Project – Final Documentation

Project motivation:

之前在升大一的暑假，為了升級電腦配備花了很多功夫，光是一個CPU就有價位、效能和相容的規格等許多必須研究的細項，之後還要去爬文尋找這個CPU的評價如何，是否有什麼災情。何況一台電腦並不是只有CPU，每個組件這樣精挑細選下來就得花上更多時間。所以我想要為正在挑選配備的人打造一個能輕鬆蒐集並整理組件的販售資訊與相關評價的程式。

Plan description:

那首先，必須先獲取市場上大部分零件的資訊，所以呢，第一步就是先用爬蟲的方式從原價屋的網頁獲取商品的種類、價格、廠牌與規格型號等等，基本上只要在原價屋估價頁面看的到的都先抓下來。

再者是關於評價的部分，我個人在需要查找這類評價時都會優先從PTT的PC\_shopping版下手，所以會用到其他人做好放在Github上的PyPtt這個API套件。主要目標是從PTT版上取下相關的討論、評價或是開箱文，可能只做到標題相關，能否內文相關還不確定。

最後，由於原價屋不會附上一些較細節的規格，必須去官網才找的到，而且有時只參考PTT版上的討論是不夠的，最後會導入Google custom search這項API提供使用者更多元豐富的資訊來源，大概顯示5~10項結果與其URL，預期可能的內容有產品官網、其他的購物網站或來自PTT以外的評價文章。

User interaction:

而在程式使用上呢，先要求使用者輸入想查找的商品類型及型號，並從原價屋取出並顯示所有販售中的品項，接著可再輸入指令在原價屋資訊、PTT相關討論或google搜尋結果三者間切換。在搜尋產品時的輸入必須能無視大小寫，可能會規定特殊的輸入格式，若查無此項商品必須顯示查無此項。切換到PTT文章時，若文章大於一定數量則導入翻頁機制，當然，對於錯誤的輸入必須有所回應。

Update 1:

1. DONE:

爬取原價屋網頁(克服編碼、網頁架構問題)，將資料儲存成可轉換成json檔的格式、測試pyptt套件能否使用。

1. CHANGE:

pyptt套件需要有ptt帳號才可使用，目前找不到其他可替代的API，改變計畫移除顯示ptt文章的功能。用regular expression擷取價格後用google chart API 顯示此類型商品的價位分布。

Update 2:

CHANGE:

Google Chart API似乎是開發給html的API，所以用Matplotlib取代價格分析的圖表部分，並引入Google sheet API以添加收藏清單的功能，將Google custom search的搜尋範圍限制在PTT的PC\_shopping版以取代本來的PyPtt

Run:

1. 安裝套件:
2. Matplotlib:

python -m pip install -U pip

python -m pip install -U matplotlib

1. Gspread & oauth2clien:

pip install gspread oauth2client

1. 程式執行:

請將code和triple-access-301406-17297ed9c321.json檔案放在同資料夾便可執行

(備註 - 執行環境:僅在Windows系統執行過，以VScode執行確認可以，此外附上在Jupyter執行成功後下載的.py檔)

Timeline(updated):

1. 確認上述每項提到的內容能否成功地與網頁互動並抓取資料下來

(Before 12/20)(√)

1. 爬取原價屋的網站，並提取出商品資訊，視進度將其作為第一次Demo

(Before 12/24) (√)

1. 完成Progress report (Before 12/25) (√)
2. 搜索PTT的文章，將所有相關討論抓取並顯示 (X) (Removed)
3. 測試並使用google chart API，整理價格資訊 (X) (Removed)
4. 利用Matplotlib繪製價格分析圖(Before 1/7) (√)
5. 借助google sheet API建立收藏清單(Before 1/10) (√)
6. 透過google search API提供其他的相關資訊(Before 1/10) (√)
7. 完善使用介面，進行Bug偵測(Before 1/14) (√)