

# Etude des correcteurs orthographiques français

D'après des recherches internet, 3 bibliothèques peuvent être intéressantes, testons les :

```
In [4]: # Importations
from autocorrect import Speller
from spellchecker import SpellChecker
from textblob import TextBlob
```

## Tout d'abord essayons de corriger des mots seuls :

### La bibliothèque Speller

```
In [36]: # Limitation à la langue française
spell = Speller(lang="fr")

# Echantillon de mots mals orthographiés
misspelled = ["ventre", "douleure", "gastrau", "respiratoair", "migrene",
"fievre", "fatige", "diarhée", "eruption", "tou", "essoufflement"]
for word in misspelled:
    print("mot mal orthographié : " + word)
    print("mot corrigé : " + spell(word))
```

```
mot mal orthographié : ventre
mot corrigé : ventre
mot mal orthographié : douleure
mot corrigé : douleur
mot mal orthographié : gastrau
mot corrigé : gastro
mot mal orthographié : respiratoair
mot corrigé : respiratoire
mot mal orthographié : migrene
mot corrigé : migrent
mot mal orthographié : fievre
mot corrigé : fièvre
mot mal orthographié : fatigue
mot corrigé : fatigue
mot mal orthographié : diarhée
mot corrigé : diarrhée
mot mal orthographié : eruption
mot corrigé : éruption
mot mal orthographié : tou
mot corrigé : tou
mot mal orthographié : essoufflement
mot corrigé : essoufflement
```

Sur cet échantillon de mots, on peut voir que la plus part sont corrigés (7 / 11) mais qu'il reste des mots non corrigés (3 / 11) (en particulier les mots où il manque des accents) et qu'il y a même un mot mal corrigé (1 /11)(migrene devient migrent).

## La bibliothèque SpellChecker

```
In [39]: # Limitation à la langue française
spell = SpellChecker(language='fr')

misspelled = ["ventre", "douleure", "gastrau", "respiratoair", "migrene",
"fievre", "fatige", "diarhée", "eruption", "tou", "essoufflement"]
for word in misspelled:
    print("mot mal orthographié : " + word)
    print("mot corrigé : " + spell.correction(word))
```

```
mot mal orthographié : ventre
mot corrigé : ventre
mot mal orthographié : douleure
mot corrigé : douleur
mot mal orthographié : gastrau
mot corrigé : gateau
mot mal orthographié : respiratoair
mot corrigé : respiratoire
mot mal orthographié : migrene
mot corrigé : migrent
mot mal orthographié : fievre
mot corrigé : fièvre
mot mal orthographié : fatige
mot corrigé : fatigue
mot mal orthographié : diarhée
mot corrigé : diarrhée
mot mal orthographié : eruption
mot corrigé : éruption
mot mal orthographié : tou
mot corrigé : tou
mot mal orthographié : essoufflement
```

```
-----
TypeError                                Traceback (most recent call last)
Cell In[39], line 7
      5 for word in misspelled:
      6     print("mot mal orthographié : " + word)
----> 7     print("mot corrigé : " + spell.correction(word))
```

```
TypeError: can only concatenate str (not "NoneType") to str
```

Sur cet échantillon de mots, on peut voir que la plus part sont corrigés (7 / 11) mais qu'il reste des mots non corrigés (1 / 11), qu'il y a même des mots mal corrigés (2 /11) (migrene devient migrent et gastrau devient gateau) et qu'un mot n'est pas reconnu et devient un objet "None" (1 / 11). Contrairement à Speller, il corrige les mots avec accents.

