

# 1<sup>er</sup> projet pré-entreprise

Sujet : Traitement du langage naturel.

Vous êtes embauché au sein d'une entreprise chargée de d'analyser les tendances actuelles. Le projet en cours est la mise en place d'un système d'analyse des discussions sur un forum. On souhaite savoir si les messages sont positifs ou négatifs . On commence par ajouter via une API un plugin d'analyse des messages afin de permettre de mieux cerner les tendances d'opinions. Vous devrez développer le noyau du plugin : Votre travail consiste à réaliser une application qui prenne en entrée un texte et donne un taux de positivité ou négativité du langage. Votre travail sera fait impérativement via une base de données MongoDB ou Neo4j, celles qui sont installées au sein de l'entreprise.

- Pour commencer votre travail :

**PAS de GÉNÉRATION AUTOMATIQUE DE TEXTES (GAT OU NLG-NATURAL LANGUAGE GENERATION), PAS DE GRAPHEME, PAS DE CHATBOT.**

## A) Définitions

Donnez les définitions des termes suivant :

Traitement du langage naturel  
tokenisation  
racinisation  
corpus  
lemme  
lemmatisation  
thésaurus  
n-gramme  
analyseur lexical  
ontologie  
Wordnet & WOLF & eurowordnet  
Analyse lexicale  
Analyse syntaxique  
Analyse sémantique  
Analyse pragmatique

## B) Algorithme et code source

Proposer un algorithme+code et le décrire étape par étape. Présentez les états de départ, intermédiaire et d'arrivée ainsi que les traitements)

### C) Agrégation de données

L'entreprise vous fourni :

Une liste de termes classifiés/pondérés : [lien](#) et Deux listes de termes triés : [lien](#) et [lien](#)

Les données triées sont à nettoyer et classifier avec le premier dataset.

### D) Champs des rendus

Vous devrez vous cantonner à produire les services suivant :

Analyse de sentiment (texte positif ou texte négatif)\*

La recommandation automatique de la réponse (si vous êtes avancés)

*\*En TALN, les étapes sont en générales, de manière simplifiée:*

1. *L'analyse lexicale*
2. *L'analyse syntaxique*
3. *L'analyse sémantique*
4. *L'analyse pragmatique*

*On souhaite s'arrêter à l'analyse lexicale, votre application devra s'appuyer exclusivement dessus pour son résultat.*

### E) Données à traiter

Les discussions à traiter sont : [lien](#)

### F) Rendu

Le rendu se fera au travers d'un rapport présentant votre solution (et en répondant aux éléments au dessus), votre méthodologie, votre implémentation ainsi que d'une démonstration.

Puis un test sera effectué par l'entreprise sur des échantillons qu'elle vous fournira.