

20240516Meeting

312707003 黃鈺婷

Retrieval-Augmented Generation(RAG)

- 檢索式模型+生成式模型
- 利用外部知識來增強生成的文本，使生成的文本更加準確
- 不需要重新訓練模型
- 適用於有大量資料，但多數資料未分類或標記
- 開放領域的問答任務、資訊檢索

RAG vs. Fine-tuning

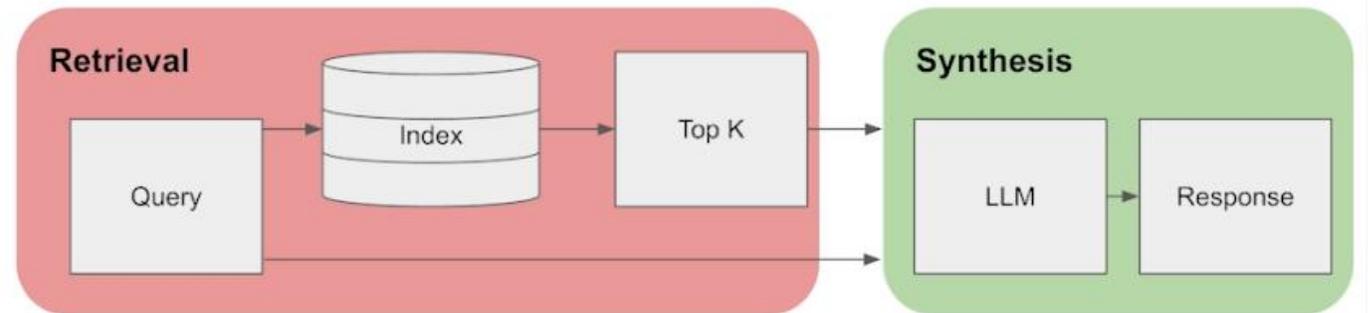
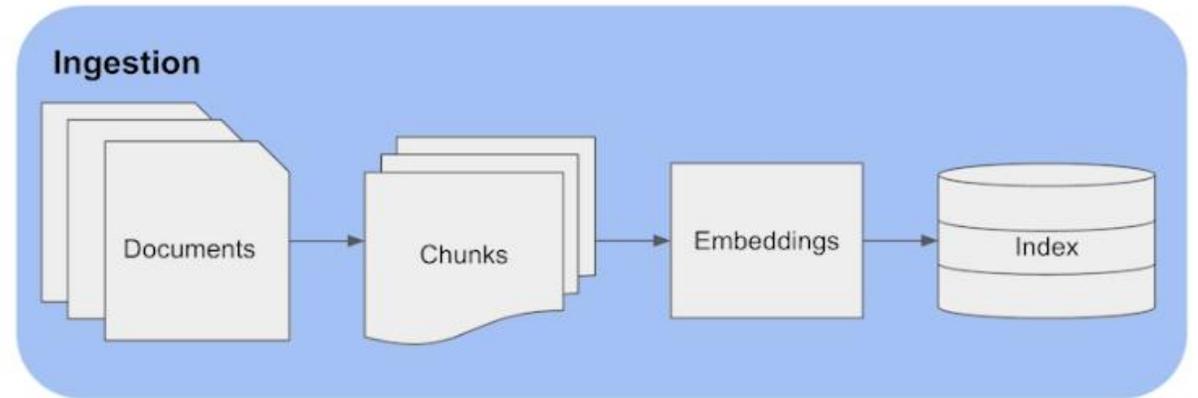
RAG	Fine tuning
直接更新檢索知識庫	需要重新訓練
需要少的數據處理	有限的資料集無法有顯著效能的提升
答案可以追溯到特定的資料來源	像黑盒子，無法知道模型為何以某種方式做出反應
無法自訂模型的行為或風格	允許根據特定語氣或術語調整模型的風格

Neural Retrieval

- Encode the query and documents into dense vector representations
 - compute cosine similarity
 - determine which documents are most relevant to a query
- Advantage:
 - adept at dealing with long and complex queries
- Challenge:
 - performance depends on the data they are trained on

Process of RAG

1. Vector Database Creation
2. User Input
3. Information Retrieval
4. Combining Data
5. Generating Text



Chunking

- Fixed-size chunking
- Context-aware chunking
 1. Sentence Splitting
 - Naïve Splitting
 - NLTK
 - spaCy
 2. Recursive Chunking

Recursive Chunking

```
text_splitter = RecursiveCharacterTextSplitter(  
    separators=[  
        "\n\n",  
        "\n",  
        " ",  
        ".",  
        ",",  
        "\u200b", # Zero-width space  
        "\uff0c", # Fullwidth comma  
        "\u3001", # Ideographic comma  
        "\uff0e", # Fullwidth full stop  
        "\u3002", # Ideographic full stop  
        "" ,  
    ],  
    # Existing args  
)
```

中文判斷順序:

\n\n → \n → " " → , → 、 → . → 。 → ""

Recursive Chunking(設 chunk size=50)

text = ""
第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。

需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。

本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開始及終止之時間。

依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。

哺（集）乳時間，視為工作時間。

員工為撫育未滿三歲子女，得請求下列所定事項之一：

一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。

二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影響其全勤獎金、考績或為其他不利之處分。

第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。）

第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦理。""

