

請回答以下問題（使用 PDF 檔案, 學號_RAG_HW_01.pdf）

1. 請告訴我固定大小切塊和滑動視窗切塊你的參數設定多少？

固定大小切塊:

chunk_size: 500

chunk_overlap: 0

滑動視窗切塊:

chunk_size: 500

chunk_overlap: 250

語意切塊:

SEMANTIC_THRESHOLD = 0.5

2. 請簡述哪一種切塊方法效果最好（平均值分數最高）？

效果最好的切塊方法是滑動視窗切塊，平均是 0.6939，高於固定大小切塊的 0.6841 和語意切塊的 0.6834

以下是執行的結果圖

方法	總分	平均分
方法 : 滑動視窗	13.8779	0.6939
方法 : 固定大小	13.6826	0.6841
方法 : 語意切塊	13.6675	0.6834
表現最好的方法 : 滑動視窗		

固定大小切塊:

```
(vllm-env) shiyi@SY:~/Project/HW/day5$ python 1111132028_RAG_HW_01.py
❖ 開始執行 RAG 作業流程 ❖
=====
🚀 正在執行方法：固定大小 ...
📦 固定大小 總共切出 49 個區塊
• Q1 (固定大小): 分數 0.5806, 來源 data_02.txt
• Q2 (固定大小): 分數 0.7490, 來源 data_03.txt
• Q3 (固定大小): 分數 0.6919, 來源 data_01.txt
• Q4 (固定大小): 分數 0.7305, 來源 data_01.txt
• Q5 (固定大小): 分數 0.6909, 來源 data_05.txt
• Q6 (固定大小): 分數 0.5918, 來源 data_05.txt
• Q7 (固定大小): 分數 0.7061, 來源 data_01.txt
• Q8 (固定大小): 分數 0.6982, 來源 data_04.txt
• Q9 (固定大小): 分數 0.7031, 來源 data_02.txt
• Q10 (固定大小): 分數 0.5874, 來源 data_03.txt
• Q11 (固定大小): 分數 0.8530, 來源 data_01.txt
• Q12 (固定大小): 分數 0.7446, 來源 data_02.txt
• Q13 (固定大小): 分數 0.6699, 來源 data_04.txt
• Q14 (固定大小): 分數 0.7163, 來源 data_02.txt
• Q15 (固定大小): 分數 0.5718, 來源 data_03.txt
• Q16 (固定大小): 分數 0.7354, 來源 data_05.txt
• Q17 (固定大小): 分數 0.6445, 來源 data_03.txt
• Q18 (固定大小): 分數 0.6841, 來源 data_04.txt
• Q19 (固定大小): 分數 0.6157, 來源 data_04.txt
• Q20 (固定大小): 分數 0.7178, 來源 data_02.txt
✓ 固定大小 執行完畢，總得分：13.6826，平均得分：0.6841
```

滑動視窗切塊:

```
🚀 正在執行方法：滑動視窗 ...
📦 滑動視窗 總共切出 64 個區塊
• Q1 (滑動視窗): 分數 0.6685, 來源 data_02.txt
• Q2 (滑動視窗): 分數 0.7285, 來源 data_03.txt
• Q3 (滑動視窗): 分數 0.7012, 來源 data_01.txt
• Q4 (滑動視窗): 分數 0.7358, 來源 data_01.txt
• Q5 (滑動視窗): 分數 0.6909, 來源 data_05.txt
• Q6 (滑動視窗): 分數 0.5947, 來源 data_05.txt
• Q7 (滑動視窗): 分數 0.7285, 來源 data_01.txt
• Q8 (滑動視窗): 分數 0.7051, 來源 data_04.txt
• Q9 (滑動視窗): 分數 0.7925, 來源 data_02.txt
• Q10 (滑動視窗): 分數 0.5366, 來源 data_03.txt
• Q11 (滑動視窗): 分數 0.8530, 來源 data_01.txt
• Q12 (滑動視窗): 分數 0.7822, 來源 data_02.txt
• Q13 (滑動視窗): 分數 0.6763, 來源 data_04.txt
• Q14 (滑動視窗): 分數 0.7720, 來源 data_02.txt
• Q15 (滑動視窗): 分數 0.5811, 來源 data_03.txt
• Q16 (滑動視窗): 分數 0.7261, 來源 data_05.txt
• Q17 (滑動視窗): 分數 0.6377, 來源 data_03.txt
• Q18 (滑動視窗): 分數 0.6880, 來源 data_04.txt
• Q19 (滑動視窗): 分數 0.5928, 來源 data_04.txt
• Q20 (滑動視窗): 分數 0.6865, 來源 data_02.txt
✓ 滑動視窗 執行完畢，總得分：13.8779，平均得分：0.6939
```

語意切塊:

```
🚀 正在執行方法 : 語意切塊 ...
📦 語意切塊 總共切出 82 個區塊
• Q1 (語意切塊): 分數 0.6025, 來源 data_02.txt
• Q2 (語意切塊): 分數 0.7085, 來源 data_03.txt
• Q3 (語意切塊): 分數 0.6064, 來源 data_01.txt
• Q4 (語意切塊): 分數 0.7378, 來源 data_01.txt
• Q5 (語意切塊): 分數 0.6675, 來源 data_05.txt
• Q6 (語意切塊): 分數 0.6152, 來源 data_05.txt
• Q7 (語意切塊): 分數 0.7129, 來源 data_01.txt
• Q8 (語意切塊): 分數 0.7788, 來源 data_04.txt
• Q9 (語意切塊): 分數 0.7085, 來源 data_02.txt
• Q10 (語意切塊): 分數 0.6533, 來源 data_03.txt
• Q11 (語意切塊): 分數 0.8223, 來源 data_01.txt
• Q12 (語意切塊): 分數 0.7495, 來源 data_02.txt
• Q13 (語意切塊): 分數 0.6074, 來源 data_04.txt
• Q14 (語意切塊): 分數 0.6953, 來源 data_05.txt
• Q15 (語意切塊): 分數 0.5962, 來源 data_03.txt
• Q16 (語意切塊): 分數 0.6885, 來源 data_05.txt
• Q17 (語意切塊): 分數 0.6641, 來源 data_03.txt
• Q18 (語意切塊): 分數 0.6343, 來源 data_04.txt
• Q19 (語意切塊): 分數 0.6929, 來源 data_04.txt
• Q20 (語意切塊): 分數 0.7256, 來源 data_05.txt
✓ 語意切塊 執行完畢，總得分：13.6675，平均得分：0.6834
```