CE6155 Homework1 Report

                         student ID：112522087       name：黃懷萱

Can answer in Chinese or English(可以使用中文或英文回答)

1. (15%) Please record the performance statistics of the web crawler, for example: how long does it take to crawl 100 web pages, how long does it take to crawl 500 web pages? Then, how do you speed up the process, and what is the resulting performance after speeding up?

在進行網頁爬蟲的過程中，我使用了兩種不同的方法來搜尋網站：動態查詢（利用Selenium）和靜態查詢（利用BeautifulSoup）。最初，當我嘗試搜尋「數學系」相關的網站時，我注意到僅使用靜態查詢會遇到一個問題：網頁內容因為沒有等待JavaScript渲染完成而導致不完整，這使得網頁看起來內容一致。為了解決這個問題，我起初選擇了動態查詢的方法。

隨後，當我將搜索深度調整為2時，我發現許多數學系的網站存在著404錯誤，顯示未維護。這促使我轉向另一個搜索目標：「學習與教學研究所」。在這一階段，我發現目標網站是靜態網站，且內容已經完成渲染。因此，我決定改用BeautifulSoup進行查詢，這大大提高了查詢速度。

1. (15%) Based on the URLs you have crawled, analyze the structure and layout of the website, and draw an approximate sitemap of the website.
2. (20%) Please compare the differences in query results using different query methods. Please attempt to use embedding models for vector search.

