

國立台灣科技大學資訊工程系
一百零七學年度第一學期專題研究

總報告

商品智慧訂閱系統

研 究 組 員

廖晨竹 B10415031

謝明軒 B10415038

林崇恩 B10415046

陳重佑 B10415047

指導老師：邱舉明 教授

中 華 民 國 107 年 10 月 22 日

目錄

目錄	I
第一章 緒論	1
1.1 研究動機	1
1.2 研究目的	1
1.3 研究簡介	1
1.4 報告組織	2
第二章 相關研究	3
2.1 Python	3
2.2 PHP	3
2.3 CSS	4
2.4 HTML	4
2.5 SQLite	5
2.6 網路爬蟲	5
2.7 聊天機器人Chatbot	6
2.8 Selenium	7
2.9 BeautifulSoup	7
2.10 sqlite3	7
2.11 Requests	7
2.12 re	8
2.13 string	8
2.14 smtplib	8
第三章 系統操作架構	8
3.1 執行流程	8
3.1.1系統主流程	8
3.1.2網頁執行流程	9
3.1.3 Line Chatbot執行流程	9
3.1.4 爬蟲執行流程	10
3.2程式架構	11
3.2.1 搜尋及查詢資料庫(演算法)	11
3.2.2 寄信與資料庫比對	12
3.2.3 網路爬蟲演算法	13
3.2.4 LINE Chatbot基礎架構	15
第四章 結果與討論	16
4.1前端畫面	16
4.1.1 首頁畫面	16
4.1.2 會員註冊畫面	16
4.1.3 登入畫面	17
4.1.4 商品搜尋	17
4.1.5 搜尋結果	18
4.2 Line ChatBot 介面	19
4.2.1 基本操作選單	19
4.2.2 指令	19

4.2.3 登入	20
4.2.4 登出	20
4.2.5 訂閱	21
4.2.6 查看	21
4.2.7 刪除	22
4.3 寄信範例	23
4.4 結論	23
參考文獻	24

第一章 緒論

1.1 研究動機

線上購物已成為現代人購物的主要方式之一，電子商務已然改變我們的生活方式，精打細算的客戶穿梭在不同的平台上比較價格，卻不一定可以找到符合其需求的商品。潛在客戶可能會因為沒有找到商品而流失，但也許隔幾天就有符合需求的商品出現，而且客戶的需求可能不全是突發性的，而是有一段需求的區間，例如購買生日禮物、機票等。

現有的購物網站，使用者需要每天固定去搜尋，他才能確認是否有他所需要的商品，而我們提供他們以需求的方式搜尋商品，使用者僅需要設定下商品、價格等相關資訊，我們即可為他提供定期搜尋，並且於需求符合時通知他。研究搜尋的執行時機、需求資訊與產品之間的比對方式、演算法的改進等，以期提供使用者更佳的購物經驗與方式，也是本計畫著眼的方向。

1.2 研究目的

開發出一個系統可以自動搜尋使用者所需的物品商品，現在人們能使用的線上購物網站太多，幾乎不可能將所有的購物網站都全部看過一遍來尋找自己預算中的商品。藉此這個系統來當個小助理幫忙這樣需求的用戶。只要輸入所需的物品、價格、規格等等。系統會自動搜尋直到一有物件便會立刻用電子郵件或其他通信軟體通知使用者，讓使用者能夠在最短最少的時間內也最精確的搜尋到他所要的物件。

- 時間：客戶為此商品設定一段期望購買到它的時間
- 預算價格：客戶的預算容許多少價格的商品，甚至不需要價格只需要輸入想要的商品，也能夠搜尋。
- 商品：客戶想購買的商品。例：手機iPhoneXS、筆電 ASUS S406UA 256G

1.3 研究簡介

結合使用者操作面、資料庫系統、背後的爬蟲程式來完成我們的專題。後端的爬蟲程式會往各大拍賣網站或是購物網站爬下所有商品資料進我們的大資料庫裡，而使用者在登入網站後即可對網站下訂單也就是搜尋功能。而使用者的訂單會儲存在資料庫系統裡，而資料庫就會開始比對開始檢索是否有對應資料(商品)。如果符合使用者需求，便會寄信給使用者通知已搜尋到。

