NATURAL LANGUAGE PROCESSING

# Integrated Text Editor

Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human–computer interaction. Many challenges in NLP involve natural language understanding, that is, enabling computers to derive meaning from human or natural language input, and others involve natural language generation.

To implement: A live text editor that has the following:

1. Spell Checker/Corrector: A live automated spell checker that detects and corrects misspelt errors and also, corrects contextual errors. Underlines words that are not in the dictionary, option to add words into the dictionary:
   1. Non word error detection using a Hash Map amalgamation of a corpus.
   2. Generating candidates of the misspelt word, using Levenstein’s Edit Distance Model (incorporated an edit distance of 2).
   3. Automatic correction using a Bayesian Model and the Noisy channel Model to make the most appropriate choice of correction.
   4. Bayesian Model for word segmentation: ‘iloveindia’ to ‘I love india’
   5. Add to dictionary.
   6. Underlining of words that are misspelt
   7. Right Click feature, to generate candidate words, if an error exists.
2. Word Segmentation:   
   A live feature that check if the user, has missed out spaces between words, and tries to automatically add spaces.
3. Auto-complete: An option that continuously scans for long words and completes them automatically whenever applicable.
4. Dynamic find: That provides an option to the user to search for regular expressions
5. Topic Extraction: To provide search links to the user to arrange for the privilege of easily searching for additional information on the text being typed. This requires text analysis, topic modeling, and keyword extraction.
   1. Detection using keyword mapping, and
      1. Letter Detection
      2. Resume Detection
      3. Code Detection
   2. Topic/Keyword Extraction
      1. Essay theme
      2. Word Frequency Analysis
6. Background Processing: All features must be background processes, to provide seamless smooth operations of the editor