

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 (Fixed-size Chunking)**
 - **Chunk Size (區塊大小): 400 字元**
 - **Chunk Overlap (重疊大小): 50 字元**
- **滑動視窗切塊 (Sliding Window Chunking)**
 - **Chunk Size (區塊大小): 300 Tokens**
 - **Chunk Overlap (重疊大小): 30 Tokens**

2、 請簡述哪一種切塊方法效果最好（平均值分數最高）？

但目前實際做出來是『固定大小』分數最高。設定的 Chunk Size (400 字) 恰好涵蓋了大部分問題所需的上下文長度。使用的 GPT-4o 模型具備強大的理解與容錯能力，即使切塊邊界不完美，仍能從中提取正確答案。資料文本結構較鬆散，使得複雜的語意分割優勢不明顯，反而是簡單的覆蓋策略更有效。」

```
=== 各方法平均分數 ===  
method  
fixed_size      0.742801  
semantic        0.719385  
sliding_window  0.718680  
Name: score, dtype: float64  
  
* 表現最好的方法是：fixed_size（平均分：0.74）
```

* 執行方法: *fixed_size* (Chunk數: 62)

正在將文字轉換為向量...

[Q1] 分數: 0.85791015625, 來源: data_02.txt

[Q2] 分數: 0.7197265625, 來源: data_03.txt

[Q3] 分數: 0.65771484375, 來源: data_01.txt

[Q4] 分數: 0.8349609375, 來源: data_01.txt

[Q5] 分數: 0.81298828125, 來源: data_05.txt

[Q6] 分數: 0.75390625, 來源: data_05.txt

[Q7] 分數: 0.72802734375, 來源: data_01.txt

[Q8] 分數: 0.86083984375, 來源: data_04.txt

[Q9] 分數: 0.7900390625, 來源: data_02.txt

[Q10] 分數: 0.74951171875, 來源: data_03.txt

[Q11] 分數: 0.09918212890625, 來源: data_01.txt

[Q12] 分數: 0.67919921875, 來源: data_02.txt

[Q13] 分數: 0.76025390625, 來源: data_04.txt

[Q14] 分數: 0.927734375, 來源: data_02.txt

[Q15] 分數: 0.62158203125, 來源: data_03.txt

[Q16] 分數: 0.783203125, 來源: data_05.txt

[Q17] 分數: 0.7529296875, 來源: data_03.txt

[Q18] 分數: 0.8701171875, 來源: data_04.txt

[Q19] 分數: 0.84619140625, 來源: data_04.txt

[Q20] 分數: 0.75, 來源: data_05.txt

* 執行方法: *sliding_window* (Chunk數: 99)

正在將文字轉換為向量...

[Q1] 分數: 0.6845703125, 來源: data_02.txt

[Q2] 分數: 0.7236328125, 來源: data_03.txt

[Q3] 分數: 0.65771484375, 來源: data_01.txt

[Q4] 分數: 0.85888671875, 來源: data_01.txt

[Q5] 分數: 0.80126953125, 來源: data_05.txt

[Q6] 分數: 0.71630859375, 來源: data_02.txt

[Q7] 分數: 0.673828125, 來源: data_01.txt

[Q8] 分數: 0.81640625, 來源: data_04.txt

[Q9] 分數: 0.6923828125, 來源: data_02.txt

[Q10] 分數: 0.77587890625, 來源: data_03.txt

[Q11] 分數: 0.09918212890625, 來源: data_01.txt

[Q12] 分數: 0.83984375, 來源: data_02.txt

[Q13] 分數: 0.79296875, 來源: data_04.txt

[Q14] 分數: 0.6962890625, 來源: data_05.txt

[Q15] 分數: 0.62109375, 來源: data_03.txt

[Q16] 分數: 0.82373046875, 來源: data_05.txt

[Q17] 分數: 0.78369140625, 來源: data_03.txt

[Q18] 分數: 0.8046875, 來源: data_04.txt

[Q19] 分數: 0.81005859375, 來源: data_04.txt

[Q20] 分數: 0.701171875, 來源: data_02.txt

* 執行方法: *semantic* (Chunk數: 61)

正在將文字轉換為向量...

[Q1] 分數: 0.1376953125, 來源: data_05.txt

[Q2] 分數: 0.72216796875, 來源: data_03.txt

[Q3] 分數: 0.62548828125, 來源: data_01.txt

[Q4] 分數: 0.8330078125, 來源: data_01.txt

[Q5] 分數: 0.83251953125, 來源: data_05.txt

[Q6] 分數: 0.751953125, 來源: data_05.txt

[Q7] 分數: 0.642578125, 來源: data_01.txt

[Q8] 分數: 0.873046875, 來源: data_04.txt

[Q9] 分數: 0.80078125, 來源: data_02.txt

[Q10] 分數: 0.75, 來源: data_03.txt

[Q11] 分數: 0.6484375, 來源: data_01.txt

[Q12] 分數: 0.67919921875, 來源: data_02.txt

[Q13] 分數: 0.7880859375, 來源: data_04.txt

[Q14] 分數: 0.7021484375, 來源: data_05.txt

[Q15] 分數: 0.62158203125, 來源: data_03.txt

[Q16] 分數: 0.79052734375, 來源: data_05.txt

[Q17] 分數: 0.7783203125, 來源: data_03.txt

[Q18] 分數: 0.8251953125, 來源: data_04.txt

[Q19] 分數: 0.849609375, 來源: data_04.txt

[Q20] 分數: 0.7353515625, 來源: data_05.txt