- 使用者操作面：也就是前端網頁製作，需要把使用者的資料與資料庫結合，並且設計出使用者方便的介面。
- 資料爬蟲：也就是後端的資料來源，需要寫出爬蟲程式使其可以在各大拍賣網站或購物網站抓取資料，並且分類出名稱、樣式、價格、網址等等相關資料儲存進資料庫裡。
- 資料庫系統：結合了前端與後端的中樞橋梁，包括銜接程式運作與健全的資料庫。需要利用前端的使用者需求搜尋後端的爬蟲資料庫，進而把結果告知使用者，並且使用者能隨時修改需求時間及商品。資料庫要能隨時面對使用者不同的需求產生不同的結果。
- Line Chat Bot：鑒於現在智慧型手機普及，我們也加入了使用Line Chat Bot來做操作介面，讓使用者能直接在LINE上登入帳號並搜尋商品，而找到的商品也會直接用LINE來通知使用者，讓使用者又多出一個方便的管道。

1.4 報告組織

報告將針對相關研究、系統操作架構、結果與討論進行說明。第一章緒論，說明研究的動機目的與簡介此研究。第二章相關研究，簡述我們所採用的程式語言、框架。第三章系統操作架構部分，說明使用者之操作方式與程式內部執行之架構。第四章結果與討論，對於計畫最終所完成的結果進行說明與討論。

第二章 相關研究

2.1 Python

是一種廣泛使用的高階程式語言，屬於通用型程式語言，由Guido van Rossum 創造，第一版釋出於1991 年。作為一種直譯式語言，Python 的設計哲學強調代碼的可讀性和簡潔的語法（尤其是使用縮排劃分代碼塊，而非使用大括弧或者關鍵詞）。相比C++ 或Java，Python 讓開發者能夠用更少的代碼表達想法。不管是小型還是大型程式，該語言都試圖讓程式的結構清晰明了。

與Scheme、Ruby、Perl、Tcl 等動態型別程式語言一樣，Python 擁有動態型別系統和垃圾回收功能，能夠自動管理記憶體使用，並且支援多種編程範式，包括物件導向、命令式、函數式和程序式編程。其本身擁有一個巨大而廣泛的標準庫。

Python 虛擬機器本身幾乎可以在所有的作業系統中執行。Python 的官方直譯器CPython，用C語言編寫，是一個由社群驅動的自由軟體，目前由Python軟體基金會管理。

2.2 PHP

PHP-全名PHP:Hypertext Preprocessor (Personal Homepage Program)

PHP是一套讓你可以製作動態網頁的工具。內嵌PHP語法的HTML網頁文件仍會被當作一般的HTML網頁文件來處理，可以用同樣的標準HTML語法來編輯HTML部分。

內嵌的意思是，將PHP的程式寫在HTML裡面，而當伺服器執行網頁程式時，會先去直譯PHP的程式，並輸出在使用者的瀏覽器上，使用者所看到的是PHP程式執行的結果，而不是原始程式。

由於開放、免費、跨平台、易學及效率高等等的優點，全世界有許多的團體和組織提供一大堆的原始碼及函式庫，可協助快速而彈性地進行開發工作。所以說，PHP 可簡單的視為一種程式語言，可以用來設計：互動式介面、留言板、討論區、電子商務、多媒體應用、聊天室等等的動態網頁。

雖然PHP算是一個後端語言，但在此次專題中，前端負責人也有利用PHP加上MySQL來進行會員機制的管理以及資料呈現的撰寫。

2.3 CSS

現在要介紹的CSS再拿人來舉例的話，就好比人的膚色頭髮身上穿的衣服等等，對於HTML來說提供了裝飾、排版等功能。能夠使HTML所架構出來的網站進行更進一步的優化。

CSS是不能單純使用的，就好比衣服放在那邊終究只是一件衣服，重點還是要將其穿在人身上，才有它實質的功用，因此HTML跟CSS可以說是一體兩面，缺一不可的程式語言。

