

Task: Given a text query, retrieve the top1 relevant documents from the document list and the top20 images from the image folder. Use the CLIP model to achieve this task.

Reference:

1. <https://github.com/openai/CLIP>
2. https://huggingface.co/docs/transformers/model_doc/clip

Expected output: generate a prediction csv file, which contains the query ID ("qid", type: str) with the retrieved document ID ("doc_id", type: str) and Top5 images ID ("img_id", type:list). Please use the exact column name as the evaluation script will read these given column names only.

Share a Notebook of your program and the prediction csv file.