Voici un nouvel algo qui lorsqu'un mot devient un objet "None" car aucune correction n'est trouvé alors on récupère le mot mal orthographié :

```
In [7]: # Limitation à la langue française
spell = SpellChecker(language='fr')

misspelled = ["ventre", "douleure", "gastrau", "respiratoair", "migrene",
               "fievre", "fatige", "diarrhée", "eruption", "tou", "essoufflement"]

for word in misspelled:
    corrected_word = spell.correction(word)
    print("Mot mal orthographié : " + word)
    if corrected_word is not None:
        print("Mot corrigé : " + corrected_word)
    else:
        print("Mot corrigé : " + word)
```

```
Mot mal orthographié : ventre
Mot corrigé : ventre
Mot mal orthographié : douleure
Mot corrigé : douleur
Mot mal orthographié : gastrau
Mot corrigé : gateau
Mot mal orthographié : respiratoair
Mot corrigé : respiratoire
Mot mal orthographié : migrene
Mot corrigé : migrent
Mot mal orthographié : fievre
Mot corrigé : fièvre
Mot mal orthographié : fatige
Mot corrigé : fatigue
Mot mal orthographié : diarrhée
Mot corrigé : diarrhée
Mot mal orthographié : eruption
Mot corrigé : éruption
Mot mal orthographié : tou
Mot corrigé : tou
Mot mal orthographié : essoufflement
Mot corrigé : essoufflement
```

## La bibliothèque TextBlob

```
In [18]: misspelled = ["ventre", "douleure", "gastrau", "respiratoair", "migrene",
    "fievre", "fatige", "diarhée", "eruption", "tou", "essoufflement"]

    for word in misspelled:
        # Créer une instance de TextBlob (une bibliothèque interne doit détecte
        r la langue française)
        blob = TextBlob(word)

        # Obtenir Le mot corrigé
        corrected_word = str(blob.correct())

        # Afficher les résultats
        print("Mot original : " + word)
        if corrected_word != word:
            print("Mot corrigé : " + corrected_word)
        else:
            print("Mot corrigé : " + word)
```

```
Mot original : ventre
Mot corrigé : centre
Mot original : douleure
Mot corrigé : douleurs
Mot original : gastrau
Mot corrigé : gastro
Mot original : respiratoair
Mot corrigé : respiratoair
Mot original : migrene
Mot corrigé : irene
Mot original : fievre
Mot corrigé : pierre
Mot original : fatige
Mot corrigé : fatigue
Mot original : diarhée
Mot corrigé : diarhée
Mot original : eruption
Mot corrigé : eruption
Mot original : tou
Mot corrigé : to
Mot original : essoufflement
Mot corrigé : essoufflement
```

Sur cet échantillon de mots, on peut voir que trop de mots sont mals corrigés ou pas corrigés du tout. Cette bibliothèque n'est donc pas à utiliser en français.

## Essayons maintenant de corriger un texte entier :

### La bibliothèque Speller

```
In [1]: def correct_speller(texte):
    # Créer une instance de Speller
```

```
spell = Speller(lang='fr')

# Corriger l'orthographe dans le texte
texte_corrigé = spell(texte)

return texte_corrigé
```

```
In [6]: # Test
texte = "J'ai une énoorme douleur au ventre je crois que j'ai la gastrau."
texte_corrigé = correct_speller(texte)

print("Texte original:", texte)
print("Texte corrigé:", texte_corrigé)
```

Texte original: J'ai une énoorme douleur au ventre je crois que j'ai la gastrau.

Texte corrigé: J'ai une énorme douleur au ventre je crois que j'ai la gastro.

4 fautes sur 5 ont été corrigées. "Croix" n'a pas été corrigé car le mot "croix" existe. Notre code ne prend pas en compte le contexte.

```
In [8]: # Test
texte = "Les brolures cutanez après la plage nécessitent l'applicassion d'une lotion appésante."
texte_corrigé = correct_speller(texte)

print("Texte original:", texte)
print("Texte corrigé:", texte_corrigé)
```

Texte original: Les brolures cutanez après la plage nécessitent l'applicassion d'une lotion appésante.

Texte corrigé: Les brulures cutane après la plage nécessitent l'application d'une lotion apposant.