而CSS也如同PHP一樣有內嵌的特性，可以於HTML用style標籤所包圍起來，或是在任一HTML標籤當中引入style這個關鍵字，如此一來網頁在編譯時會自動將此區塊內的程式碼看做是CSS的程式，但大多數工程師都會將CSS撰寫於另外的檔案內，再於HTML用引入的方式來引入CSS，如此一來的好處是能使HTML的程式碼最為簡潔，若是要對CSS或HTML進行修改都能更為直覺以及方便。

2.4 HTML

HTML-全名(Hypertext Markup Language)，是一種超文字標示語言，利用其特有的元素(Element)以及標籤(Tag)，用來控制網頁上各式各樣的樣式呈現，包括文字大小、字體粗細、圖片或超連結的呈現等等，都能透過HTML來完成，HTML提供了一個網站最最基本的樣式呈現，就好比是人類的骨架，屬於網站基底的部分。

所謂「超文字」(Hypertext)和一般傳統文件最大的不同就是，傳統文件只能讓你循序地閱讀，而Hypertext藉由一些特殊標籤（例如連結）的協助，可以让你以樹狀甚至網狀（所以叫Web）來組織你的文件。標示(Markup)是透過在文章中插入標籤(Tags)來賦予文字一些特性，如標題，段落，連結等等。標籤乃是由「小於」(<)與「大於」(>)這兩個符號所括住的文字，如果想對某段文字加上特別的意義，我們只要在前後分別加上開始與結束的標籤即可，例如<title>我是標題</title>以上這段程式碼便能產生網頁最上層的網頁標題。

而當我們在瀏覽網頁時，可以隨時按下右鍵選擇(開發者工具/檢查網頁原始碼)，便能清楚的瀏覽當下你面前的網頁的內部結構，也能透過此工具修改此網站上的HTML進行暫時性的介面修改，CSS也能透過此工具進行暫時性的更動。

2.5 SQLite

SQLite 與其他一般資料庫差異不大，一般的SQL-92 語法都能夠使用，而且不需要建立一個資料庫系統，要使用的時候，只要在編譯程式的時候將SQLite 程式庫一起編入就可以使用。另外，SQLite 的資料庫（database）都是以單一檔案的形式存於磁碟中，不需要再安裝資料庫伺服器軟體，所以要把資料庫複製或建立在你的電腦上是相單簡單快速。

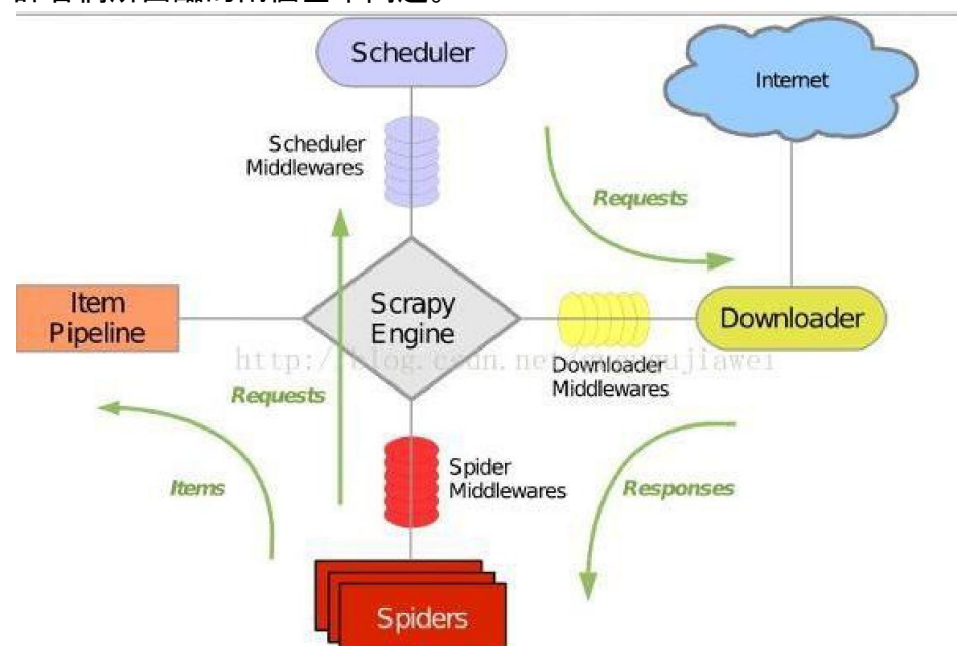
SQLite 是一個很小的C 語言程式庫，且本身就完全包含資料庫引擎的功能，而且可以嵌入至其他程式中，完全不用額外的設定。

