Prueba Técnica para desarrollador IA:
Manejo de herramienta OCR

La siguiente prueba tiene como objetivo evaluar el manejo de herramientas OCR para la extracción de datos. De esta manera consideraremos si el candidato tiene las aptitudes suficientes para comenzar con el ritmo de trabajo con el que iniciará en el área de AIA Execution de BBVA.

Instrucciones.

Realizar lo siguiente para 3 archivos pdf anexados en el correo.

- Se dará un archivo PDF de una página con una tabla de un Estado Financiero de un cliente.
- Se le proporcionará un archivo .hocr, el cual es un ejemplo de la salida de un motor de OCR al procesar el documento ejemplo.
  Este archivo HOCR tiene un formato similar a html, el cual contiene todo el texto crudo que el motor de OCR detectó. Aquí se incluye no solo el texto, sino también la ubicación de las letras (palabras) con las coordenadas de los bounding boxes y el número de la página.

Su tarea será

- Utilizar lógica y técnicas de programación para reconstruir la tabla con base en este archivo HOCR (el pdf se anexa solo para dar una mejor visualización de lo que se quiere obtener).
- Crear una función en Python que tome de entrada este archivo .hocr, es decir, el path donde está ubicado, y que tenga como salida un archivo csv con la tabla reconstruida.

Siéntase con la libertad de utilizar alguna librería que facilite o simplifique la resolución del problema.

Se evaluarán desde aspectos elementales como buenas prácticas tales como documentación y definición de variables, hasta eficiencia de código, pero principalmente se tomará en cuenta la simplificación y creatividad implementada en la resolución del problema.

Considerar temas de IA también suma puntos.