

Plagiarism-detection-in-documents

Autores

Ana Karla Caballero C-411 Alejandro Camacho C-412

Descripción del problema

Se desea detectar si existe algun porciento de plagio entre dos documentos dados.

Consideraciones tomadas

Se realiza el calculo de plagio en documentos escritos en ingles.

Explicación de la solución

plagio_detector

Este es un sistema de deteccion de plagio que utiliza el modelo BERT (Bidirectional Encoder Representations from Transformers) para calcular la similitud entre dos textos.

Func detect_plagiarism:

Esta función toma dos textos como entrada y devuelve la similitud entre ellos como un porcentaje. Utiliza la función compare_bert_embeddings para calcular la similitud entre los embeddings de los textos.

Func compare_bert_embeddings:

Calcula la similitud coseno entre los embeddings de dos textos. Primero, obtiene los embeddings de cada texto usando la función get_bert_embedding y luego calcula la similitud coseno entre ellos.

Func get_bert_embedding:

Esta función toma un texto como entrada y devuelve su embedding utilizando el modelo BERT. Primero, tokeniza el texto y luego lo pasa a través del modelo BERT para obtener el embedding. El embedding se calcula como la media de los estados ocultos de la última capa del modelo BERT.

Interfaz visual

Este código crea una interfaz gráfica de usuario (GUI) simple para una herramienta de detección de plagio utilizando Tkinter. La GUI consta de:

- Una ventana principal con el título "Plagiarism Detection". Un marco (content_frame) para organizar los elementos de la GUI.

- Dos etiquetas (`text_area1_label` y `text_area2_label`) para indicar las áreas de texto donde los usuarios pueden ingresar los textos a comparar.
- Dos áreas de texto (`text_area1` y `text_area2`) para ingresar los textos.
- Un botón (`button`) que, al ser presionado, debería activar la función `detect_plagiarism` para comparar los textos ingresados. Sin embargo, hay un error en la definición del botón, ya que la función `detect_plagiarism` se llama inmediatamente al definir el botón, en lugar de pasar una referencia a la función que se llamará cuando se presione el botón.
- Una etiqueta (`similarity_score_label`) para mostrar el resultado de la comparación de similitud entre los textos ingresados.

La GUI permite a los usuarios ingresar dos textos y compararlos para detectar plagio. Sin embargo, debido al error mencionado, el botón no funcionará como se espera. Además, la etiqueta para mostrar el resultado de la comparación no se actualiza con el resultado de la comparación, lo que significa que el usuario no puede ver el resultado de la detección de plagio.