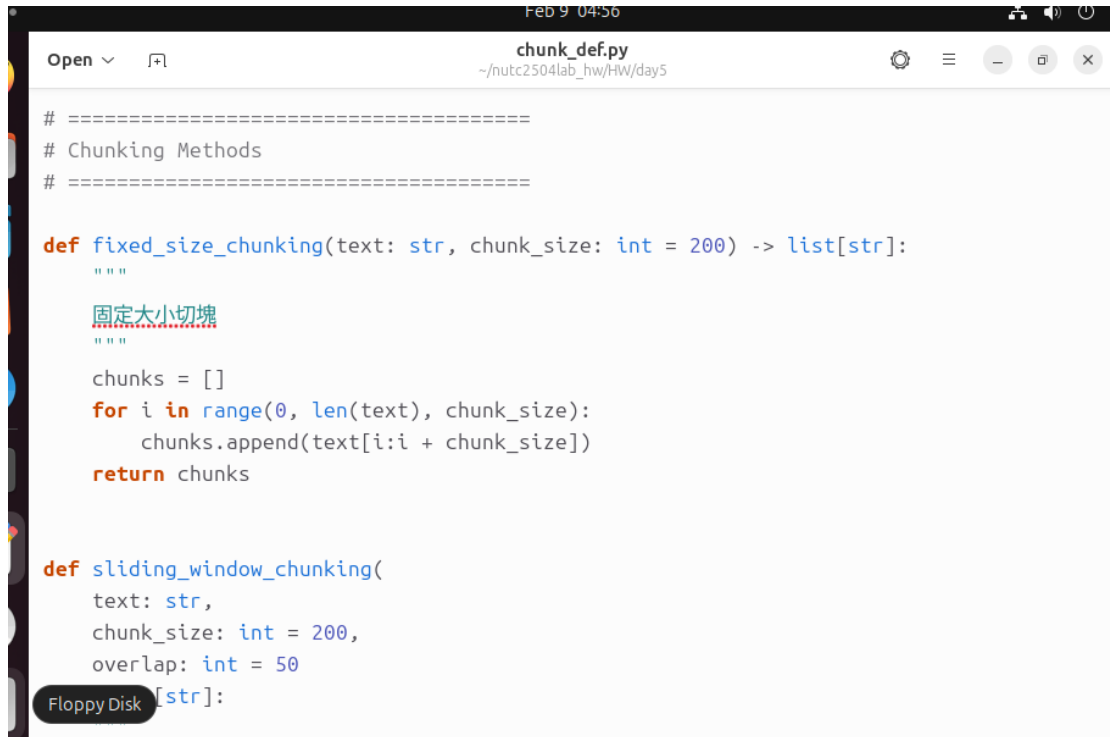


1. 請告訴我固定大小切塊和滑動視窗切塊你的參數設定多少？



```
Feb 9 04:56
chunk_def.py
~/nutc2504lab_hw/HW/day5

# =====
# Chunking Methods
# =====

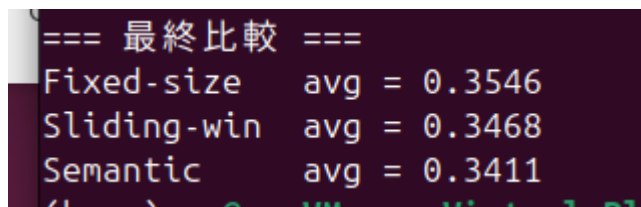
def fixed_size_chunking(text: str, chunk_size: int = 200) -> list[str]:
    """
    固定大小切塊
    """
    chunks = []
    for i in range(0, len(text), chunk_size):
        chunks.append(text[i:i + chunk_size])
    return chunks

def sliding_window_chunking(
    text: str,
    chunk_size: int = 200,
    overlap: int = 50
) -> list[str]:
```

我將固定大小切塊和滑動視窗切塊的 `chunk_size` 設為 200，這樣可以同時保留足夠語意，又不會讓 `chunk` 過長而影響 `embedding` 與召回效果。

而 `overlap` 設為 50，可以避免重要資訊剛好被切斷，提升滑動視窗切塊的語意連續性。

2. 請簡述哪一種切塊方法效果最好（平均值分數最高）？



```
=== 最終比較 ===
Fixed-size    avg = 0.3546
Sliding-win   avg = 0.3468
Semantic      avg = 0.3411
```

用固定大小的切塊方式效果最好，平均值是 0.3546。