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