# **Project Proposal**

Name: Subhasis Dutta

ID: sxd150830

# **Topic: Spelling Error detection and find most probable word suggestion using Kernighan method**

# **Overview**

The following is the project idea proposal:

Given a corpus of words with spelling mistakes find all conditional probability and apply Bayes noise channel to spellings in different sentence and detect misspelled words. Along with find the most probable replacement.

# **Data source**

The following sources would be considered for obtaining the data required to implement this project:

1. Penn Treebank/ Small data corpus with parts of speech tags.
2. Corpus of words with common spelling mistakes.

# **Input**

The following would be the input to the program:

1. A sentence as a string.

# **Expected Results**

The following would be expected as output after proper input and valid training:

1. Given a sentence detect the misspelled words and find the most probable correct word replacements.

# **Other Details**

Language – Python

Expected time for implementation – 20 Hours

# **Project Proposal**

Name: Subhasis Dutta

ID: sxd150830

# **Topic: Name Entity Recognition systems using recurrent neural network built on TensorFlow and compare with Stanford Named Entity Recognizer built with CRFClassifier**

# **Overview**

The following is the project idea proposal:

The idea is to implement a solution for Name Entity Recognition system using Recurrent Neural Network available through TensorFlow. The goal is to train a recurrent neural network on a challenging task of language modeling. The goal of the problem is to fit a probabilistic model which assigns probabilities to sentences. It does so by predicting next words in a text given a history of previous words.

If time permits will run the same sentences on StanfordNER and compare the results.

# **Data source**

The following sources would be considered for obtaining the data required to implement this project:

1. Penn Treebank/ Small data corpus of name entity tagged data set.

# **Input**

The following would be the input to the program:

1. A sentence as a string.

# **Expected Results**

The following would be expected as output after proper input and valid training:

1. Generate the parse thee for given sentence along with entity tagging of person and organization.

If time permits compare the NER built with neural network with NER built with sequence models from Stanford NER.

# **Other Details**

Language – Python

Expected time for implementation – 25 Hours