NLP HW4

NCKU F64091130 楊采語

|  |  |  |
| --- | --- | --- |
| 訓練模型 | Running environment | Python version |
| Colab  T4 GPU | Colab  Python3 |

大部分code皆有與Claude討論過再修改

1. describe the details of your implementation for the RAG system
   1. What’s in your RAG system?

|  |  |
| --- | --- |
| Document Database | 150條貓咪知識片段，每條都是獨立的句子 |
| Embeddings Model | 使用 "jinaai/jina-embeddings-v2-base-en"的 HuggingFaceEmbeddings |
| Vector Store | 使用 Chroma 進行文件儲存與檢索 |
| LLM | 透過 Ollama 使用 Llama 3.2 1B 模型 |
| Retriever | Vector store retriever configured with MMR search |
| Chain Components | Question-answering chain combined with retrieval chain |

* 1. Which retrieval model you use?

|  |  |
| --- | --- |
| Main Vector Store | Chroma with MMR (Maximal Marginal Relevance) search strategy |
| Search Configuration | k=1 (number of documents to return)  fetch\_k=5 (number of documents to fetch for MMR)  score\_threshold=0.9 (similarity threshold) |

* 1. What’s your prompt?

system\_prompt = """Answer using exact phrase from context.

Return ONLY the specific value or term needed - no additional words or numbers.

Context: {context}"""

* 1. What’s new in your code in comparison with the code from our lab course in this assignment and list your best score for the ten questions

主要的修改在於Search Configuration的地方，調整了k值，因為原本與助教提供的code一樣是將k設為3，發現模型很容易出現多個答案，所以將k改成1；另外多加了score\_threshold = 0.9的限制，讓模型可以只給出相似度夠高的文本，再下去選擇答案。

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述

(其實答案的意思很相近，但都會缺一些冠詞，但修改提示字後，會變成每個東西都加上奇怪的冠詞等等)

另外在測試第3題的different retrieval models發現，使用相同的prompts，而retrieval models 改成SIMILARITY，也可以取得相對好的成果。( search\_type="similarity", search\_kwargs={"k": 3})，推測是因為MMR可以提供較多樣性的結果，但本次的評估方式只有一個準確的答案，而MMR可能會提供不同不同答案，但也類似的回答(如單位換算、原始答案是數字，但他會轉成文字等等)導致評估出來的分數降低。

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述

1. Please provide analysis for the RAG performance using different prompts
   1. 如果加入 If the exact answer isn't in the context, say "I don't know"，則模型會很容易回答出"I don't know"

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. 如果提示字給他Good answer與Bad answer的format，有時候會學到奇怪的回答格式如下

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

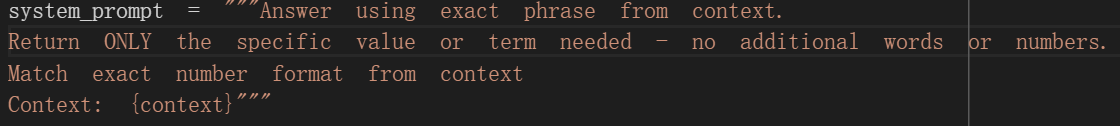
自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. 有時候精簡的提示字反而會得到更正確的回答，但回答格式可能會與正確答案不一樣

如: "bezoar" 應該是 "Bezoar"、"31" 應該是 "31 mph" 或 "49 km"



一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. 複雜的提示字常常把回答搞爛

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. Please compare the RAG performance with different retrieval models and the performance without using RAG

|  |  |  |  |
| --- | --- | --- | --- |
|  | MMR | SIMILARITY | DIRECT\_LLM |
| Accuracy | 7/10 | 9/10 | 2/10 |
| responses | 簡短 | 較完整的答案 | 缺乏精確陳述 |
| format | 某些改寫了原文 | 保持原文表述 | 加入未驗證資訊 |
| Q6(定冠詞) | 會忽略”the” | 會忽略”the” | 有”the” |

