

INALCO - Licence LMFA - TNM L3, Algorithmique et structures de données

Caroline Koudoro-Parfait

27 Mars 2023

Nom de l'étudiant :
n° d'étudiant :

Projet

1 Modalités :

Le projet :

- est individuel,
- sera à terme, constitué :
 - d'un programme python commenté dans ses grandes lignes,
 - d'un fichier PDF de 1 à 3 pages répondant aux questions ci-dessous,
- il doit être envoyé par email le 03 avril 2023 au plus tard, à caroline.parfait@outlook.fr.
- en cas de problèmes pour la remise du fichier en parler avant le 03 avril 2023.

2 But du projet

La problématique pour le projet à programmer consiste à réaliser un outil d'évaluation et de comparaison des résultats produits par différents traducteurs automatiques pour un même texte source.

3 Constitution du corpus

Vous réaliserez ce TD à partir du corpus constitué dans les étapes précédentes du projet.

4 Comparaisons Manuelles

Produire les différentes traductions avec les traducteurs automatiques choisis, puis comparer les différentes traductions avec la traduction de référence. Représenter sous forme de tableau les différences et similitudes entre les différentes traductions par rapport à la traduction de référence.

5 Comparaisons automatiques : Similarités et Distances avec scikit-learn

installer scikit-learn :

- <https://pypi.org/project/scikit-learn/>
- <https://scikit-learn.org/stable/install.html>

À partir de la documentation de scikit-learn : <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.DistanceMetric.html>

- Vous utiliserez différentes métriques,
- Vous élaborerez ce programme pour stocker au format json :
 - le nom des textes comparés,
 - les noms des métriques utilisées,
 - les résultats correspondants, et non pas les tableaux de résultats.

```
{ "Nom_fichier-ref -- Nom_fichier-hyp" : {  
  "jaccard": [  
    0.6041933418693982  
  ],  
  "braycurtis": [  
    0.2020727818420559  
  ],  
  "dice": [  
    0.4328632037610366  
  ],  
  "cosinus": [  
    0.026162890836085473  
  ]  
}
```

- Vous représenterez les résultats obtenus sous forme de graphiques. La visualisation des résultats doit permettre de pouvoir identifier quel est le meilleur traducteur automatique selon les résultats des métriques utilisées. Pour ce faire vous pouvez utiliser les librairies :

- scikit-learn (https://scikit-learn.org/stable/auto_examples/release_highlights/plot_release_highlights_0_24_0.html#sphx-glr-auto-examples-rele)
- Matplotlib (<https://python.doctor/page-creer-graphiques-scientifiques-python-apprend>)

6 Analyse

Vous écrirez quelques analyses des résultats obtenus par comparaison automatiques par rapport aux comparaisons manuelles. Vous semblent-ils que les résultats des comparaisons automatiques soient le reflet de ce que vous observez manuellement ?