SYNOPSIS

Project Details:

- Group Members: Prashant goyal (1504506)
- Supervisor: **Assistant Professor AK Singh (CSE Department)**
- Departmental Supervisor: **Professor Pradeep Ahuja (Chemical Department)**
- Project Area: Natural Language Processing / Deep Learning
- Project Topic: **Word Complexity Identification**

Work Progress:

- Our group is pursuing with the following courses as advised by our Supervisor as pre-requisites for the project.
    1. **CS224n-Natural Language Processing with Deep Learning** (Stanford University) by Richard Socher.
    2. **Introduction to Natural Language Processing (Stanford University)**- Chris Manning and Dan Jurafsky

- The important topics/concepts to learn are:
    1. Text Tokenization
    2. Naïve Bayes Modelling
    3. N-gram language modelling
    4. Tagging Problems
    5. Parsing Problems
    6. Neural Networks (RNN, LSTM)
    7. Markov Processes
    8. Language Models
    9. Hidden Markov Models

Project Overview:

Complex Word Identification consists in determining which words in a given sentence can challenge the readers of a certain target audience. The goal of the proposed Complex Word Identification shared task is to provide a framework for the evaluation of methods for this first step in a Lexical Simplification pipeline. This is a simple, well-defined and yet challenging task in the community of Lexical and Text Simplification.

Project Goals:

- To learn **which words** challenge non-native English speakers, and to understand what are their **defining characteristics**.
- To know how well one's **individual vocabulary limitations** can be predicted from the overall **limitations of a group** which they are part of.
- To introduce a new resource to be used not only in **Text Simplification**, but also in any task related to **Topic Modelling** and **Semantics**.

Prof. P Ahuja

(Co-Supervisor)

Asst. Prof. AK Singh

(Supervisor)