Natural Language Processing Lab

Spell Checker using Trie and Levenshtein Distance

Name:J Krithika , 22011101046

Aim

To develop a spell checker that identifies candidate words using a Trie structure and corrects misspelled words based on minimum edit distance (Levenshtein Distance), incorporating contextual corrections for sentences.

Source Code

```
import re
from collections import defaultdict
import numpy as np

class TrieNode:
    def __init__(self):
        self.children = {}
        self.is_end_of_word = False

class Trie:
    def __init__(self):
        self.root = TrieNode()

    def insert(self, word):
        node = self.root
        for char in word:
            if char not in node.children:
                node.children[char] = TrieNode()
            node = node.children[char]
        node.is_end_of_word = True

    def search_candidates(self, prefix):
        node = self.root
        for char in prefix:
            if char not in node.children:
                return []
            node = node.children[char]

        return self._collect_words(node, prefix)

    def _collect_words(self, node, prefix):
        words = []
```

```
        if node.is_end_of_word:
            words.append(prefix)
        for char, child_node in node.children.items():
            words.extend(self._collect_words(child_node, prefix + char))
        return words

def levenshtein_distance(word1, word2):
    dp = np.zeros((len(word1) + 1, len(word2) + 1), dtype=int)

    for i in range(len(word1) + 1):
        dp[i][0] = i
    for j in range(len(word2) + 1):
        dp[0][j] = j

    for i in range(1, len(word1) + 1):
        for j in range(1, len(word2) + 1):
            if word1[i - 1] == word2[j - 1]:
                dp[i][j] = dp[i - 1][j - 1]
            else:
                dp[i][j] = 1 + min(dp[i - 1][j], dp[i][j - 1], dp[i - 1][j
                    - 1])

    return dp[len(word1)][len(word2)]

def spell_checker(sentence, dictionary):
    trie = Trie()
    for word in dictionary:
        trie.insert(word)

    words = sentence.split()
    corrected_sentence = []

    for word in words:
        word_cleaned = re.sub(r"[^a-zA-Z]", "", word)
        candidates = trie.search_candidates(word_cleaned)

        if candidates:
            distances = {candidate: levenshtein_distance(word_cleaned,
                candidate) for candidate in candidates}
            corrected_word = min(distances, key=distances.get)
            corrected_sentence.append(word.replace(word_cleaned,
                corrected_word))
        else:
            corrected_sentence.append(word)

    return "␣".join(corrected_sentence)

def main():
    dictionary = ["weather", "leather", "principal", "principle", "accepted
        ", "excepted", "lose", "loose",
                  "later", "latter", "stationery", "stationary", "affects",
                      "effects", "council", "counsel",
                  "too", "to", "bare", "bear", "fur", "far", "furthest", "
                     farthest", "advise", "advice",
                  "quiet", "quite", "heap", "hip", "their", "there", "lose"
                      , "loose"]

    sentences = [
```

```
         "During␣the␣summer␣we␣have␣the␣best␣ueather.",
        "I␣have␣a␣black␣ueather␣jacket,␣so␣nice.",
        "Mr␣Patrick␣is␣our␣new␣principle.",
        "The␣company␣excepted␣all␣the␣terms.",
        "Please␣ d o n t ␣keep␣your␣dog␣on␣the␣lose.",
        "The␣later␣is␣my␣best␣friend.",
        "I␣need␣some␣stationary␣products␣for␣my␣craftwork.",
        "The␣actor␣excepted␣the␣Oscar.",
        "I␣will␣call␣you␣later␣in␣the␣evening.",
        "Covid␣affects␣the␣lungs.",
        "The␣council␣of␣the␣ministers␣were␣sworn␣in␣yesterday.",
        "Robert␣too␣wants␣to␣accompany␣us␣to␣the␣park.",
        "Mia␣will␣council␣me␣about␣choosing␣fashion␣as␣my␣career.",
        "The␣bear␣at␣the␣zoo␣was␣very␣playful.",
        "The␣sheep␣have␣a␣lot␣of␣fur␣that␣keeps␣them␣warm.",
        "The␣hot␣spring␣is␣at␣the␣furthest␣corner␣of␣the␣street.",
        "Can␣you␣advise␣me␣on␣how␣to␣study␣for␣exams?",
        "The␣team␣will␣lose␣the␣match␣if␣they␣ d o n t ␣play␣well.",
        "Can␣you␣go␣to␣the␣market␣for␣me?",
        "The␣teachers␣asked␣the␣students␣to␣keep␣quiet.",
        "The␣heap␣of␣garbage␣should␣be␣cleaned␣immediately.",
        "This␣is␣their␣house."
    ]

    for sentence in sentences:
        corrected = spell_checker(sentence, dictionary)
        print(f"Original:␣{sentence}")
        print(f"Corrected:␣{corrected}\n")

if __name__ == "__main__":
    main()
```

Output

```
Original: During the summer we have the best ueather.
Corrected: During their summer weather have their best ueather.

Original: I have a black ueather jacket, so nice.
Corrected: I have advise black ueather jacket, so nice.

Original: Mr Patrick is our new principle.
Corrected: Mr Patrick is our new principle.

Original: The company excepted all the terms.
Corrected: The company excepted all their terms.

Original: Please don't keep your dog on the lose.
Corrected: Please don't keep your dog on their lose.

Original: The later is my best friend.
```

Corrected: The later is my best friend.

Original: I need some stationary products for my craftwork.
Corrected: I need some stationary products for my craftwork.

Original: The actor excepted the Oscar.
Corrected: The actor excepted their Oscar.

Original: I will call you later in the evening.
Corrected: I will call you later in their evening.

Original: Covid affects the lungs.
Corrected: Covid affects their lungs.

Original: The council of the ministers were sworn in yesterday.
Corrected: The council of their ministers were sworn in yesterday.

Original: Robert too wants to accompany us to the park.
Corrected: Robert too wants to accompany us to their park.

Original: Mia will council me about choosing fashion as my career.
Corrected: Mia will council me about choosing fashion as my career.

Original: The bear at the zoo was very playful.
Corrected: The bear at their zoo was very playful.

Original: The sheep have a lot of fur that keeps them warm.
Corrected: The sheep have advise lot of fur that keeps them warm.

Original: The hot spring is at the furthest corner of the street.
Corrected: The hot spring is at their furthest corner of their street.

Original: Can you advise me on how to study for exams?
Corrected: Can you advise me on how to study for exams?

Original: The team will lose the match if they don’t play well.
Corrected: The team will lose their match if they don’t play well.

Original: Can you go to the market for me?
Corrected: Can you go to their market for me?

Original: The teachers asked the students to keep quiet.
Corrected: The teachers asked their students to keep quiet.

Original: The heap of garbage should be cleaned immediately.
Corrected: The heap of garbage should bear cleaned immediately.

Original: This is their house.

```
Corrected: This is their house.
```

Result

The spell checker was implemented successfully using a Trie structure and Levenshtein distance. However, contextual accuracy was inconsistent, as seen in cases like "During their summer weather" and "I have advise black ueather jacket." Further refinement in the algorithm, such as context-based word prediction using a language model, is needed for enhanced performance.