

RAG 文本切塊與檢索評估報告

HW Day5 - 學號 1411232019

Q1：固定大小切塊和滑動視窗切塊的參數設定多少？

一、固定大小切塊 (Fixed-size Chunking)

使用 LangChain 的

CharacterTextSplitter，以純字元數進行切割，不考慮語意邊界，可能在句子中間截斷。

參數	設定值	說明
chunk_size	300	每個切塊最多 300 個字元
chunk_overlap	0	切塊之間無重疊
separator	"" (空字串)	不使用分隔符，純粹按字數切割

二、滑動視窗切塊 (Sliding Window Chunking)

使用 LangChain 的 RecursiveCharacterTextSplitter，按照中文語意邊界進行遞迴切割，並透過 overlap 保留上下文。

參數	設定值	說明
chunk_size	300	每個切塊最多 300 個字元
chunk_overlap	100	相鄰切塊重疊 100 個字元
separators	"\n\n", "\n", ". " , "！", "？", "；", "，", "“”，"”	優先在段落、句號等語意邊界處切割

三、語意切塊 (Semantic Chunking)

先按中文句號切分成句子，再透過 Embedding API (維度

4096) 計算相鄰句子的餘弦相似度，當相似度低於門檻時斷開，形成語意段落。過長段落再以 RecursiveCharacterTextSplitter 細切。

參數	設定值	說明
similarity_threshold	0.5	相鄰句子相似度低於此值時斷開
max_chunk_length	500	超過 500 字的語意段落再細切
sub_chunk_size	400	細切時的 chunk_size
sub_chunk_overlap	50	細切時的 overlap

四、各方法切塊數量統計

資料檔案	字元數	固定大小	滑動視窗	語意切塊

data_01.txt	8,038	27	40	46
data_02.txt	2,189	8	11	10
data_03.txt	1,109	4	5	6
data_04.txt	2,178	8	11	16
data_05.txt	5,206	18	28	26
合計	18,720	65	95	104

Q2：哪一種切塊方法效果最好（平均值分數最高）？

一、評分方式說明

本次使用老師提供的評分

API (https://hw-01.wade0426.me/submit_answer) 進行評分。每題將檢索到的 retrieve_text 作為 student_answer 提交，API 回傳 0~1 之間的分數，衡量檢索結果與標準答案的匹配程度。此評分方式較單純的向量相似度更能反映答案品質。

二、各方法平均分數比較

切塊方法	平均分數	切塊總數	排名
滑動視窗 (Sliding Window)	0.702954	95	第 1 名
語意切塊 (Semantic)	0.693921	104	第 2 名
固定大小 (Fixed-size)	0.692383	65	第 3 名

結論：滑動視窗效果最好，平均分數 0.702954。

三、20 題各方法分數明細

Q	題目摘要	固定大小	滑動視窗	語意切塊
1	校正流程分為哪兩種？	0.6792	0.6743	0.7109
2	衛福部長石崇良首波改革方向	0.6772	0.6904	0.7119
3	伊朗2025年大規模抗爭經濟與環境主因	0.6206	0.6274	0.4851
4	「復辟巴勒維王朝」口號背後民意	0.7671	0.7915	0.6846
5	何謂 OSINT？運作方式為何？	0.7173	0.7046	0.7031
6	衛星影像監測油槽可分析什麼情報？	0.5698	0.7559	0.5703
7	美國與海灣盟邦對軍事介入伊朗保留態度	0.6538	0.6099	0.6426
8	二戰後全球核試驗次數與地點特徵	0.6597	0.6851	0.7622
9	守望亞洲計畫功能與應用舉例	0.7891	0.7690	0.7588
10	健保兆元時代面臨哪些挑戰？	0.6851	0.6899	0.7510
11	伊朗2026年1月鎮壓的具體手段	0.7290	0.7280	0.8501
12	福衛九號的技術特色	0.7480	0.7349	0.7891
13	魯尼特圓目前面臨什麼危機？	0.5264	0.5454	0.8086

14	福衛八號自主研發的突破	0.7710	0.6577	0.6157
15	六大皆空是指哪六個科別？	0.6699	0.6646	0.6538
16	為何需自製衛星而非向美國購買？	0.7388	0.7646	0.6392
17	何謂「直美（Chokubi）」現象？	0.6592	0.7593	0.8264
18	鈾-90 與鈍-239 對人體的危害	0.7397	0.7554	0.6377
19	2026年為何是核武威脅關鍵時間點？	0.7764	0.8228	0.5854
20	福衛八號星系對國安監測能力提升	0.6704	0.6284	0.6919

四、各題最佳方法分佈統計

最佳方法	獲勝題數	佔比
語意切塊	9 題	45%
滑動視窗	6 題	30%
固定大小	5 題	25%

五、分析與討論

(1) 滑動視窗為何表現最佳？

滑動視窗使用中文語意邊界（句號、問號等）作為分隔符，搭配 100 字元的 overlap 保留上下文連貫性。這種方法在檢索時能提供較完整的段落資訊，使得提交給評分 API 的答案涵蓋了足夠的上下文。例如 Q19（核武威脅關鍵時間點）滑動視窗得分 0.8228 為三種方法最高，Q4（巴勒維王朝口號）得分 0.7915 也領先其他方法。overlap 機制讓關鍵資訊不會因切割而遺失，是其優勢所在。

(2) 語意切塊的優劣分析

語意切塊平均分數 0.6939 位居第二，與滑動視窗差距約 0.009。語意切塊在部分題目表現突出，例如 Q11（伊朗鎮壓手段）得分 0.8501 為全場最高，Q13（魯尼特圓頂）得分 0.8086 也大幅領先。然而語意切塊在某些題目表現不穩定，例如 Q3（伊朗抗爭主因）僅 0.4851、Q19（核武威脅）僅 0.5854，可能因語意斷點不理想導致關鍵內容被切分至不同段落。這顯示語意切塊的效果高度依賴 threshold 設定與文本結構。

(3) 固定大小的表現

固定大小平均分數 0.6924 位居第三，與語意切塊僅差 0.0015。雖然沒有 overlap 且不考慮語意邊界，但在某些題目反而因為切塊恰好涵蓋了完整的答案段落而表現不錯，例如 Q9（守望亞洲）得分 0.7891 為該題最高。固定大小的優勢在於簡單穩定，不會因為演算法邏輯問題而產生極端低分。

(4) 評分方式的影響

本次使用老師的評分 API（submit_answer）進行評分，而非純向量相似度。API 評分更側重於答案內容是否準確涵蓋標準答案的關鍵資訊，因此切塊內容的完整性與精確度比單純的語意相似度更重要。

這也解釋了為何滑動視窗的 overlap

機制在此評分標準下表現較好——它能保留更多上下文，讓檢索結果包含更完整的答案片段。

六、總結

在本次作業的 5 份資料、20 題檢索任務中，使用評分 API 進行評估，滑動視窗以平均分數 0.702954 表現最佳，其次為語意切塊 (0.693921) ，固定大小 (0.692383) 最低。三種方法的分數差距不大（最大差距僅 0.0106） ，顯示在當前的資料規模與問題設計下，切塊方法的選擇雖有影響，但差異不至於懸殊。滑動視窗因兼顧了語意邊界切割與 overlap

上下文保留，在需要完整答案片段的評分標準下表現最為穩健。