多執行緒簡易網頁爬蟲報告

**本報告說明本次作業之實作內容與原理。**

# 作業說明

本程式使用 Python 撰寫，目的是實作一個簡單的多執行緒爬蟲。

本程式完全由本人原創，無抄襲、使用 AI 協助。

程式使用 Python 內建 threading 套件搭配 requests 套件來實作多執行緒的網頁下載功能。

# 原理說明

爬蟲是一種自動化工具，用來下載網頁資料。本程式中，先定義一組網址清單，接著為每個網址開一條執行緒，並使用 requests.get() 下載該頁內容，最後統一等待所有執行緒完成。這樣可以同時爬取多個網頁，效率比逐一處理更高。

# 使用方式

1. 確保已安裝 Python 和 requests 套件。

2. 將程式碼儲存為 crawler.py。

3. 在命令列執行： python crawler.py

# 程式碼

import threading  
import requests  
  
urls = [  
 "https://example.com",  
 "https://openai.com",  
 "https://www.python.org",  
 "https://www.wikipedia.org",  
 "https://github.com"  
]  
  
def fetch(url):  
 try:  
 res = requests.get(url, timeout=10)  
 print(f"{url} 下載成功，大小：{len(res.text)} bytes")  
 except Exception as e:  
 print(f"{url} 下載失敗：{e}")  
  
threads = []  
for url in urls:  
 t = threading.Thread(target=fetch, args=(url,))  
 threads.append(t)  
 t.start()  
  
for t in threads:  
 t.join()  
  
print("全部下載完成。")