Assignment -2

Token Healing

Submitted by:

Devansh Verma

Introduction:

The Token Healing algorithm is a technique used in natural language processing to handle out-of-vocabulary (OOV) or unknown words by replacing them with similar known words. This technique is commonly used in editors apps like Microsoft word, PowerPoint, one note etc. The token healing algorithm can be achieved by using various existing natural language processing (NLP) libraries like NLTK, spacy or hugging face’s transformers to assist in tokenization, error detection and correction.

Approach

My answer to the given problem is divided into 2 phases namely spell check and grammar check algorithm.

The spell check algorithm uses ‘pyspellchecker’ library to perform spell check on a given word. ‘pyspellchecker’ is a pure Python spell checking library that uses Peter Norvig's algorithm. It provides simple and efficient spell-checking capabilities.

After working with many different words, it is found that the algorithm works with reasonable accuracy and efficiency.

The grammar check algorithm uses ‘language\_tool\_python’ library to perform grammar check on a given sentence. The language\_tool\_python library is a Python wrapper for the Language Tool grammar checking tool. It allows you to perform grammar and style checking on text using Language Tool's extensive set of rules. The algorithm takes a sentence in input and gives the correct sentence after correcting all the grammatical errors and identifies the type of grammatical error in the input sentence.

After working with many different sentences, it is found that the algorithm works with reasonable accuracy and efficiency