**CSCE 5290**

**NATURAL LANGUAGE PROCESSING**

**IN CLASS ASSAIGNMENT 1**

Name – Kirankumar Danthala ID – 11514845

1. **Why we use stopwords? Why stopwords are not necessary for NLP frequency distribution.?**

Generally, we use stop words in Text mining and NLP to eliminate the words which are used very frequently that carry very little useful information. So when we apply frequency distribution technique on a particular url we may end up with the words which are stopwords, as we write everything in English we will use stopwords definitely and mostly to describe anything. But we don’t need those stopwords to know the exact information what the url contains.

1. **Based on high frequency words what information you can extract from the graph?**

As we removed stop words, now it is easy to find out best high frequency words. According to my data from the link <https://en.wikipedia.org/wiki/SpaceX> , word ‘original’ stands at top with word count slightly greater than 170, and word ‘SpaceX’ is in second place with word count around 145. Only these both above mentioned words have word count greater than 100, rest all the other words are having word count less than 100. Word ‘march’ is staying low in the frequency distribution chart I have created.

1. **Can you provide different visualization for frequency distribution? If yes, please perform. If no, why?**

Yes, I have created Pie-chart visualization which is mentioned in the code.

**Github link:**

<https://github.com/Danthalakirankumar/NLP-ICE-1>

<https://github.com/Danthalakirankumar/NLP-ICE-1/wiki>