

## 阮荏浩 Day5\_hw

(1) 請告訴我固定大小切塊和滑動視窗切塊你的參數設定是多少？

在本次實驗中，固定大小切塊 (Fixed-size Chunking) 與滑動視窗切塊 (Sliding Window Chunking) 皆採用相同的 chunk 大小設定，以確保比較的公平性。

- 固定大小切塊 (Fixed-size Chunking)

- chunk\_size = 256 tokens
- overlap = 0

此方法將文件依固定長度進行切分，各區塊之間不重疊，實作簡單且計算成本較低。

- 滑動視窗切塊 (Sliding Window Chunking)

- chunk\_size = 256 tokens
- overlap = 48 tokens

此方法在相鄰區塊之間保留部分重疊內容，以降低重要資訊被切斷的風險，提升檢索時上下文的連貫性。

(2) 請簡述哪一種切塊方法效果最好 (平均值分數最高)？

根據本次實驗結果，比較三種切塊方法 (固定大小、滑動視窗、語意切塊) 在 20 題問題上的平均得分，結果如下：

- 語意切塊 (Semantic Chunking)：平均分數最高
- 滑動視窗切塊 (Sliding Window Chunking)：次之
- 固定大小切塊 (Fixed-size Chunking)：相對較低

實驗結果顯示，語意切塊在檢索效果上表現最佳。主要原因在於語意切塊能依據內容語意進行分段，使每個區塊更集中於單一主題，有助於向量檢索模型在 Qdrant 中精準匹配相關段落，進而提升整體檢索與回答品質。