Final Project for the
Biomedical Information Retrieval Course

Deadline: the week of Jan 7, 2025 or before

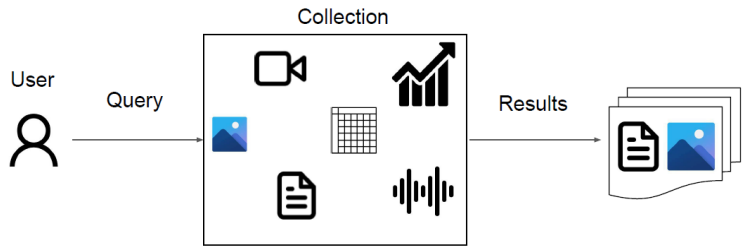
Overview

多模態模型（Multimodal Models）中「多模態」指的是多種資料型態，模型能夠處理並整合來自不同資料型態（例如文字、圖像、聲音等）的資訊，用來理解和擷取跨模態的內容。多模態模型能夠整合多種數據的能力，幫助其應用在更多元複雜的決策場景，能夠應用在製造業、服務業、醫學、自動駕駛...等領域

Implement and analyze the *multi-model information retrieval* based on *word embedding technology* representation. For example, “text” or “image” can be represented as a vector. Each group builds the multi-model IR system as the recipe(食譜) retrieval kernel for smart food search.

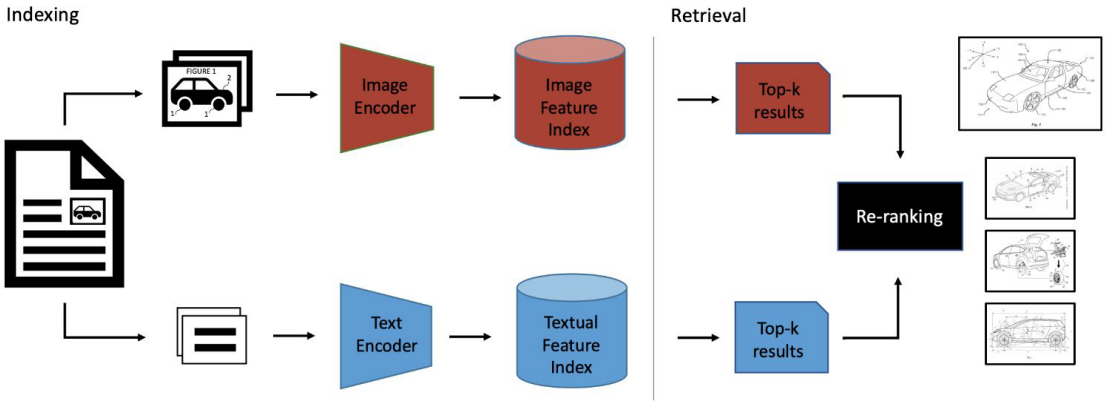
System Description

Following is an example of *multi-model information retrieval*



Multi-model information retrieval approaches:

- Combining search results from individual search systems for each data type, such as text and image
- Retrieval in a unified framework



Focused on using food and cooking images, which are valuable information sources to predict recipe

- For textual information, the text data is obtained from sentence transformer model, such as BERT or RoBERTa(with Pretraining)
- To extract image features, U-Net or Contrastive Language-Image Pre-training (CLIP) can be used
- The individual retrieval results are based on the cosine similarity of vector representations
- To use both of the features, the individual retrieval results are combined using average or maximum scores, etc.