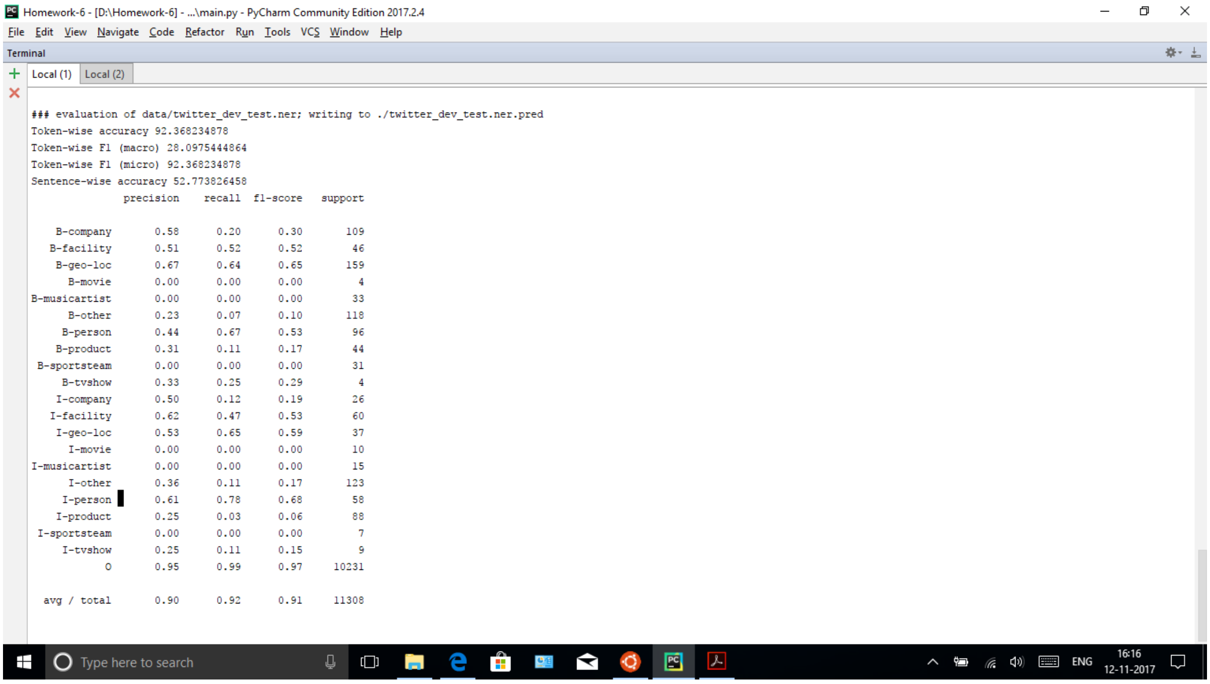
Homework-6 - [D:\Homework-6] - ...\main.py - PyCharm Community Edition 2017.2.4
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Run main

```
      I-tvshow       1.00      0.62      0.76        21
             O       0.99      1.00      0.99     33091

   avg / total       0.99      0.99      0.99     34932

### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 92.3328616908
Token-wise F1 (macro) 24.7557440656
Token-wise F1 (micro) 92.3328616908
Sentence-wise accuracy 51.920341394
               precision    recall  f1-score   support

    B-company       0.70      0.19      0.30       109
   B-facility       0.59      0.37      0.45        46
    B-geo-loc       0.74      0.60      0.67       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.13      0.03      0.04       118
     B-person       0.47      0.59      0.53        96
    B-product       0.25      0.05      0.08        44
 B-sportsteam       0.00      0.00      0.00        31
     B-tvshow       0.00      0.00      0.00         4
    I-company       1.00      0.12      0.21        26
   I-facility       0.56      0.30      0.39        60
    I-geo-loc       0.69      0.54      0.61        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.37      0.19      0.25       123
     I-person       0.66      0.69      0.67        58
    I-product       0.25      0.02      0.04        88
 I-sportsteam       0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.94      0.99      0.97     10231

  avg / total       0.90      0.92      0.90     11308


Process finished with exit code 0
```

