NLP Assignment 1 Report

Compilation:

To compile the program, extract the zip file, move into the PA1 directory and use the following command:

```
python2 SpamLord.py data_dev/dev/ data_dev/devGOLD
```

Version of python used: Python 2.7.12

Results and Analysis:

For the Dev set:

True Positives	False Positives	False Negatives
59	0	0

The program identifies email ids and Phone numbers for all formats that are present in the dev set. The program also works for several other general scenarios that may not be present in the dev set. With respect to the dev set, the program has 100% precision and 100% recall.

Known bugs/Limitations/problems:

There are no known bugs/problems with the program. However, the main limitation is that the program can only fetch expressions that "look like" an e-mail id. There is no way of verifying if the email id exists or is a fake one.

Also, the program can be extended further to carry out a more exhaustive search and identify more types of email id, including organizational email ids and obfuscated email ids, which are beyond the scope of the program.