

Spelling Corrector

Name of team members:

1. Ameya Parab

UTD ID: 2021166954

email: axp132530@utdallas.edu

2. Binita Sinha

UTD ID: 2021196721

email: bxs133130@utdallas.edu

Project Description:

Spelling correction is a very important component of any system processing text. Creation of the text is prone to errors, people make typing errors or they sometimes ignore the correct spelling of the word. Therefore they need spelling correctors. Another issue where spelling correction is highly necessary is optical character recognition systems. These systems often make errors during recognition due to bad quality of input, change of fonts or insufficient training. Today, many spelling correction functions are available in many languages like English.

Project Functional Requirements:

- First a large set of training set is needed to train the corpus for unigrams, bigrams.
- Unigrams are used to find the set of nearest close misspelled words.
- Bigrams are used for getting the probability of the best match by using an algorithm.

Working

- The trained dataset is to be used for testing a sentence for any misspelled words.
- The program will have a sentence boundary.
- The program will use bigram data to predict the best match case and will highlight the corrected word.
- The Accuracy of the project to be 80-90%.
- For new words a smoothing algorithm to be implemented in-order to give some weightage.