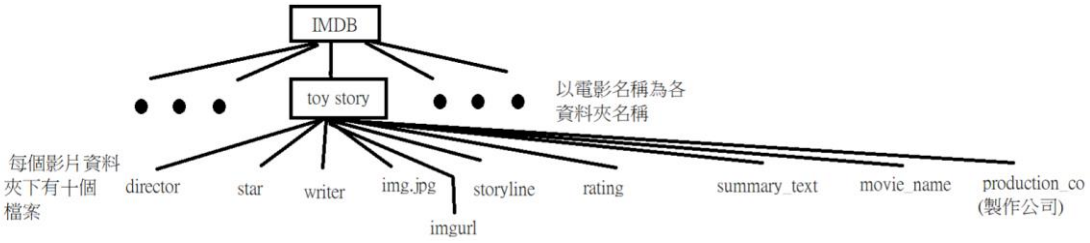
*資料來源: IMDB
https://www.imdb.com/

*使用工具:
python
requests
pyquery

*爬取檔案架構:

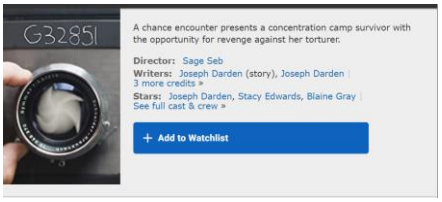



*流程
分為兩個檔案
一個檔案處理大多數的三種紀錄格式，並將未取的的電影資訊輸出成 NonList 後傳給另一份檔案進行特殊處理 (比如格式特殊、缺少資料 ….)

首先先載入要爬入的電影資訊清單 (MovieLens 1M dataset)
擷取電影名稱

接著分析 IMDB 的 URL 搜尋時的 pattern，然後利用 requests.get 取得電影頁面資料

大部分情況下可依 img 所在的位置判斷是主流格式的哪一種
比如:


然後就可以爬取該頁資訊
並當沒有爬取到 img 或其他資料時，輸出到 NonList 中，以及印出資訊通知使用者