

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？

| 檔案 | 檔案長度 (Chars) | 文本 類型 | Character Split | Token Split | Semantic Split |
|-------------|-----------------|----------|--------------------|----------------------------|-----------------------|
| data_01.txt | 8,038 | 長文 本 | Size: 500 | Size: 500 (Overlap 100) | Min: 200, Max: 400 |
| data_02.txt | 2,189 | 較長 文本 | Size: 300 | Size: 300 (Overlap 75) | Min: 150, Max: 300 |
| data_03.txt | 1,109 | 中等 文本 | Size: 200 | Size: 200 (Overlap 50) | Min: 100, Max: 200 |
| data_04.txt | 2,178 | 較長 文本 | Size: 300 | Size: 300 (Overlap 75) | Min: 150, Max: 300 |
| data_05.txt | 5,206 | 長文 本 | Size: 500 | Size: 500 (Overlap 100) | Min: 200, Max: 400 |

參數說明 (由 get_dynamic_split_params 決定)

這些參數是依據字數門檻動態決定的：

1. **長文本** (≥ 5000) (如 `data_01` , `data_05`)
 - Character: 設定為 500，區塊最大。
 - Token: 設定為 500 (Overlap 100)，保留最多上下文。
 - Semantic: 範圍在 200~400 之間，彈性最大。
2. **較長文本** (2000 ~ 5000) (如 `data_02` , `data_04`)
 - Character: 設定為 300。
 - Token: 設定為 300 (Overlap 75)。
 - Semantic: 範圍在 150~300 之間。
3. **中等文本** (500 ~ 2000) (如 `data_03`)
 - Character: 設定為 200。
 - Token: 設定為 200 (Overlap 50)。
 - Semantic: 範圍在 100~200 之間，區塊較小以求精確。

2. 請簡述哪一種切塊方法效果最好（平均值分數最高）？

第三種方法語意分塊，平均分數為:075775

固定切分:0.7316

滑動切分:0.7396