2.6 網路爬蟲

網路爬蟲也叫網路蜘蛛（spider），是一種用來自動瀏覽全球資訊網的網路機器人。其目的為編纂網路索引。本專題採用基於python語言的scrapy。

網路爬蟲始於一張被稱作種子的統一資源位址（URL）列表。當網路爬蟲存取這些統一資源定位器時，它們會甄別出頁面上所有的超連結，並將它們寫入一張「待訪列表」，所謂爬行疆域（crawl frontier）。此疆域上的諸URL將被按照一套策略迴圈存取。

網際網路資源多如繁星，這意味著網路爬蟲在一定時間內只能下載有限數量的網頁，因此它需要最佳化它的下載方式。網際網路資源瞬息萬變，這也意味著網路爬蟲下載的網頁在使用前就已經被修改甚至是刪除了。這是網路爬蟲設計者們所面臨的兩個基本問題。



2.7 聊天機器人Chatbot

聊天機器人是透過文字、圖片等進行交談的電腦軟體，是近幾年來發展最熱門且應用在無論是商業或生活上最廣泛的技術之一，其實早在1966年麻省理工學院就有發展出當時世界上第一個聊天機器人(ELIZA)，隨著人工智慧、語意分析等技術的大幅進步，加上社加平台的全面普及化，聊天機器人又再一次活躍起來。目前聊天機器人的應用範圍包括客戶對答服務或資訊獲取應用等，有些聊天機器人會搭配自然語言處理系統，讓聊天機器人能自己找尋最合適的應答文字，讓使用者誤以為和真人在交談，這樣不僅能節省人力的成本，更能提升市場價值。

現今的聊天機器人廣泛運用在各大即時通訊消息平台，例如LINE、臉書Messenger、WeChat、Kik等，通常是以網路行銷、客戶服務和休閒娛樂為主。各大品牌廠商也運用聊天機器人來進行販售產品、回答問題、推廣品牌等；也能運用在教育上，聊天機器人能針對學生的問題進行回答，讓學習更加有效率且更方便。找尋聊天機器人最適合運作的場合，並加強與使用者溝通流暢性的行為，是目前聊天機器人重要發展的議題。

2.8 Selenium

是自動化測試工具。它支持各種瀏覽器，包括Chrome, Safari, Firefox 等主流界面式瀏覽器，如果你在這些瀏覽器裡面安裝一個Selenium 的插件，那麼可以方便地完成對Web界面的測試。換句話說叫Selenium 支援這些瀏覽器驅動。Selenium支持多種語言開發，比如Java, C, Ruby、Python。安裝時在cmd中輸入pip install selenium 即可。

2.9 BeautifulSoup

Beautiful Soup是一個Python包，功能包括解析HTML、XML文檔、修復含有未閉合標籤等錯誤的文檔（此種文檔常被稱為tag soup）。這個擴展包為待解析的頁面建立一棵樹，以便提取其中的數據，這在網絡數據採集時非常有用。

2.10 sqlite3

python對sqlite資料庫操作的library，主要使用包含connect、cursor、fetch、commit等函式。

connect()：即是資料庫的連接，可以利用它來切換連接到不同的資料庫。

cursor()：類似操作介面，是用來操作connect後的回傳值經cursor呼叫execute()即可使用sqlite的關鍵字。

fetch()：就是把execute()後的資料抓取出來，存成python列表的形式，有fetchone()、fetchall()兩種。

`commit()`：即許可，許可前面所有尚未被許可的操作但僅限於cursor所對應的資料庫。

2.11 Requests

是在Python上的一個HTTP library由Kenneth Reitz所撰寫，是用來做網路爬蟲非常簡單易懂的一個工具，它可以將一個網頁用僅僅幾行程式碼就將其抓下並用可讀的格式輸出處理，雖然requests並不是萬能的，可是對於大部分的網站卻是沒有問題的。

