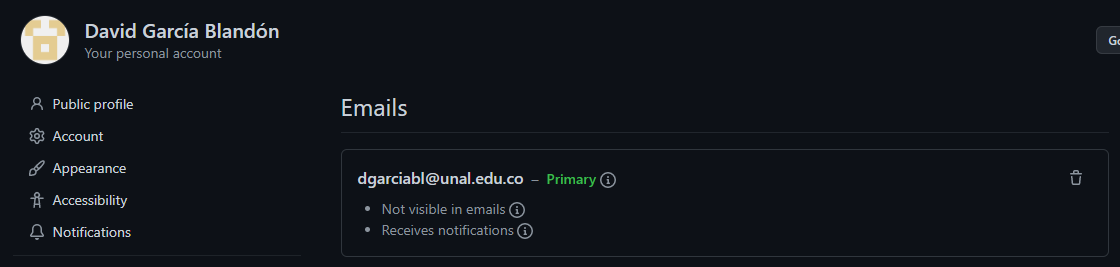
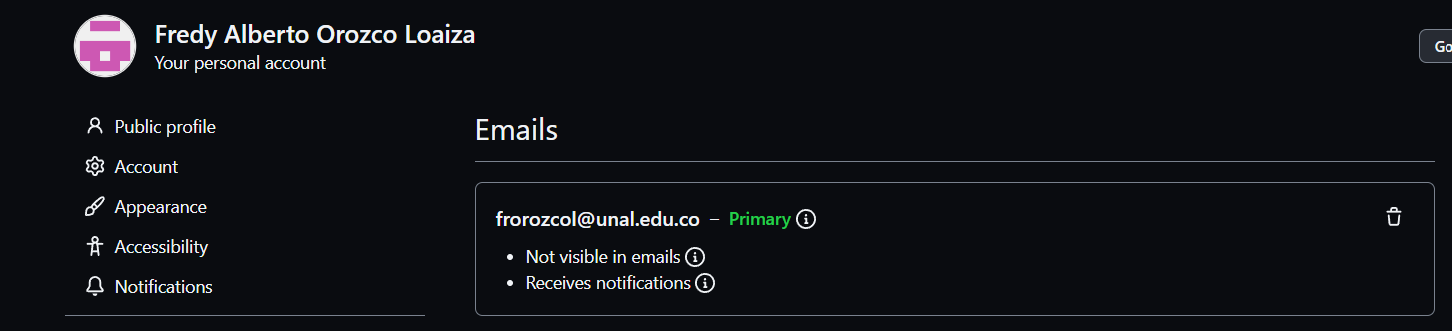
**Integrantes:**

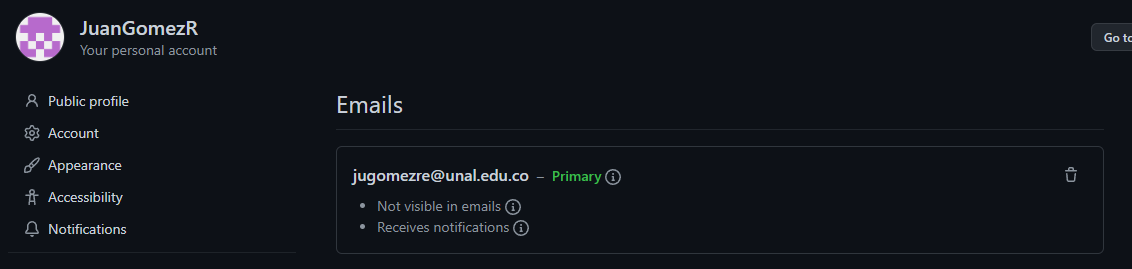
David Garcia Blandon:



[Fredy Alberto Orozco Loaiza](mailto:frorozcol@unal.edu.co)



Juan Pablo Gómez Reyes:



1. **Nombre del proyecto:**

TextIA

1. **Objetivo del proyecto:**

El proyecto TextIA busca desarrollar un aplicativo de procesamiento de lenguaje natural para análisis de texto en artículos de noticias, utilizando modelos pre-entrenados open source para la identificación de sentimientos, temas y entidades, así como para generar resúmenes. El objetivo es ofrecer una herramienta útil y fácil de usar para el análisis y comprensión de textos en lenguaje natural.

1. **Información con la que cuenta para iniciar a trabajar en el proyecto:**

El procesamiento de lenguaje natural (PLN) es una rama de la inteligencia artificial que se enfoca en la interacción entre los humanos y los computadores mediante el uso de lenguaje natural. Su objetivo es permitir a las máquinas entender, interpretar y generar lenguaje humano de manera efectiva. El PLN utiliza técnicas de lingüística computacional, aprendizaje automático y estadística para procesar y analizar textos, voz y otros datos en lenguaje natural. Algunas de sus aplicaciones más comunes incluyen el análisis de sentimiento, el reconocimiento de voz, la traducción automática, el resumen de texto, entre otras.

Análisis de sentimiento: Es una técnica de procesamiento de lenguaje natural que permite determinar la actitud o emoción expresada en un texto. Se utiliza para analizar opiniones en redes sociales, comentarios de clientes en reseñas de productos, entre otros.

Reconocimiento de entidades nombradas: Es una técnica de procesamiento de lenguaje natural que se utiliza para identificar y clasificar diferentes tipos de entidades en un texto, como personas, organizaciones, ubicaciones, fechas, entre otras.

Resumen de texto: Es una técnica de procesamiento de lenguaje natural que consiste en resumir el contenido de un texto en un formato más breve y conciso. El objetivo es preservar la información más importante y relevante del texto original.

Modelado de temas: Es una técnica de procesamiento de lenguaje natural que se utiliza para identificar los temas principales en un conjunto de documentos o textos. El objetivo es agrupar los textos en categorías temáticas para facilitar su análisis y comprensión.

La librería HugginFace cuenta con una gran cantidad de modelos pre-entrenados por la comunidades y por empresas de tecnología, como lo es OpenIA, Facebook, Microsoft, Google entre otras. Cabe resaltar que la misión de HuigginFace es la democratización de la inteligencia artificial, por lo que su contenido es Open Source y también es una librería de alto nivel y muy popular para el desarrollo de PLN. Además, la librería de Streamlit permite crear una interfaz gráfica para que el usuario final pueda ingresar el texto y mostrar los resultados obtenidos por el modelo.

1. **¿Cómo obtendrá el aplicativo la información para operar (entradas o inputs)?**

El usuario ingresará el texto de un artículo de noticias en la interfaz creada con Streamlit.

1. **¿Qué resultados de salida generará el aplicativo cuando sea ejecutado (salidas o outputs)?**

El aplicativo generará una interfaz gráfica que mostrará los siguientes resultados del análisis del texto del artículo de noticias ingresado por el usuario:

* Análisis de sentimiento (positivo, negativo o neutral)
* Identificación de entidades nombradas relevantes en el texto (personas, organizaciones, ubicaciones, fechas, entre otras).
* Generación de un resumen del texto que capture la información más importante y relevante del mismo.
* Modelado de temas para identificar los temas principales en el texto.

1. **¿Qué cálculos o procesos debe realizar el aplicativo para convertir las entradas en salidas?**

* Interfaz gráfica: el aplicativo debe contar con una interfaz gráfica simple y fácil de usar para que el usuario pueda ingresar el texto a analizar.
* Preprocesamiento del texto: se deben realizar una serie de procesos para limpiar y preparar el texto para su análisis, como la eliminación de signos de puntuación, la eliminación de palabras vacías (stopwords) y la normalización de las palabras (por ejemplo, convertir todas las letras a minúsculas).
* Tokenización: se debe dividir el texto en palabras o frases más pequeñas, para poder analizarlas de manera individual.
* Análisis de sentimientos: se deben utilizar modelos pre-entrenados para el análisis de sentimientos, que permitan determinar si el texto tiene una connotación positiva, negativa o neutral.
* Reconocimiento de entidades nombradas: se deben utilizar modelos pre-entrenados para el reconocimiento de entidades nombradas, que permitan identificar y clasificar diferentes tipos de entidades en el texto, como personas, organizaciones, ubicaciones, fechas, entre otras.
* Resumen de texto: se deben utilizar técnicas de resumen de texto para extraer la información más importante y relevante del texto original, y presentarla de manera clara y concisa.
* Modelado de temas: se deben utilizar técnicas de modelado de temas para identificar los temas principales en el texto, y agrupar los textos en categorías temáticas para facilitar su análisis y comprensión.
* Post Procesamiento: se deben procesar los resultados obtenidos para presentarlos de manera clara y concisa en la interfaz gráfica, utilizando gráficos y visualizaciones de datos para facilitar su comprensión.