

SPELL CHECKER and SUGGESTER

Guided by:

Resp. Rachna Ma'am

Presented by-

Aryan Ojha (2021BIT006)

Ronak Pawar (2021BIT022)

Sharvari Salodkar (2021BIT027)

Shraddha Lokhande (2021BIT041)

Overview

Introduction

Objective

Literature Review

Methodology

Applications

Future Scope

References

Software Used

Conclusion

Introduction

NEED STATEMENT:

We had to rely on our own prior knowledge of spelling, while typing documents, letters, reports, etc.

This process is time consuming and often results in many spelling mistakes making our documents unprofessional.

PROBLEM STATEMENT:

In order to solve this problem, we came up with the idea of developing a SPELL CHECKER and SUGGESTER.

Introduction

- Spell checker is the software feature that checks for Miss spelling in a text as well as gives the correct suggestion for the same.
- Spell checker is the important feature that we can use in various software services like search engines, emails, dictionaries, word-processors, etc.
- Spell checkers mostly use approximate string-matching algorithm to find correct spelling of misspelled words.
- We are specifically using Levenshtein Distance Algorithm(Edit distance problem) for our project.

Objective

- Objective of our project is to make a Spell checker which will correct spelling error in any written text.
- Our Spell checker uses Levenshtein Distance Algorithm and a dictionary to scan the text and compare each word with a database of correctly spelled words.
- If a word is not found in the dictionary, the spell checker highlights it as a potential spelling error and suggests one or more possible corrections.

Literature review

- The first spell checker that accessed 10,000 acceptable words was made by Les Earnest in 1961.
- Ralph Gorin and graduate student under Earnest at that time created the first true spelling checker program written as an application program in assembly language.
- It was made at Stanford University's artificial intelligence laboratory in February 1971.

Evolution of SpellCheckers

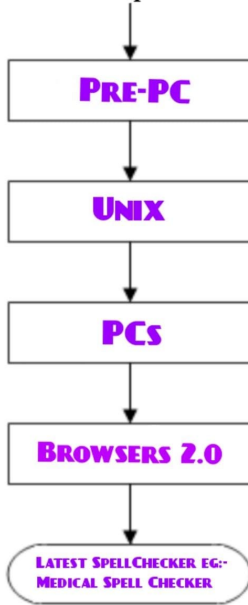


Figure: Caption

Methodology

DATA COLLECTION

- We have taken the data set of words ie.Dictionary from kaggle.
- We took the help of GfG to understand the Edit-distance algorithm.

DATA PREPARATION

- In data preparation we sorted the raw data(words) and structured it properly.

Algorithm

- The algorithm that we have used in our project is- Levenshtein Distance Algorithm.
- The Levenshtein Distance, between two words is the minimum number of single character edits.
- Insertion, deletion and substitution are the main three operations required to change one word into other.
- It is also known as Edit-distance.

$$\text{lev}(a, b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \text{lev}(\text{tail}(a), \text{tail}(b)) & \text{if } a[0] = b[0], \\ 1 + \min \begin{cases} \text{lev}(\text{tail}(a), b) \\ \text{lev}(a, \text{tail}(b)) \\ \text{lev}(\text{tail}(a), \text{tail}(b)) \end{cases} & \text{otherwise,} \end{cases}$$

Figure: Caption

Algorithm

A. flowchart

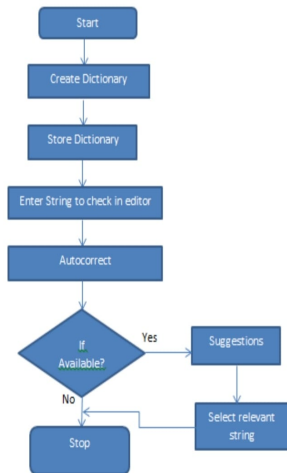


Figure: Caption

Algorithm

		H	Y	U	N	D	A	I
	0	1	2	3	4	5	6	7
H	1	0	1	2	3	4	5	6
O	2	1	1	2	3	4	5	6
N	3	2	2	2	2	3	4	5
D	4	3	3	3	3	2	3	4
A	5	4	4	4	4	3	2	3

Figure: Caption

Applications

Field/Industry	Application of Spell Checker
Writing	Detect and correct spelling errors in documents, emails, and other written communications.
Education	Assist students in improving their spelling skills and provide feedback on written assignments.
Publishing	Ensure the accuracy and consistency of spelling in books, newspapers, and
Medical Records	Ensure accurate spelling of medical terminology in patient records and reports.
Legal Documents	Ensure accuracy in spelling legal terminology in legal documents, contracts, and agreements.

- Multilingual support: Spell checker could be developed to support multiple languages, which would expand its utility in diverse settings.
- Real-time error detection: The spell checker could be designed to detect and highlight errors in real-time, enabling users to make necessary corrections as they type.
- User-customized dictionaries: Users could be provided with the option to create their own customized dictionaries, which could include industry-specific terms or regional slangs, improving the accuracy of the spell checker.

References

a) Data Set of Dictionary of Words:

<https://www.kaggle.com/datasets/therohk/urban-dictionary-words-dataset>

b) GeeksforGeeks for Levenshtein:

<https://www.geeksforgeeks.org/java-program-to-implement-levenshtein-distance-computing-algorithm/>

c) Video to understand Levenshtein Algorithm:

<https://youtu.be/MiqoA-yF-0M>

c) Information about Spell Checker:

https://en.wikipedia.org/wiki/Spell_checker

Software Used

- 1) VS Code
- 2) Kaggle

Conclusion

In conclusion, spell checkers are an essential tool for anyone who wants to improve the accuracy and readability of their writing. They can help users identify and correct spelling errors quickly and easily.

Thank
You!

A close-up photograph of a right hand holding a black marker, positioned to finish writing the word 'You!' in a cursive script. The word 'Thank' is already written above it. A faint 'dreamstime' watermark is visible in the background.

Figure: Caption