中文判斷順序：\n\n → \n → " " → , → 、 → . → 。 → ""

將text先拆成6部分，並判斷每部分的長度是否超過chunk size

第一部分長度為319

第二部分長度為15

第三部分長度為24

第四部分長度為98

第五部分長度為49

第六部分長度為51

再依序處理第一四六部分

Document(page_content=' 哺（集）乳時間，視為工作時間。\\n\\n員工為撫育未滿三歲子女，得請求下列所定事項之一：'),

Document(page_content=' 一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。'),

Document(page_content=' 二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影'),

Document(page_content=' 響其全勤獎金、考績或為其他不利之處分。'),

Document(page_content=' 第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。)'),

Recursive Chunking(設 chunk size=50)

text = ""
第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。

需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。

本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開始及終止之時間。

依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。
子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。

哺（集）乳時間，視為工作時間。

員工為撫育未滿三歲子女，得請求下列所定事項之一：

一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。

二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影響其全勤獎金、考績或為其他不利之處分。

第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。）

第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦理。""

中文判斷順序：\n\n → \n → " " → ， → 、 → . → 。 → ""

處理第六部分

Document(page_content='第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦'),
Document(page_content='理。')]

Recursive Chunking(設 chunk size=50)

text = ""
第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。

需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。

本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開始及終止之時間。

依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。
子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。

哺（集）乳時間，視為工作時間。

員工為撫育未滿三歲子女，得請求下列所定事項之一：

一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。

二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影響其全勤獎金、考績或為其他不利之處分。

第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。）

第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦理。""

中文判斷順序：\n\n → \n → " " → , → 、 → . → 。 → ""

處理第四部分

第一句長度:29

第二句長度:68 → 再繼續處理第二句

Document(page_content='一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。'),

Document(page_content='二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影'),

Document(page_content='響其全勤獎金、考績或為其他不利之處分。'),

Recursive Chunking(設 chunk size=50)

text = ""
第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。

需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。

本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開始及終止之時間。

依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。
子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。

哺（集）乳時間，視為工作時間。

員工為撫育未滿三歲子女，得請求下列所定事項之一：

一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。

二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影響其全勤獎金、考績或為其他不利之處分。

第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。）

第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦理。""

中文判斷順序：\n\n → \n → " " → , → 、 → . → 。 → ""

處理第一部分

第一句長度:40

第二句長度:40

第三句長度:57

第四句長度:90

第五句長度:88

[Document(page_content='第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。'),

Document(page_content='需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。'),

Document(page_content='本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開'),

Document(page_content='始及終止之時間。'),

Document(page_content='依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫'),

Document(page_content='師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。'),

Document(page_content='子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每'),

Document(page_content='日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。'),

Recursive Chunking(設 chunk size=50)

```
text = """第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。  
需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。  
本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開始及終止之時間。  
依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。  
子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。  
  
哺（集）乳時間，視為工作時間。  
  
員工為撫育未滿三歲子女，得請求下列所定事項之一：  
  
一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。  
二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影響其全勤獎金、考績或為其他不利之處分。  
  
第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。）  
  
第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦理。"""
```

全部處理完後，會有15個chunk，每個chunk長度都小於50

```
[Document(page_content='第二十一條：（工作時間）員工正常工作時間每日不超過八小時，每週不得超過四十小時。'),  
Document(page_content='需晝夜輪班工作之各班工作時間、班次更換及非輪班工作之實際工作起訖時間均另行定之。'),  
Document(page_content='本公司得視員工照顧家庭成員需要，允許員工於不變更每日正常工作時數下，在一小時範圍內，彈性調整工作開'),  
Document(page_content='始及終止之時間。'),  
Document(page_content='依職業安全衛生法第二十一條、第二十九條及第三十一條規定，經醫師建議應縮短工作時間者，本公司將參採醫'),  
Document(page_content='師之建議，調整員工之工作時間；其他法規另有規定者，本公司亦將遵守法令規定調整工時。'),  
Document(page_content='子女未滿二歲須員工親自哺（集）乳者，除規定之休息時間外，本公司將每日另給哺（集）乳時間六十分鐘；每'),  
Document(page_content='日正常工作時間以外之延長工作時間達一小時以上者，另給予哺（集）乳時間三十分鐘。'),  
Document(page_content='哺（集）乳時間，視為工作時間。'),  
Document(page_content='員工為撫育未滿三歲子女，得請求下列所定事項之一：'),  
Document(page_content='一、每天減少工作時間一小時；減少之工作時間，不得請求報酬。'),  
Document(page_content='二、調整工作時間。員工為前二項哺（集）乳時間、減少或調整工時之請求時，本公司不得拒絕或視為缺勤而影'),  
Document(page_content='響其全勤獎金、考績或為其他不利之處分。'),  
Document(page_content='第二十一條之一：（工廠員工工作時間）工廠員工上班時間為8：00到17：00。（含休息時間1小時。)'),  
Document(page_content='第二十一條之二：（台北員工工作時間）台北員工之上下班時間及其他規定詳本公司「彈性上下班時間辦法」辦'),  
Document(page_content='理。')] ]
```