

Diseño de un Modelo grande de lenguaje (*LLM*), para la comprensión de textos

Asignatura:

Redes de Computadores-Grupo 5

Profesor:

Octavio José Salcedo Parra

Integrantes:

Ivan Yesid Sepulveda Paez

Fecha: 31 de agosto de 2023 - Semestre 2023 – II

Conceptos importantes para el desarrollo del proyecto:

Metodología CRISP-DM: La elección de la metodología CRISP-DM para el desarrollo del proyecto es un punto clave, ya que brinda una estructura organizada para la minería de datos y el proceso de desarrollo.

Modelado de Temas: La aplicación de modelado de temas en el procesamiento de texto permite identificar patrones y temas clave en los documentos, lo que es esencial para entender y organizar la información contenida en ellos.

Tecnología de Web Scraping: La extracción automática de contenido web a través del web scraping es esencial para recopilar los datos y documentos necesarios para el proyecto.

Interacción Texto a Texto con LLM: La interacción entre el usuario y el modelo de lenguaje mediante consultas en formato de texto es una característica central del proyecto que permitirá una experiencia de usuario eficaz.

Selección de la técnica de modelado y uso de la metodología CRISP-DM: Uso de ChromaDB y LangChain para procesamiento de texto y consultas. Ajuste de hiperparámetros y configuración del modelo. Evaluación de resultados y revisión del proceso.

Objetivo General:

Mejorar la interacción y accesibilidad de los aspirantes y demás usuarios a la normativa de admisiones estudiantiles de la Universidad Nacional mediante la implementación de un sistema de procesamiento de texto que facilite la búsqueda y obtención de información relevante.

Objetivos Específicos:

Crear un modelo de procesamiento de lenguaje natural utilizando la metodología CRISP-DM.

Implementar el modelado de temas para identificar patrones en los documentos o textos.

Aplicar técnicas de extracción de datos web para recopilar documentos e información.

Establecer un método de interacción texto a texto para obtener respuestas precisas a las consultas de los usuarios.

Evaluar y mejorar continuamente el sistema en función de la retroalimentación de los usuarios y la eficacia en la búsqueda de información.

Referencias

[Application of Large Language Models to Software Engineering Tasks: Opportunities, Risks, and Implications | IEEE Journals & Magazine | IEEE Xplore](#)

Artificial Corner. (n.d.). GPT4All is the local ChatGPT for your documents and it is free. Retrieved from <https://artificialcorner.com/gpt4all-is-the-local-chatgpt-for-your-documents-and-it-is-free-df1016bc335>

[Web Scraping Approaches and their Performance on Modern Websites | IEEE Conference Publication | IEEE Xplore](#)