ENSE 480

Dillan Zurowski

Project Proposal

Jan 8, 2023

Introduction

As a 4th year student, I have written many lengthy reports and have taken advantage of the use of autocorrect countless times. For my ENSE 480 final project, I will implement an autocorrect AI algorithm that will take a sentence or paragraph and make automatic corrections to it. I will use Python to create this project as I am learning it simultaneously in a different class and can use this project to further challenge myself using this language. The main intent of this project is to help those writing large reports or papers by allowing them to focus more on the topic and less on spelling and grammatical mistakes.

Scope

The program's goal is to do the following:

- Take an input string of variable length
- Compare each word to a pre-made library of a list of English words.
- If the word is not in the dictionary, perform one of 4 options
    - Insert (will add a letter)
    - Delete (will remove a letter)
    - Switch (it will swap two nearby letters)
    - Replace (exchange one letter to another one)
- Replace the word in the original string with the correct word
- Give the string sentence back to the user.

Optional

- Correct punctuation mistakes such as commas, periods, and capitalization
- Check if the sentence makes sense
- Give the user a list of mistakes with an option to change each one
    - Give more then one option

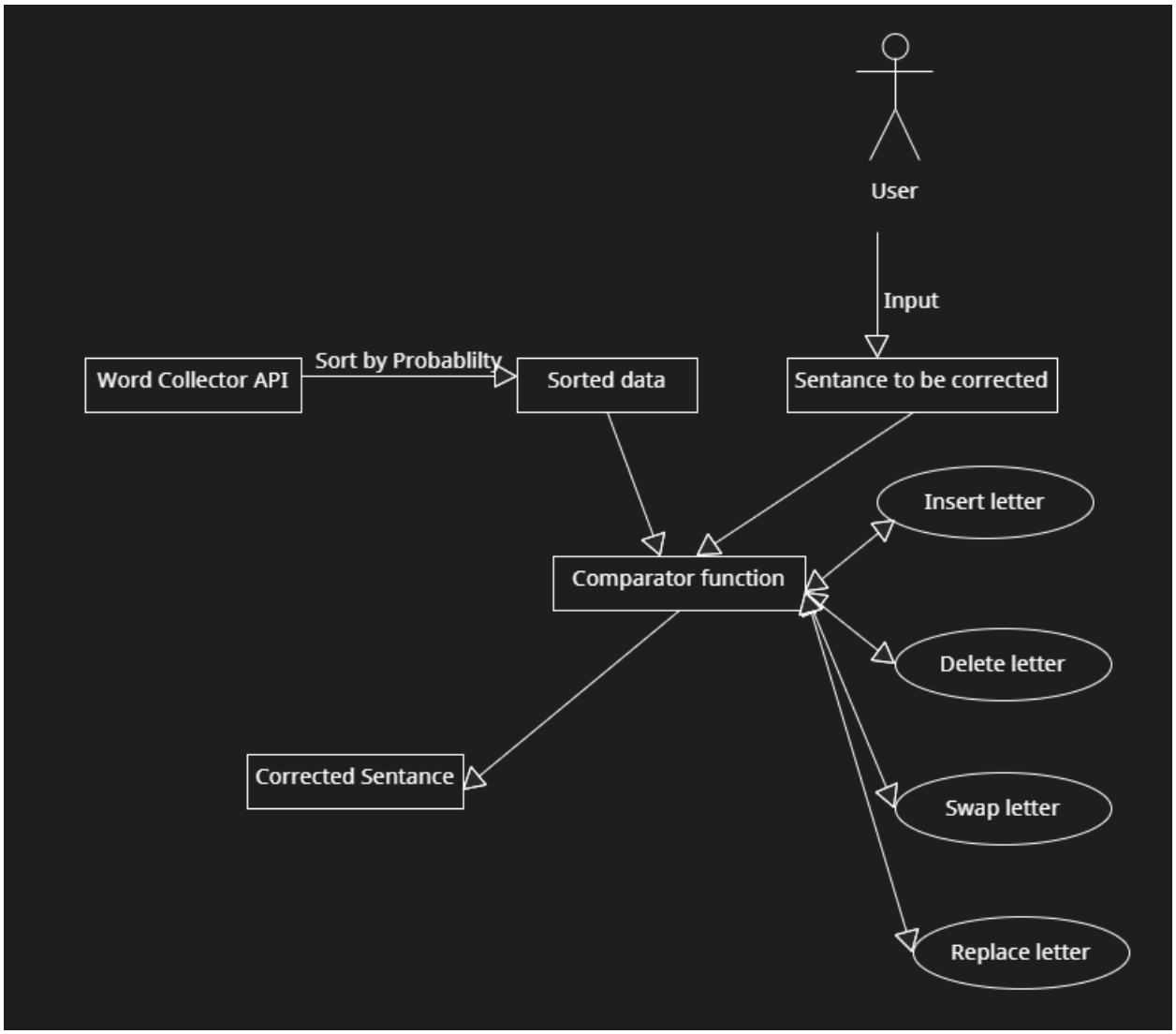
Representation and Data Structures

I plan on using a string for the input which can be manipulated in the same way as an array. I chose this as it is the easiest to create an input and can be changed or replaced as needed as well. I may also use a list to represent a temporary storage for a word if it needs to be tested by manipulating each letter.

Techniques and Algorithms

1. Collection of english words
    a. English-words 2.0.0 or if too large
    b. https://github.com/dentex22/Autocorrect_System/blob/main/sample.txt
2. Computational linguistics algorithm: **Natural Language Processing** (Python package) Python package that contains several functions related to finding word frequency which is used to determine probability of a word and select the most-likely suggestion.

A Structure Diagram



References

Kadlaskar, Amruta. “Autocorrect Feature Using NLP in Python: What Is Autocorrect.” *Analytics Vidhya*, 8 Nov. 2021, https://www.analyticsvidhya.com/blog/2021/11/autocorrect-feature-using-nlp

-in-python/#:~:text=Now%20let%20us%20get%20started,as%20f%3A%20file_name_data%20%3D%20f

"Building an Autocorrect Feature Using NLP with Python." *Section*, 4 Dec. 2021, https://www.section.io/engineering-education/building-autocorrect-feature-using-nlp-with-python/

"English-Words." *PyPI*, 5 Jan. 2023, https://pypi.org/project/english-words/#:~:text=To%20add%20a%20word%20list,run%20the%20script%20process_raw_data.py%20