2.12 re

re就是所謂的正規化，正規化可以幫助我們更容易擷取的字串中我們所需要的部份。例如在爬蟲的時候，因為網頁都是用html寫的，所以可能會包含很多我們不想要的內容，這時候就可以用`re.search`，抓到我們要的，也可以利用`re.split`，切字串的方式，還有我們在做商品需求的字串比對的時候，為了預防可能有人亂輸入名稱，也許有特殊符號或是空格，所以我們會先用re模塊做一些字串上的整理。

2.13 string

string模組也是在對字串處理上有很大的幫助，跟re模塊可以相互幫助使用，像是`punctuation`的功能，可以將符號都刪除掉，`replace`則可以把字元替換成我們想要的。

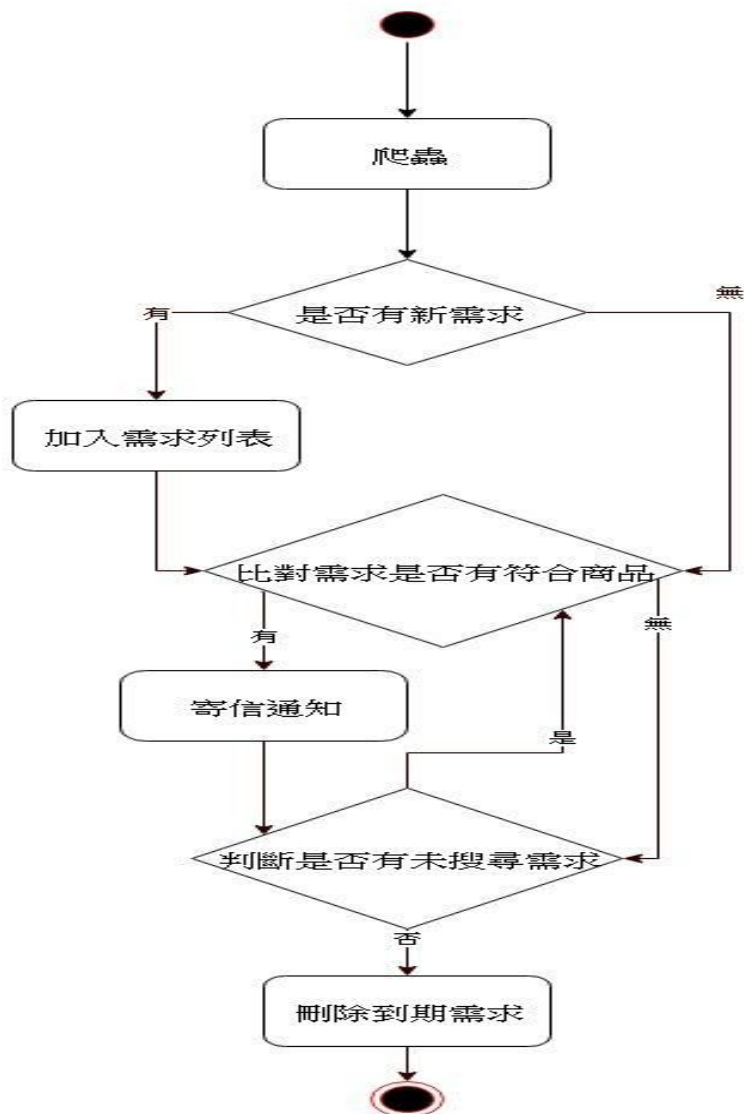
2.14 smtplib

SMTP(Simple Mail Transfer Protocol)即簡單郵件傳輸協議，python對smtp協議進行簡單的封裝提供很方便的途徑發送email。

