

Proyecto Final

Entrega 1: Lunes 5 de Junio de 2023

Entrega 2: Jueves 15 de Junio de 2023

Máximo 4 estudiantes por grupo

Intr. Sistemas Inteligentes - 2023-1

Objetivo:

Desarrollar una aplicación que permita interactuar con un conjunto de documentos usando un modelo grande de lenguaje (LLM).

Descripción

Para el desarrollo de la aplicación usaremos LangChain (<https://docs.langchain.com/docs/>). En particular el caso de uso "Question Answering Over Documents" (<https://docs.langchain.com/docs/use-cases/qa-docs>). Cada grupo trabajará sobre un conjunto de datos diferente, el cual debe estar en Español.

Para desarrollar el proyecto tendrán que integrar diferentes tecnologías:

- Herramientas de web scrapping. Herramientas/librerías para extraer información automáticamente de sitios web. La utilizarán para obtener los documentos base.
- Modelos grandes de lenguaje. Se sugiere usar los modelos disponibles en GPT4All (<https://gpt4all.io/index.html>).
- Embeddings. Representación de fragmentos de texto como vectores. Los embeddings están asociados al modelo de lenguaje que utilicen.
- Vector stores. Bases de datos de vectores para almacenar los documentos. Se sugiere usar Chroma (<https://docs.trychroma.com/getting-started>).
- LangChain: Un framework para desarrollar aplicaciones basadas en LLMs.

El proyecto se desarrollará utilizando el lenguaje de programación Python y debe correr sobre un notebook de Colab. Para la documentación usaremos la metodología CRISP-DM.

Entregable 1

Debe entregarse un documento pdf con las siguientes secciones (revisar el documento de la metodología CRISP-DM):

1. Comprensión del negocio. Descripción del problema que queremos resolver desde el punto de vista del cliente/negocio.
2. Comprensión de los datos. Descripción de los datos y sus fuentes.
3. Preparación de datos. Descripción del proceso de web scrapping, del procesamiento de los documentos para extraer el texto y del conjunto de datos resultante. Los datos se deben poner en una carpeta compartida de Google Drive y el documento debe tener la dirección del mismo.

Se debe enviar un archivo pdf con el informe antes del mediodía de la fecha límite a través del siguiente link: <https://www.dropbox.com/request/joy0i2OWqlr7056yLYN6>. El archivo debe nombrarse isi-proj1--usern1-usern2-usern3-usern4.pdf, donde usern es el nombre de usuario asignado por la universidad (incluir los nombres de usuario de todos los miembros del grupo).

Entregable 2

Se debe enviar un archivo zip con los documentos descritos más adelante antes del mediodía de la fecha límite a través del siguiente link: <https://www.dropbox.com/request/Xx0wqGGF371Wy5LgEboO> . El archivo debe nombrarse isi-proj2--usern1-usern2-usern3-usern4.zip, donde usern es el nombre de usuario asignado por la universidad (incluir los nombres de usuario de todos los miembros del grupo).

El archivo zip debe incluir los siguientes archivos:

- **Jupyter Notebook:** con todo el código del proyecto. El notebook debe estar debidamente explicado usando celdas de texto. Todos los pasos de carga, embedding, construcción de la base de vectores, carga del modelo de lenguaje y prueba. Asegúrese antes de hacer el envío de que el notebook esté libre de errores, se guardó correctamente y se visualiza apropiadamente al volver a cargarse.

- **Reporte:** un documento pdf que complete el documento de la entrega 1 con las siguientes secciones adicionales:
 - a. Modelado. Se debe describir de manera detallada cada uno de los componentes de la solución y cómo se integraron.
 - b. Evaluación. Se deben mostrar diferentes ejemplos de interacciones con el sistema y evaluar cualitativamente su desempeño.
 - c. Despliegue. Discutir cómo se podría desplegar el sistema en producción. Opcional hacer una despliegue como aplicación web.
- **Póster:** Un archivo pdf con un póster que presente el proyecto.

Por favor, no incluya imágenes o archivos binarios diferentes a los solicitados.

Adicionalmente se debe preparar un video de máximo 10 minutos. Presentando el proyecto, todos los integrantes del grupo deben participar. La primera parte debe mostrar el las generalidades del problema y cómo se resolvió apoyándose en el póster. La segunda parte debe mostrar la ejecución del código sobre el notebook mostrando y explicando todos los pasos. Se deben mostrar ejemplos de ejecución con diferentes queries y explicar los resultados. El video debe subirse a YouTube y se debe incluir el enlace tanto en el reporte como en el póster. Cerciórese de que el video pueda ser accedido por terceras personas con el link. **No incluya el video en archivo que se debe subir a Dropbox.**

Bibliografía

IBM. "Manual CRISP-DM de IBM SPSS Modeler." *CRISP-DM*, 2012,

https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf . Accessed 13 11

2020.