Sinhala Auto-correct in Python

Introduction

Make a python program to automatically detect the spelling mistakes and correct it.

Objective is to implement the auto correct algorithm to find the misspelled word and give the most probable correct words list.

How Sinhala word is converting in computer

let's consider a full word like "අමාරු", which means "difficult" in Sinhala. The Unicode points for each character are:

අ (U+0D85)

ම (U+0DB8)

ා (U+0DCB, a vowel sign extending "ā")

ර (U+0DBB)

ු (U+0DD4, a vowel sign extending "u`)

So, when you type "අමාරු," the computer stores it as a sequence of Unicode points:

U+0D85 U+0DB8 U+0DCB U+0DBB U+0DD4

When relevant keys pressed the computer convert to the Unicode points and relevant characters are rendered according to Unicode.

Sinhala Regx for get only matching ones

"sinhala": r"[\u0D80-\u0DFF]+"

* \u0D80 this letter is Unicode starting letter of sinhala letters
* \u0DFF this letter is Unicode ending letter of sinhala letters