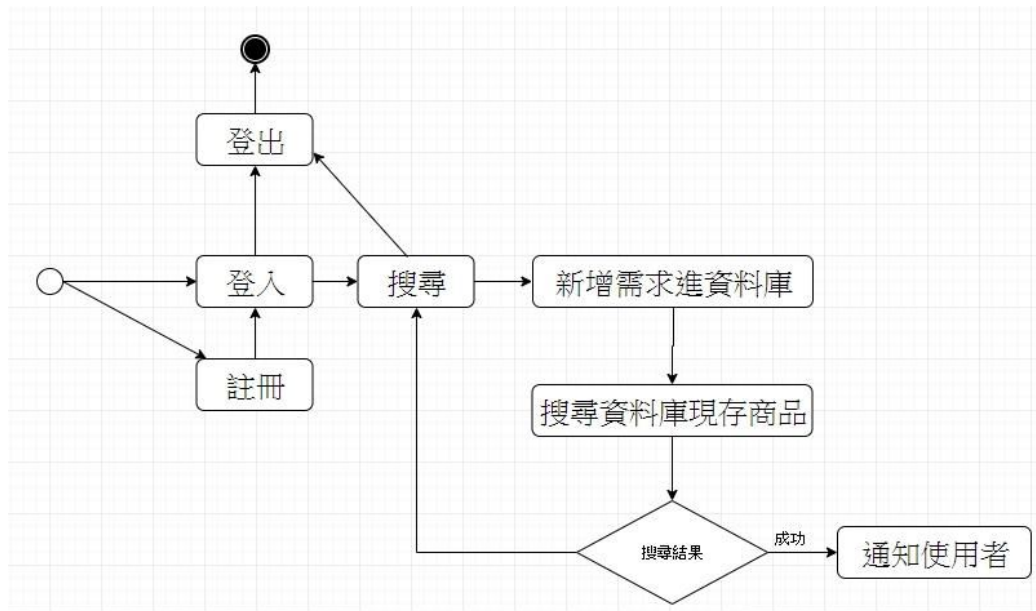
第三章 系統操作架構

3.1 執行流程

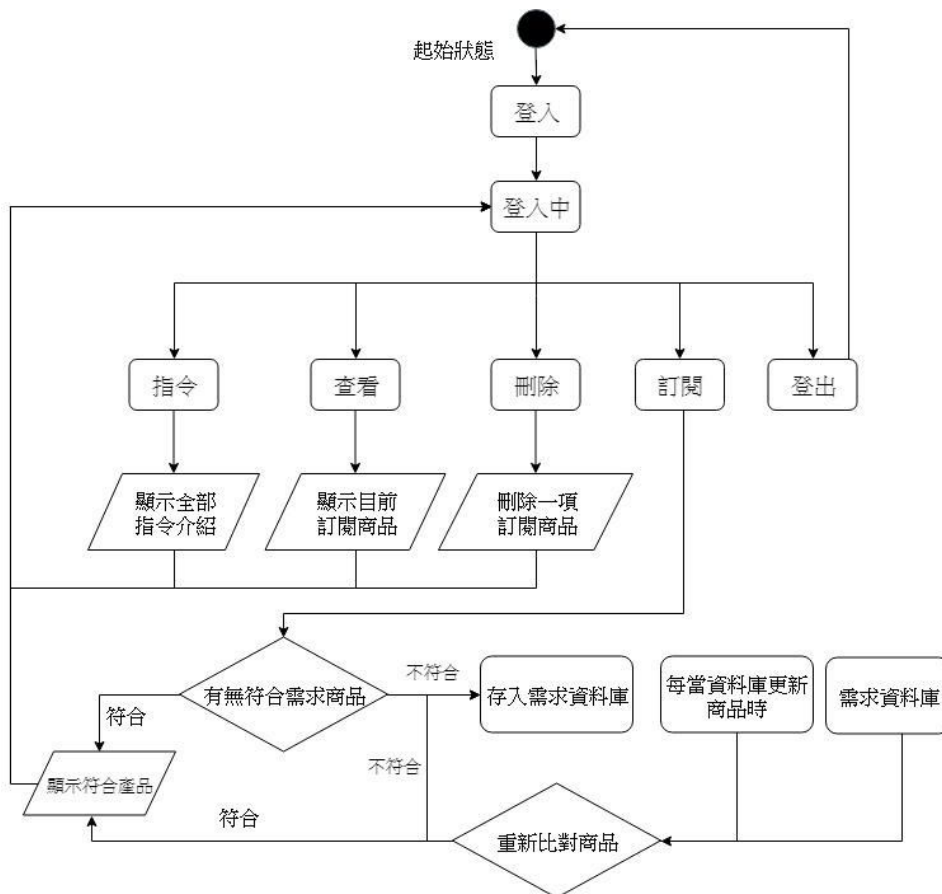
3.1.1系統主流程



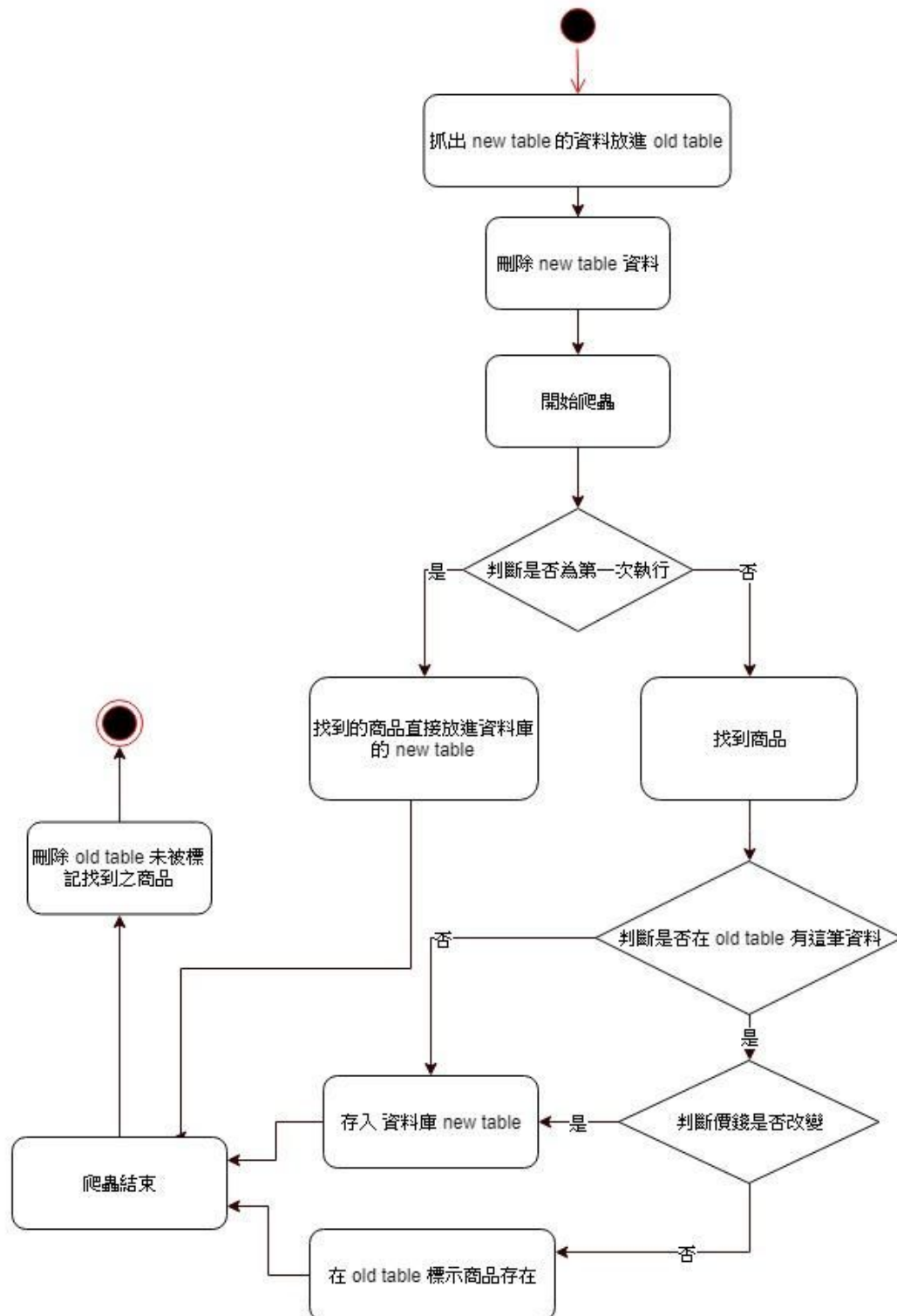
3.1.2 網頁執行流程



3.1.3 Line Chatbot執行流程



3.1.4 爬蟲執行流程



3.2 程式架構

3.2.1 搜尋及查詢資料庫(演算法)

a. 系統邏輯

1. 前端協作
2. 後端協作

main() 中執行的順序：

1. 呼叫爬蟲
2. 確認是否有新商品
3. 確認是否有新需求
4. 搜尋商品是否有符合需求的，有則寄信
5. 刪除到期需求(目前設定期限為兩周)

