CSCE 638: Natural Language Processing

Assignment 1 Report

Submitted By – Sameer Kumar Behera   UIN: 526004296

System Requirements:

- Python must be installed.
- Installation Link: https://www.python.org/downloads/

Compile and Run Method:

1. Open a Terminal

2. Go to the Project Folder, i.e. PA1-638

3. Run > python SpamLord.py data_dev/dev/ data_dev/devGOLD

Result and Analysis:

Summary of the Results:

(TP) True Positives: 59

(FP) False Positives: 0

(FN) False Negatives: 0

Some Analytical Parameters:

Precision = TP/(TP+FP) = 1

True Positive Rate = Recall = TP/(TP+FN) = 1

Screenshot Attached:

```
beherasameer@VirtualBox:/media/sf_Shared_Drive_Ubuntu/nlp/PA1-638$ python SpamLord.py data_dev/dev/ data_dev/devGOLD
True Positives (59):
set([('ashishg', 'e', 'ashishg@stanford.edu'),
    ('ashishg', 'e', 'rozm@stanford.edu'),
    ('ashishg', 'p', '650-723-1614'),
    ('ashishg', 'p', '650-723-4173'),
    ('ashishg', 'p', '650-814-1478'),
    ('balaji', 'e', 'balaji@stanford.edu'),
    ('bgirod', 'p', '650-723-4539'),
    ('bgirod', 'p', '650-724-3648'),
    ('bgirod', 'p', '650-724-6354'),
    ('cheriton', 'e', 'cheriton@cs.stanford.edu'),
    ('cheriton', 'e', 'uma@cs.stanford.edu'),
    ('cheriton', 'p', '650-723-1131'),
    ('cheriton', 'p', '650-725-3726'),
    ('dabo', 'e', 'dabo@cs.stanford.edu'),
    ('dabo', 'p', '650-725-3897'),
    ('dabo', 'p', '650-725-4671'),
    ('dlwh', 'e', 'dlwh@stanford.edu'),
    ('engler', 'e', 'engler@lcs.mit.edu'),
    ('engler', 'e', 'engler@stanford.edu'),
    ('eroberts', 'e', 'eroberts@cs.stanford.edu'),
    ('eroberts', 'p', '650-723-3642'),
    ('eroberts', 'p', '650-723-6092'),
    ('fedkiw', 'e', 'fedkiw@cs.stanford.edu'),
    ('hager', 'e', 'hager@cs.jhu.edu'),
    ('hager', 'p', '410-516-5521'),
    ('hager', 'p', '410-516-5553'),
    ('hager', 'p', '410-516-8000'),
    ('hanrahan', 'e', 'hanrahan@cs.stanford.edu'),
    ('hanrahan', 'p', '650-723-0033'),
    ('hanrahan', 'p', '650-723-8530'),
    ('horowitz', 'p', '650-725-3707'),
    ('horowitz', 'p', '650-725-6949'),
    ('jks', 'e', 'jks@robotics.stanford.edu'),
    ('jurafsky', 'e', 'jurafsky@stanford.edu'),
    ('jurafsky', 'p', '650-723-5666'),
    ('kosecka', 'e', 'kosecka@cs.gmu.edu'),
    ('kosecka', 'p', '703-993-1710'),
    ('kosecka', 'p', '703-993-1876'),
    ('kunle', 'e', 'darlene@csl.stanford.edu'),
    ('kunle', 'e', 'kunle@ogun.stanford.edu'),
    ('kunle', 'p', '650-723-1430'),
    ('kunle', 'p', '650-725-3713'),
    ('kunle', 'p', '650-725-6949'),
    ('lam', 'e', 'lam@cs.stanford.edu'),
    ('lam', 'p', '650-725-3714'),
    ('lam', 'p', '650-725-6949'),
    ('latombe', 'e', 'asandra@cs.stanford.edu'),
    ('latombe', 'e', 'latombe@cs.stanford.edu'),
    ('latombe', 'e', 'liliana@cs.stanford.edu'),
    ('latombe', 'p', '650-721-6625'),
    ('latombe', 'p', '650-723-0350'),
    ('latombe', 'p', '650-723-4137'),
    ('latombe', 'p', '650-725-1449'),
    ('levoy', 'e', 'ada@graphics.stanford.edu'),
    ('levoy', 'e', 'melissa@graphics.stanford.edu'),
    ('levoy', 'p', '650-723-0033'),
    ('levoy', 'p', '650-724-6865'),
    ('levoy', 'p', '650-725-3724'),
    ('levoy', 'p', '650-725-4089')])
False Positives (0):
set([])
False Negatives (0):
set([])
Summary: tp=59, fp=0, fn=0
beherasameer@VirtualBox:/media/sf_Shared_Drive_Ubuntu/nlp/PA1-638$
```

Known Bugs & Limitations:

No bugs were found to the best of my knowledge. The python script with the regex cases added is able to correctly handle all the possible scenarios for emails and phone numbers with the given training data set. Additionally, some more plausible cases are also added.