

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Larbi Ben M'Hidi Oum El Bouaghi

Faculté des sciences exactes et sciences de la nature et de la vie
Département MI
Année : 2024-2025

LE : 13/05/2025
Niveau : M1 IA S2
Session normale

Examen en Traitement du langage naturel

Exercice 01 : (4 pts)

Appliquer les techniques de tokenisation, extraction des entités nommées (NER), Part of Speech tagging (POS) et lemmatisation sur le corpus suivant :

« Apple est une entreprise technologique fondée par Steve Jobs en 1976. Ils sont situés à Cupertino, en Californie. »

Exercice 02 :(7 pts)

Considérez le corpus suivant composé de 3 documents :

"Medical diagnosis is crucial."

"Medical data is confidential."

"Medical treatment requires accurate data"

- 1) Effectuez un prétraitement du corpus
- 2) Construisez un vocabulaire en utilisant la méthode Bag of Words et représentez chaque phrase sous forme d'un vecteur basé sur le vocabulaire construit.
- 3) Construisez la matrice TF-IDF correspondant à ce corpus.

Exercice 03 :(7 pts)

"le professeur explique la leçon. la leçon est difficile. le professeur corrige les copies."

- 1) Listez tous les bigrammes présents dans ce corpus.
- 2) Calculez la fréquence des bigrammes et des unigrammes.
- 3) Calculez les probabilités $P(w_i \mid w_{i-1})$ pour les bigrammes.
- 4) Calculez la probabilité de la phrase : la leçon est difficile

Exercice 04 :(2 pts)

Pour cette phrase : *"Tesla a lancé son nouveau modèle de voiture électrique, la Model Y, lors de la conférence Tesla 2025 à Los Angeles."*

Créez un graphe d'ontologie représentant les relations existantes.

Bonne chance