La correction est mieux orthographiée que le texte d'origine, en revanche des erreurs sont induites tel que "appésante" devient "apposant".

## La bibliothèque SpellChecker

```
In [2]: def correct_spellChecker(text):
# Créer une instance de SpellChecker
spell = SpellChecker(language='fr')

# Séparer le texte en mots
words = text.split()

# Trouver les mots mal orthographiés
misspelled = spell.unknown(words)

# Correction des mots mal orthographiés
for word in misspelled:
    # Obtenir la suggestion de correction
    correction = spell.correction(word)

    # Vérifier si la correction n'est pas None
    if correction is not None:
        # Remplacer le mot mal orthographié par la suggestion la plus p
        robable
        text = text.replace(word, correction)

    return text
```

```
In [7]: # Test
texte = "J'ai une énoorme douleur au ventre je crois que j'ai la gastrau."
texte_corrige = correct_spellChecker(texte)

print("Texte original:", texte)
print("Texte corrigé:", texte_corrige)
```

Texte original: J'ai une énoorme douleur au ventre je crois que j'ai la gastrau.

Texte corrigé: J'ai une énorme douleur au ventre je crois que j'ai la gastrau.

Nous avons dans cette phrase 3 mots corrigés sur 5. Le mot "crois" n'a pas été corrigé car c'est un mot qui existe. Comme la bibliothèque précédente : Notre code ne prend pas en compte le contexte. Une erreur a été induite : "j'ai" est devenu "jai".

```
In [9]: # Test
texte = "Les brolures cutanez après la plage nécessitent l'applicass ion d'une lotion appésante."
texte_corrige = correct_spellChecker(texte)

print("Texte original:", texte)
print("Texte corrigé:", texte_corrige)
```

Texte original: Les brolures cutanez après la plage nécessitent l'applicass ion d'une lotion appésante.

Texte corrigé: Les brûlures cutanée après la plage nécessitent l'applicassi on dune lotion appésante.

La correction est mieux orthographiée que le texte d'origine, mais n'est pas parfaite.

## Essayons de combiner les deux correcteurs pour améliorer leurs corrections :

```
In [17]: # Corrigeons le texte à l'aide de speller puis à l'aide de spellChecker
texte = "Les brolures cutanez après la plage nécessitent l'applicasson d'une lotion appésante."
texte_corrigé_1 = correct_speller(texte)
texte_corrigé_2 = correct_spellChecker(texte_corrigé_1)

print("Texte original :", texte)
print("Texte corrigé speller :", texte_corrigé_1)
print("Texte corrigé spellChecker :", texte_corrigé_2)
```

Texte original : Les brolures cutanez après la plage nécessitent l'applicasson d'une lotion appésante.  
Texte corrigé speller : Les brulures cutaneaprès la plage nécessitent l'application d'une lotion appasant.  
Texte corrigé spellChecker : Les brulures cutanée après la plage nécessitent application dune lotion opposants

```
In [18]: # Corrigeons le texte à l'aide de spellChecker puis à l'aide de speller
texte = "Les brolures cutanez après la plage nécessitent l'applicasson d'une lotion appésante."
texte_corrigé_1 = correct_spellChecker(texte)
texte_corrigé_2 = correct_speller(texte_corrigé_1)

print("Texte original :", texte)
print("Texte corrigé spellChecker :", texte_corrigé_1)
print("Texte corrigé speller :", texte_corrigé_2)
```

Texte original : Les brolures cutanez après la plage nécessitent l'applicasson d'une lotion appésante.  
Texte corrigé spellChecker : Les brûlures cutanée après la plage nécessitent l'applicasson dune lotion appésante.  
Texte corrigé speller : Les brûlures cutanée après la plage nécessitent l'application dune lotion appasant.

Utiliser plusieurs correcteurs peut dans certains cas améliorer l'orthographe de certains mots mais peut également totalement changer le sens d'un mot.

In [ ]: