

REFLEXION SOBRE LOS DESAFÍOS ENCONTRADOS

SELIN COCARCA

1. INTRODUCCIÓN

2. PROBLEMAS INICIALES Y SOLUCIONES APLICADAS

INTRODUCCIÓN:

Este documento reflexiona sobre los desafíos encontrados durante la implementación de un sistema de traducción automática utilizando modelos de la biblioteca Transformers de Hugging Face. Se abordarán los problemas iniciales enfrentados y las soluciones aplicadas para superarlos.

PROBLEMAS INICIALES Y SOLUCIONES APLICADAS:

Inicialmente, uno de los desafíos principales fue la falta de instalación de bibliotecas necesarias para el proyecto. Al intentar ejecutar el código, surgieron errores relacionados con la ausencia de librerías clave. La solución consistió en instalar las bibliotecas requeridas, como transformers y datasets, para asegurar que todas las dependencias estuvieran disponibles y el entorno de trabajo correctamente configurado.

Otro obstáculo significativo surgió al intentar cargar el dataset inicialmente utilizando el conjunto de datos OSCAR con el siguiente código: `dataset = load_dataset("oscar")`. El error reportado indicaba que era necesario especificar una configuración específica para el dataset OSCAR, ya que este tiene múltiples configuraciones disponibles para diferentes idiomas y tareas. La solución inmediata fue cambiar al dataset "facebook/flores", que ofrecía una estructura y configuración más adecuadas para la tarea de traducción automática, evitando así los problemas de carga y configuración inicial.

Posteriormente, al enfrentar la tarea de traducción al euskera, surgieron dificultades con el modelo inicialmente seleccionado, HiTZ/latxa-7b-v1. Este modelo resultó ser demasiado pesado y exigente en términos de memoria RAM, lo que afectó negativamente el rendimiento del sistema. Se intentó cambiar a otro modelo de Facebook diseñado para más idiomas, incluido el euskera, pero también se encontraron limitaciones en su desempeño. La solución definitiva fue optar por el modelo MarianMT, que demostró ser más eficiente y efectivo para la traducción al euskera, asegurando resultados de mejor calidad y evitando problemas de recursos computacionales.

Finalmente, durante la fase de implementación final del código, se observaron problemas con las traducciones al rumano, ya que el modelo no lograba realizar las traducciones correctamente en ninguno de los idiomas. La solución adoptada fue eliminar el rumano del conjunto de idiomas objetivo, enfocándose así en garantizar la precisión y coherencia en las traducciones realizadas entre inglés, español y euskera, los cuales se manejaron de manera efectiva con el modelo MarianMT seleccionado.

Estas soluciones permitieron superar los desafíos técnicos y operativos durante la implementación del sistema de traducción automática, asegurando un entorno de desarrollo estable y resultados consistentes en la tarea de traducción entre múltiples idiomas.