

Exercice 1 : Tokenisation Manuelle

Texte original : "Les étudiants de Master en informatique explorent les dernières avancées en IA."

Corrigé : ["Les", "étudiants", "de", "Master", "en", "informatique", "explorent", "les", "dernières", "avancées", "en", "IA", "."]

Texte additionnel : "L'IA, c'est-à-dire l'intelligence artificielle, est l'étude des 'agents intelligents'."

Corrigé : ["L'", "IA", ",", "c'est-à-dire", "l'", "intelligence", "artificielle", ",", "est", "l'", "étude", "des", "'", "agents", "intelligents", "'."]

Exercice 2 : Identification et Suppression de Bruit

Texte original : "Pour tout renseignement, écrivez-nous à info@univ-constantine.dz ou visitez le site www.univ-constantine.dz. #Renseignement #Université"

Corrigé : "Pour tout renseignement, écrivez-nous ou visitez le site."

Texte additionnel : "Nouveautés en TALN: voir p. 50-51 du manuel; pour plus, email: taln-expert@academie.fr. #NLP #MasterTALN"

Corrigé : "Nouveautés en TALN: voir pages du manuel; pour plus, contactez-nous."

Exercice 3 : Normalisation Textuelle

Texte original : "L'INTELLIGENCE ARTIFICIELLE révolutionne le monde !"

Corrigé : "l'intelligence artificielle révolutionne le monde !"

Texte additionnel : "M. Dupont, PhD, est spécialisé en TALN: traitement 'automatique' du langage naturel."

Corrigé : "monsieur dupont est spécialisé en traitement automatique du langage naturel."

Exercice 4 : Application Complète de Prétraitement

Texte original : "Dr. Faïd présentera sa recherche sur l'IA au colloque de Constantine. Contact : faid@chercheur.dz"

Corrigé : "docteur faïd présentera sa recherche sur l'intelligence artificielle au colloque de constantine contactez-nous."

Texte additionnel : "Conférence, 16h30: 'Impact de l'IA sur l'éco-système numérique'; inscription via le site-web."

Corrigé : "conférence à seize heures trente impact de l'intelligence artificielle sur l'écosystème numérique inscription via le site web."

Pour chaque exercice, les réponses correctes dépendent des règles et conventions spécifiques établies pour la tokenisation, la normalisation et la suppression du bruit dans votre cours. Les corrigés doivent être adaptés en fonction des directives que vous avez fournies aux étudiants.