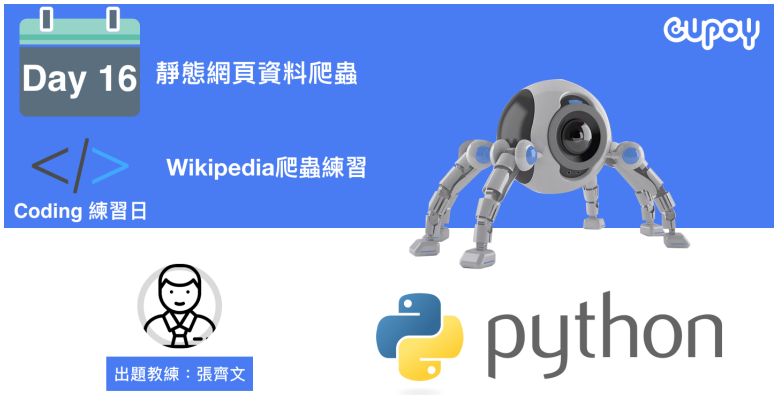
D16：實作 Day：Wiki的爬蟲實作練習

簡報閱讀 範例與作業 問題討論

Wikipedia爬蟲練習

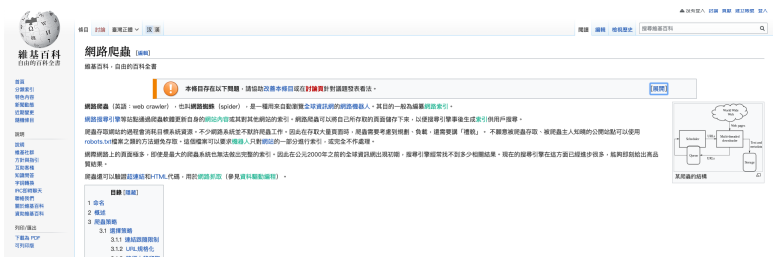


出題教練：張齊文

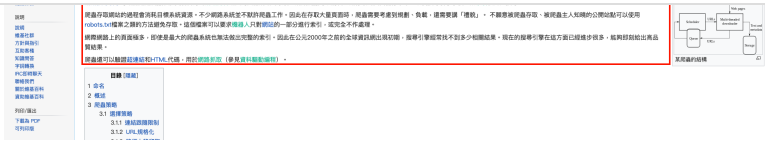
python

重要知識點

Wikipedia(維基百科)爬蟲

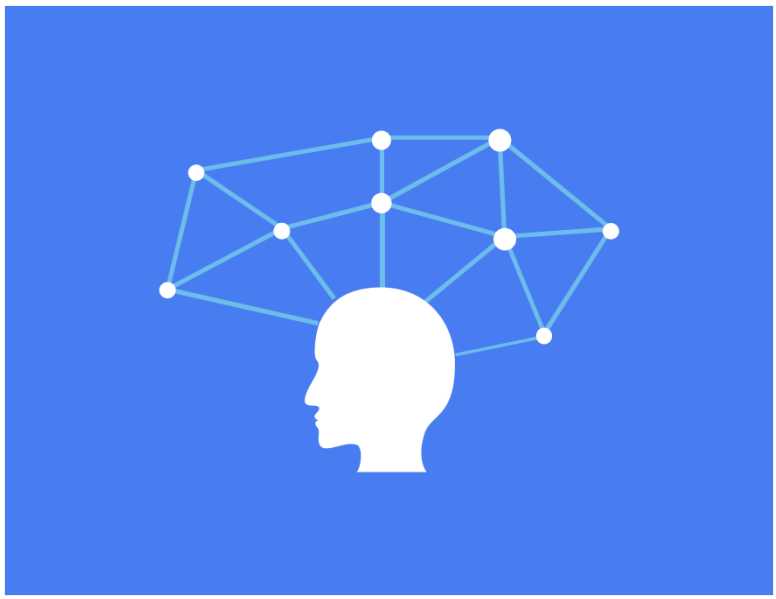


範例1：選定一個關鍵字，爬取該關鍵字的文章內容。



範例2：擷取文章中，延伸出的外部連結關鍵字。

重要知識點複習



練習：定義一個爬蟲函數，重複前面兩個步驟的流程，可遞迴爬取更多關鍵字的解釋文章。其流程如下：

- 爬取當前關鍵字的解釋，並存入檔案(因為文章內容太多會佔滿整個頁面，所以存程檔案，方便後續檢視)。
- 萃取出當前關鍵字所引用的外部連結，當作新的查詢關鍵字。

解題時間

下一步：閱讀範例與完成作業