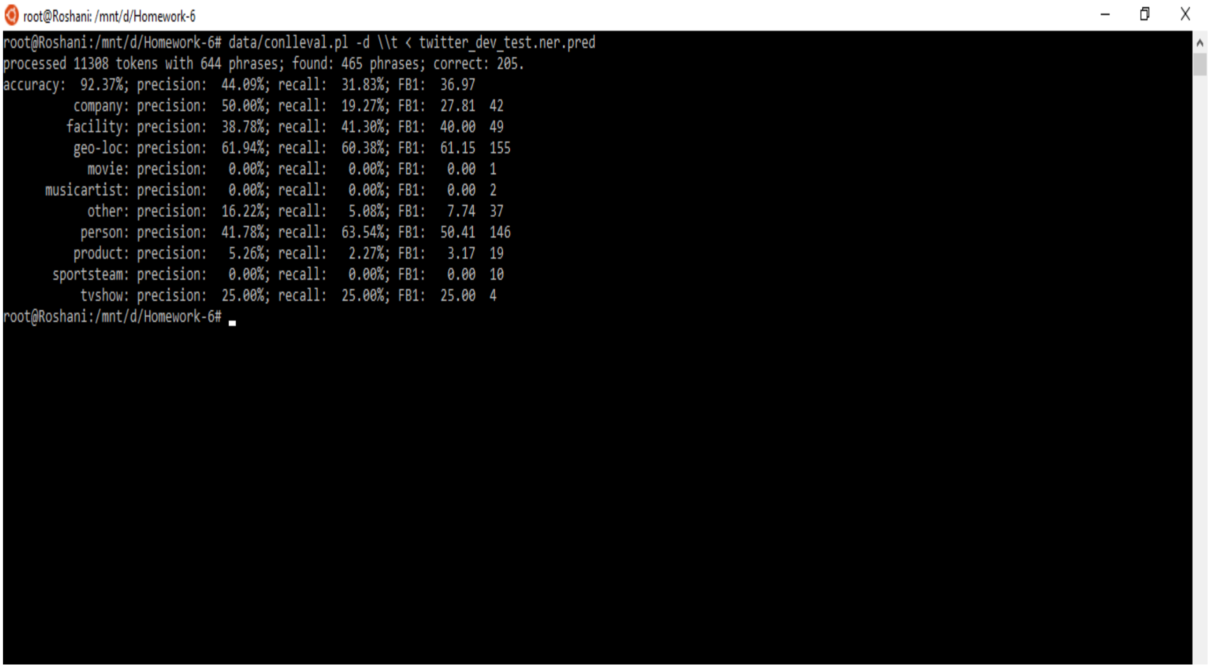


```
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 450 phrases; correct: 170.
accuracy:  92.33%; precision:  37.78%; recall:  26.40%; FB1:  31.08
          company: precision:  62.50%; recall:  18.35%; FB1:  28.37  32
         facility: precision:  12.00%; recall:  13.04%; FB1:  12.50  50
          geo-loc: precision:  65.22%; recall:  56.60%; FB1:  60.61  138
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  4
            other: precision:   0.00%; recall:   0.00%; FB1:   0.00  66
           person: precision:  39.85%; recall:  55.21%; FB1:  46.29  133
          product: precision:   6.25%; recall:   2.27%; FB1:   3.33  16
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  8
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  3
root@Roshani:/mnt/d/Homework-6# 
```

CRF: (With additional features)

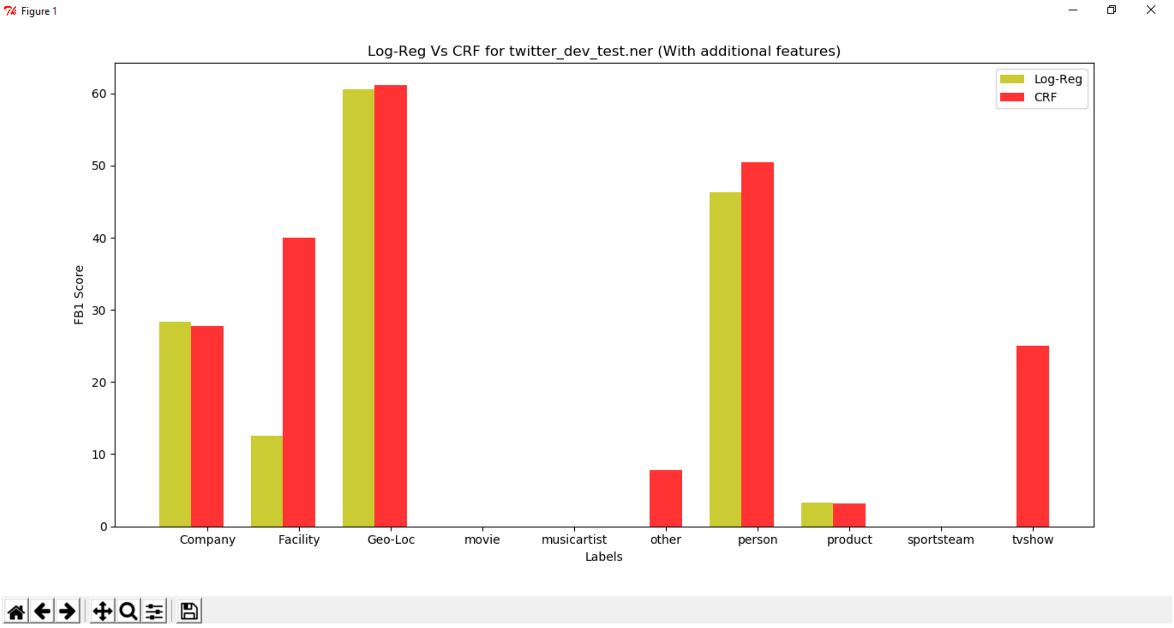

Homework-6 - [D:\Homework-6] - ...\main.py - PyCharm Community Edition 2017.2.4
File Edit View Navigate Code Refactor Run Tools VCS Window Help
Terminal
Local (1)
Local (2)
Type here to search
ENG
16:16
12-11-2017

