**Name:** Jai Kumar

**Roll. No.:** 15076

**Course :**  ECS 405

Total distinct words: 10380

Total distinct letters(including numbers): 36

1. Saved in file '*lower-austen-courpus.txt*' in .tar

2. Saved in file '*freq\_letters.txt*' and '*freq\_words.txt*' in .tar

3. Done in program in file '*hw2.py*'

4. Done in program in file '*hw2.py*'

5. Done in program in file '*hw2.py*'

6. Top five frequent words and letters

Rank 1 *the*: 12638 and *e* : 199579

Rank 2 *to* : 12166 and *t* : 136388

Rank 3 *and* : 11189 and *a* : 124196

Rank 4 *of* : 10435 and *o* : 122831

Rank 5 *a* : 6817 and *n* : 112450

Highly frequent words used are ‘*a,the*’ in english language which are there in top 5.

Every word of contains either of the five vowels (a,e,i,o,u) and top five include ‘*e,a,o’*.

7. Saved in file ‘*letter.pdf*’ and ‘*words.pdf’* in .tar file

8. Saved in file ‘*log\_letter.pdf*’ and ‘*log\_words.pdf’* in .tar file

9.Pearson correlation coefficient

words: -0.17310381832898336 (little random)

letters: -0.907222999056525 (almost straight line)

log\_words: -0.580914020071538

log\_letters: -0.7567212667416133