



UNIVERSIDAD NACIONAL DE ROSARIO

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Tecnicatura Universitaria en Inteligencia Artificial

Procesamiento de Lenguaje Natural - 2025

Comisión 1

TRABAJO PRÁCTICO

INTEGRANTES:

Grimaldi, Damián Daniel

Ejercicio 2

En este ejercicio trabaje con el texto de `Mecanica_del_juego.txt`, cuyo objetivo era fragmentar dicho texto, vectorizarlo y utilizar para el modelo de embeddings, y realizar búsqueda semántica, con una frase que se le pasaba por teclado.

Con el modelo de embedding utilizado fue de SentenceTransformer, que permitiendo realizar una semántica multilingüe. Y con la frase ingresada por teclado, se obtuvo con cual métrica daba el mejor resultado y era por similitud del coseno.

Ejercicio 3

Cargue los modelos de POS y NET, con su respectivo nlp, y obtuve del archivo de mecánicas de juego, donde extraje los sustantivos y los clasifique con net, cuales serian esos sustantivos. Luego calculo todas las similitudes de los sustantivos, y evaluo con una palabra sustantiva, para ver si está ahí y ver cuál es su similitud.

Utilizo la similitud del coseno, ya que es la que mejor metrica me daba.

Ejercicio 4

De la librería langdetect, tome detect, para poder detectar los idiomas que se encontró en todos los archivos que hay en información. Guardando el nombre del archivo, el idioma detectado y su contenido en un dataframe.

Ejercicio 5

Utilice el modelo de s-bert, no como estaba en el apunte directamente, sino que utilice con pipeline, y cargue como el modelo directamente sin procesar, y donde clasifique como comentarios negativos que tuvieran 2 estrellas o menos, 2 como neutro, y 3 o más, como positivo.

Utilice el modelo de embedding, utilizado en el ejercicio 2, como para procesar la consulta, y los comentarios según el sentimiento, y por similitud de coseno, retorne la mejor semejanza con el comentario y sentimiento a buscar.

Ejercicio 6

Para este ejercicio, utilice el modelo de regresión logaritmo, como me daba las mejores métricas para predecir en que clasificación irían. Además de que es mas rápido de entrenar