<u>README</u>

The python program files for each part of the assignment are as follows:

1) Part1: Tokenisation and POS tagging: **Part_1.py**
2) Part2: Regular Expressions, FSAs and FSTs: **Part_2.py**
3) Part3: N-Gram Models: **Part_3.py**

All these files can run with the following simple command:

python3 filename

Note: While running Part_3.py, please ensure that the .txt files are in the same working directory.

Also, it is assumed that the python compiler is already equipped with all the packages used in the program, namely: nltk, BeautifulSoup4, pandas, collections. If not, kindly install it from the command prompt using the following command:

pip install packagename