

Practica Unidad 3 -2024 - Procesamiento del Lenguaje

Ejercicio 1

Utilice las librerías `spacy` y `stanza` para realizar un POS sobre los siguientes textos:

Genere una funcion que extraiga la cantidad de sustantivos y los verbos encontrados. Compare los resultados de ambas librerías.

1- "Cuidadoso es mi paso por la senda,
donde el paso del tiempo se hace eterno.
Cada huella que dejo me encomienda,
a un futuro que paso a paso yo gobierno.
"

2- "En la mañana, yo canto con alegría,
el canto del jilguero me acompaña.
Mis notas se elevan, su melodía,
entreteje esperanzas en la montaña.
"

3- "En la quietud de la noche, él espera,
bajo la luz tenue de una vieja farola,
la espera se torna eterna y sincera,
mientras el viento de otoño desconsola
"

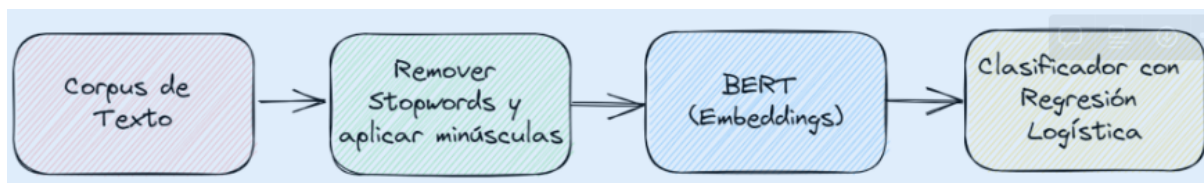
Ejercicio 2

Genere una webscraping de 5 categorías de frases para realizar un dataset a partir de las frases de la pagina <https://proverbia.net/>. Recuede las categorías deben ser balanceadas.

Una vez reunido el dataset entrene un

clasificador utilizando **sklearn** para las categorías seleccionadas con los datos vectorizados. (Pruebe vectorizar con **distintos modelos de embeddings**)

Obtenga las métricas de cada ajuste.



Luego grafique las palabras más influyentes para nuestros clasificador.

Ejercicio 3

Con lo aprendido extraiga el texto de los siguientes archivos:

https://drive.google.com/file/d/182Byc4nATNQpoyxZTLOW0PjSCnceoU_M/view?usp=drive_link

https://drive.google.com/file/d/1tFvGT9VjUIBQykkXilqfuoxoYoH6iAba/view?usp=drive_link

Luego aplique un modelo para extraer las Entidades Nombradas mas relevantes. Observe los resultados. Juegue con el valor de umbral o parametros de cada libreria utilizada. (Si lo cree necesario, realice los tratamientos que considere para mejorar los resultados: stopwords, llevar a minúsculas, etc.)

Posteriormente genere un resumen de cada texto con alguno de los scripts presentados en clase. Si es posible evalúe con distintos modelos y parametrizaciones. Analice los resultados obtenidos.

Ejercicio 4:

Mediante el uso de las técnicas de análisis de sentimientos, explore los textos de reseñas de un producto se encuentran en el siguiente archivo:

https://drive.google.com/file/d/17amue7bWaiUojX0hZwYsVioWPIVwrk8j/view?usp=drive_link

Analice el valor de cada comentario y también genere una descripción de las estadísticas principales del conjunto. Utilice algún análisis gráfico para su resolución.