CSCE-638, Programming Assignment #1 Report

About the Assignment

The program uses regular expressions that were coded to extract phone numbers and email addresses.

Contents of Submission

The deliverables of this assignment has been tabulated as below:

logs.txt	Contains the output logs.
sources/PA1	Contains the source files including py script, dev test cases
sources/PA1/SpamLord.py	Main program for regular expressions
PA1_Report_RizuJain_430000753.pdf	Assignment Report

Steps of Compilation and Execution

Platform Used: Python (Version 2.7)

Host PC: Ubuntu 18.04

*** Please note that the source file has only been developed for non-Mac OS and hence the _MACOSX directory has not been modified. The assignment would not be reflected in source files from _MACOSX.

Steps to setup the compilation and execution of the python program

1. Unzip the PA1_RizuJain.zip
2. Go to sources/PA1/ directory using the commands and run the following commands to execute the program SpamLord.py
    - `$ cd <path to sources/PA1/ directory>`
    - `$ python2.7 SpamLord.py data_dev/dev data_dev/devGOLD`

Result and Analysis

Upon the execution of the program the following logs can be obtained on the console:

```
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 (1):
set([('jure', 'e', 'server@cs.stanford.edu')])
```

```
False Negatives (0):
set([])
Summary: tp=59, fp=1, fn=0
```

INFERENCE:

- **The program is able to classify 59 developmental cases as true positive, correctly matching the patterns.**
- **There is 1 false positive where the program has incorrectly predicted a negative class as positive.**
- **The program has zero false negatives implying it has not detected any positive class as incorrect.**

Known Limitations

- The program output results into one false positive for a case of extracting email address from data_dev/dev/jure developmental test case. An email address pattern couldn't be found in its HTML page:

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html><head>
<title>403 Forbidden</title>
</head><body>
<h1>Forbidden</h1>
<p>You don't have permission to access /people/~jure/
on this server.</p>
<hr>
<address>Apache/2.2.4 (Fedora) Server at cs.stanford.edu Port
80</address>
</body></html>
```