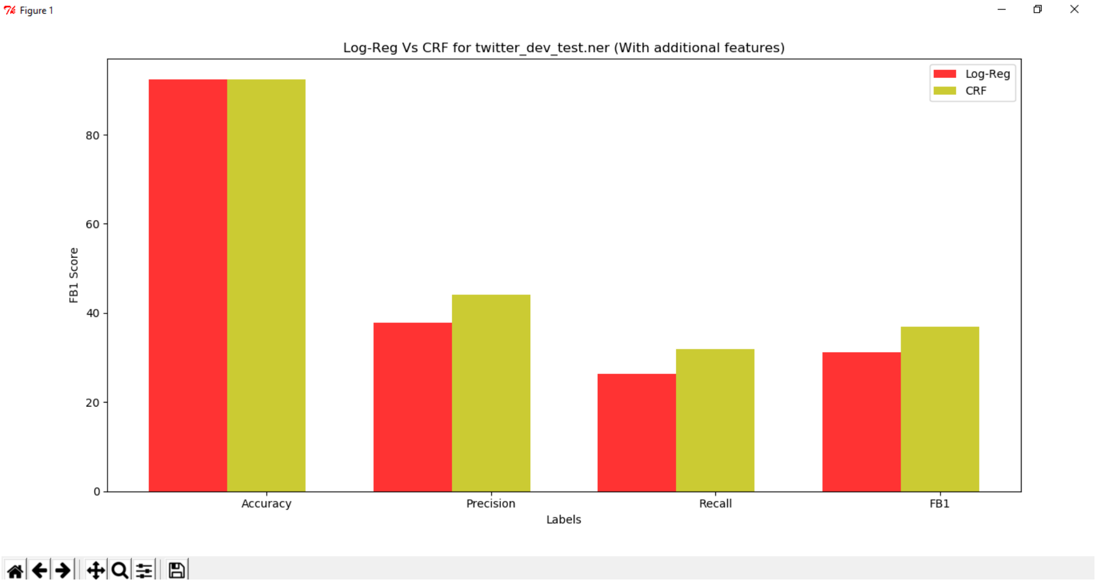
```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 92.368234878
Token-wise F1 (macro) 28.0975444864
Token-wise F1 (micro) 92.368234878
Sentence-wise accuracy 52.773826458
              precision    recall  f1-score   support

   B-company       0.58      0.20      0.30       109
  B-facility       0.51      0.52      0.52        46
   B-geo-loc       0.67      0.64      0.65       159
     B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
     B-other       0.23      0.07      0.10       118
    B-person       0.44      0.67      0.53        96
   B-product       0.31      0.11      0.17        44
 B-sportsteam       0.00      0.00      0.00        31
    B-tvshow       0.33      0.25      0.29         4
   I-company       0.50      0.12      0.19        26
  I-facility       0.62      0.47      0.53        60
   I-geo-loc       0.53      0.65      0.59        37
     I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
     I-other       0.36      0.11      0.17       123
    I-person       0.61      0.78      0.68        58
   I-product       0.25      0.03      0.06        88
 I-sportsteam       0.00      0.00      0.00         7
    I-tvshow       0.25      0.11      0.15         9
           O       0.95      0.99      0.97     10231

 avg / total       0.90      0.92      0.91     11308
```



```
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 465 phrases; correct: 205.
accuracy:  92.37%; precision:  44.09%; recall:  31.83%; FB1:  36.97
          company: precision:  50.00%; recall:  19.27%; FB1:  27.81  42
         facility: precision:  38.78%; recall:  41.30%; FB1:  40.00  49
          geo-loc: precision:  61.94%; recall:  60.38%; FB1:  61.15  155
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  2
            other: precision:  16.22%; recall:   5.08%; FB1:   7.74  37
           person: precision:  41.78%; recall:  63.54%; FB1:  50.41  146
          product: precision:   5.26%; recall:   2.27%; FB1:   3.17  19
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  10
           tvshow: precision:  25.00%; recall:  25.00%; FB1:  25.00  4
root@Roshani:/mnt/d/Homework-6# _
```

COMPARISION GRAPHS:





EXPLAINATION:

Here for each output 2 graphs are shown:

- **First graph shows the bars for log-reg Vs CRF scores for categories. Where the CRF score bars are mostly longer than log reg hence we can say that CRF tagger gives better results than logistic regression. (There are a few cases where log-reg gives zeros as scores but CRF still gives positive values). This holds true for with and without additional features.**
- **Second graph shows the bars for log-reg Vs CRF scores for evaluation factors such as accuracy, precision, recall and FB1 scores. Here also, we see that most times CRF gives better accuracy than log-reg in almost all cases. For accuracy it gives almost similar bar lengths.**

Being stated that CRF gives better results, here is the reason why?

Unlike log-reg that only considers the specific token and its context, CRF tagger also considers its neighbours context along with the given token.

So CRF would see the features and outputs of the previously tagged inputs and then predict the answer using Viterbi algorithm.