|  |  |
| --- | --- |
| Q1: How much of a day do cats spend sleeping on average? | |
| ANSWER | 2/3 |
| MMR | Two-thirds |
| SIMILARITY | 2/3 |
| DIRECT\_LLM | On average, domestic cats spend around 16 to 18 hours per day sleeping. This can range from 12 to 20 hours in some cases, depending on factors such as age, lifestyle, and individual cat behavior.  Cats are known for their love of sleep, and they spend a significant portion of their day snoozing away. Their sleep patterns can vary depending on their environment, health, and other factors, but 16 to 18 hours per day is the average range for domestic cats.  It's worth noting that cats have a unique sleep pattern called "polyphasic" sleep, meaning they take multiple short naps throughout the day rather than one long period of sleep at night. This can help them recharge and adapt to their environment more efficiently. |

|  |  |
| --- | --- |
| Q2: What is the technical term for a cat's hairball? | |
| ANSWER | Bezoar |
| MMR | Bezoar. |
| SIMILARITY | Bezoar. |
| DIRECT\_LLM | The technical term for a cat's hairball is a "trichobezoar." A trichobezoar is a small, hard mass of swallowed hair and other debris that accumulates in the stomach or intestines. It can be a sign of underlying health issues, such as gastrointestinal problems, pancreatitis, or other conditions that affect the digestive system. |

|  |  |
| --- | --- |
| Q3: What do scientists believe caused cats to lose their sweet tooth? | |
| ANSWER | a mutation in a key taste receptor |
| MMR | key taste receptor |
| SIMILARITY | a mutation in a key taste receptor |
| DIRECT\_LLM | There is no scientific evidence to suggest that cats ever had a sweet tooth in the first place. Cats are obligate carnivores, which means they have evolved to thrive on a diet rich in protein and fat from animal sources, rather than carbohydrates.  Cats have a unique physiology that allows them to metabolize fat more efficiently than many other animals, but this doesn't necessarily mean they need or want to eat sweet treats. In fact, cats can be quite picky eaters and may not even enjoy the taste or texture of sweet foods.  It's possible that you may have come across a humorous article or social media post suggesting that cats lost their sweet tooth due to some unknown factor. However, this is likely an exaggeration or a fictional reference rather than a scientific fact.  If you're concerned about your cat's eating habits or overall health, it's always best to consult with a veterinarian who can provide personalized advice and guidance based on your cat's specific needs and circumstances. |

|  |  |
| --- | --- |
| Q6: Which wildcat is considered the ancestor of all domestic cats? | |
| ANSWER | the African Wild Cat |
| MMR | African Wild Cat |
| SIMILARITY | African Wild Cat |
| DIRECT\_LLM | The African wildcat (Felis silvestris) is considered to be the ancestor of all domestic cats. |

1. 為什麼MMR在某些情況下可能不如純相似度檢索?

MMR演算法又叫最大邊界相關演算法，此演算法在設計初是用來計算Query文字與被搜尋文件之間的相似度，然後對文件進行rank排序的演算法。演算法公式如下：

一張含有 文字, 字型, 螢幕擷取畫面, 文件 的圖片

自動產生的描述

Q、C：代表整篇文檔、R：已求得相似度的集合

D：待選文檔

Sim(Q,D)：查詢和文檔的相似度

Sim(D,D')：已選文檔之間的相似度

λ：控制相關性和多樣性的權衡參數(0 ≤ λ ≤ 1)

MMR 背後的想法是，它試圖減少結果中的冗餘並增加多樣性，並用於文字摘要。MMR 根據查詢相關性和資訊新穎性的組合標準在最終關鍵字清單中選擇短語。

所以在選擇過程中，首先會選擇與查詢最相關的文檔，而對於後續文檔，不僅考慮其與查詢的相關性，還會懲罰其與已選文檔的相似度以減少相似的答案，提供更多樣化的檢索結果。但在本次任務中，我們只需求得文本中最精確的答案，使用MMR有時會返回不太精確但相關的答案，導致準確度不如只使用相似度檢索的方法。

參考資料：

[最大边界相关算法MMR(Maximal Marginal Relevance) 实践-CSDN博客](https://blog.csdn.net/levy_cui/article/details/117289694)

[Maximal Marginal Relevance to Re-rank results in Unsupervised KeyPhrase Extraction | by Aditya Kumar | tech-that-works | Medium](https://medium.com/tech-that-works/maximal-marginal-relevance-to-rerank-results-in-unsupervised-keyphrase-extraction-22d95015c7c5)