

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 字符

chunk\_overlap: 0 字符

滑動視窗切塊：

chunk\_size: 500 字符

chunk\_overlap: 250 字符(Step 為 250)

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，總分是 13.8706，高於固定大小切塊的 13.7256 和語意切塊的 12.1182

以下是執行的結果圖

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 ...
• Q1 (固定大小): 分數 0.5063, 來源 data_02.txt
• Q2 (固定大小): 分數 0.6904, 來源 data_03.txt
• Q3 (固定大小): 分數 0.6274, 來源 data_01.txt
• Q4 (固定大小): 分數 0.7593, 來源 data_01.txt
• Q5 (固定大小): 分數 0.7056, 來源 data_05.txt
• Q6 (固定大小): 分數 0.5952, 來源 data_05.txt
• Q7 (固定大小): 分數 0.6665, 來源 data_01.txt
• Q8 (固定大小): 分數 0.7002, 來源 data_04.txt
• Q9 (固定大小): 分數 0.6646, 來源 data_02.txt
• Q10 (固定大小): 分數 0.6899, 來源 data_03.txt
• Q11 (固定大小): 分數 0.8291, 來源 data_01.txt
• Q12 (固定大小): 分數 0.7407, 來源 data_02.txt
• Q13 (固定大小): 分數 0.7197, 來源 data_04.txt
• Q14 (固定大小): 分數 0.7217, 來源 data_02.txt
• Q15 (固定大小): 分數 0.6606, 來源 data_03.txt
• Q16 (固定大小): 分數 0.7456, 來源 data_05.txt
• Q17 (固定大小): 分數 0.7139, 來源 data_03.txt
• Q18 (固定大小): 分數 0.7422, 來源 data_04.txt
• Q19 (固定大小): 分數 0.5781, 來源 data_04.txt
• Q20 (固定大小): 分數 0.6685, 來源 data_02.txt
✓ 固定大小 執行完畢，總得分： 13.7256
```

滑動視窗切塊:

```
🚀 正在執行方法 : 滑動視窗 ...
• Q1 (滑動視窗): 分數 0.6846, 來源 data_02.txt
• Q2 (滑動視窗): 分數 0.6836, 來源 data_03.txt
• Q3 (滑動視窗): 分數 0.6084, 來源 data_01.txt
• Q4 (滑動視窗): 分數 0.7119, 來源 data_01.txt
• Q5 (滑動視窗): 分數 0.7056, 來源 data_05.txt
• Q6 (滑動視窗): 分數 0.5952, 來源 data_05.txt
• Q7 (滑動視窗): 分數 0.6885, 來源 data_01.txt
• Q8 (滑動視窗): 分數 0.7002, 來源 data_04.txt
• Q9 (滑動視窗): 分數 0.8018, 來源 data_02.txt
• Q10 (滑動視窗): 分數 0.5630, 來源 data_03.txt
• Q11 (滑動視窗): 分數 0.8291, 來源 data_01.txt
• Q12 (滑動視窗): 分數 0.7827, 來源 data_02.txt
• Q13 (滑動視窗): 分數 0.7197, 來源 data_04.txt
• Q14 (滑動視窗): 分數 0.7510, 來源 data_02.txt
• Q15 (滑動視窗): 分數 0.6606, 來源 data_03.txt
• Q16 (滑動視窗): 分數 0.7500, 來源 data_05.txt
• Q17 (滑動視窗): 分數 0.6289, 來源 data_03.txt
• Q18 (滑動視窗): 分數 0.7422, 來源 data_04.txt
• Q19 (滑動視窗): 分數 0.5781, 來源 data_04.txt
• Q20 (滑動視窗): 分數 0.6855, 來源 data_02.txt
✓ 滑動視窗 執行完畢，總得分： 13.8706
```

語意切塊:

```
🚀 正在執行方法 : 語意切塊 ...
• Q1 (語意切塊): 分數 0.5596, 來源 data_02.txt
• Q2 (語意切塊): 分數 0.6934, 來源 data_03.txt
• Q3 (語意切塊): 分數 0.6509, 來源 data_01.txt
• Q4 (語意切塊): 分數 0.6387, 來源 data_01.txt
• Q5 (語意切塊): 分數 0.5205, 來源 data_05.txt
• Q6 (語意切塊): 分數 0.5527, 來源 data_05.txt
• Q7 (語意切塊): 分數 0.6694, 來源 data_01.txt
• Q8 (語意切塊): 分數 0.7119, 來源 data_04.txt
• Q9 (語意切塊): 分數 0.6499, 來源 data_02.txt
• Q10 (語意切塊): 分數 0.4888, 來源 data_03.txt
• Q11 (語意切塊): 分數 0.6953, 來源 data_01.txt
• Q12 (語意切塊): 分數 0.6255, 來源 data_02.txt
• Q13 (語意切塊): 分數 0.4800, 來源 data_04.txt
• Q14 (語意切塊): 分數 0.6201, 來源 data_02.txt
• Q15 (語意切塊): 分數 0.5757, 來源 data_03.txt
• Q16 (語意切塊): 分數 0.6387, 來源 data_05.txt
• Q17 (語意切塊): 分數 0.6030, 來源 data_03.txt
• Q18 (語意切塊): 分數 0.5625, 來源 data_04.txt
• Q19 (語意切塊): 分數 0.4751, 來源 data_04.txt
• Q20 (語意切塊): 分數 0.7065, 來源 data_05.txt
✓ 語意切塊 執行完畢，總得分： 12.1182
```