**NLP project Round1 report**

Submitted by Members of Team Language Revolution

Jatin Dahiya 19ucs033

Rishabh Sahu 19ucs129

Naman Dhanotia 19ucs128

**Problem Description**

To take two books from <http://gutenburg.org> in .txt format and perform the following Natural Language processing operations on them

* Apply Data preprocessing on the text
* Generating frequency distributions of the words
* Creating word clouds from the text before and after removing stopwords
* Evaluating relationship between word length and frequency
* Parts of Speech tagging for the words in the text

**Data Preprocessing steps**

We performed the following data preprocessing steps

Removing chapter number and chapter Headings

Removing all punctuation marks

**Data Preparation**

For preparing data to be analysed we performed the following steps

1. Tokenising the text into a list of words

**Problem Statement**

**Illustrations (Plots ,tables and figures and output)**

**Inferences**