由於商品價格不會每天都改變，有的甚至一兩個月才會有改變，所以爬蟲這部分是採取 windows 工作排程器，定期呼叫程式去執行(目前設定一天一次)

b. 搜尋方法

1. 字串分析(英文數字)

目前可搜尋商品分為三類：手機、筆電、平板，假設今天要搜尋的商品名稱為 iphone7 plus，在 iphone 跟 7 中間訂閱者輸入時並沒有空格，會造成後端在資料庫內搜尋商品時，會因為空格的關係，明明要比對到的產品，但因為字串比對不到，造成比對不精確，所以遇到數字的部份我們會進行切割，像是上面的例子會被切成 iphone、7 以及 plus 這三個字，在資料庫中商品字串要有符合這三個才會被比對到，英文的部分就是利用空格去做切割。

2. 字串分析(中文)

中文的部分，因為要讓系統知道什麼樣單字、句字才會符合語意，所以我們利用「結巴分詞」(jieba)。

結巴分詞是目前大家認為做最好的 python 中文分詞套件，支援的模式有三種：「精確模式」、「全模式」以及「搜尋引擎模式」，精確模式顧名思義就是做最精確的切開，而全模式則是速度較快但不能解決歧義，搜尋引擎模式則是在精確的基礎上，對較長的詞語再次切分，提高召回率。

結巴有自己的字典(dict)，裡面有上萬的詞語，包含詞彙出現的次數和詞性，使用者也可以依照自己的需求，將裡面沒有的詞彙加入其中，或是改變出現次數等等，因為結巴分詞屬於概率語言模型，這個模型的任務就是在全切分所得的結果，求某個切分方案 S，使 $P(S)$ 最大。

3.2.2 寄信與資料庫比對

1.資料刪除

購物網站的商品如果下架，必須要在我們的資料庫商除這筆資料，避免訂閱者的需求搜尋到一筆不存在的資料。

2.資料更新

商品價格如果有變動，我們並不會建立新的資料，而是更新資料庫舊有的資料。

3.需求到期刪除

需求的搜尋期限為兩周，原本採用 (現在時間 - 發起需求的时间) 是否等於14，這樣的方式去做判斷，不過這樣要多做一步運算的步驟，雖然影響不是很大但多少還是會影響程式效率，所以最後就是在需求建立同時設定到期時間，只要比對到期時間是否等於現在時間。

4.需求符合寄信

import python 原有套件 smtplib 來做寄信功能，符合的商品可能不只一件，所以會先把搜尋到的商品放入陣列，待全部比對完成後再做寄信，寄信內容為商品名稱、價錢以及網址，使用 gmail 來寄信。

5.寄信通知:

當商品符合需求時，會以寄信方式來通知使用者。寄信的程式使用SMTP (Simple Mail Transfer Protocol) 的郵件傳輸協議，它是用一組源地址傳送到目的地址的傳送郵件規則。首先先建立SMTP連線，接著跟Gmail Sever進行溝通，並使用TLS的安全認證機制，然後登入Gmail傳輸資料到使用者的mail中，完成後關閉連線。

3.2.3 網路爬蟲演算法

a. 總流程及比對資料庫

1. 程式最開始會先初始化資料庫，如果是第一次執行則會建立資料庫以及所需要的 table(手機、筆電、平板)，裡面的欄位分別為：title, price, website, time, flag，前三個都是商品的資訊，time 則是爬到商品的時間，flag 是用來分辨是哪一個購物網站抓取下來的資料。
2. 接著執行 parse()，parse function 一開始會先執行 initFunc，這個 function 會把舊的資料抓出來，之後在爬蟲時會需要他們去比對是否已經存在，之後因為這邊使用 selenium，所以就會打開瀏覽器開始爬蟲。
3. 爬蟲找每一筆資料時，若是第一次執行則會無條件加入資料庫，否則會先和網址去比對是否有一樣的，如果沒有表示是新上架商品，直接存入資料庫的 new table，如果有再去比對價錢是否一樣，若一樣代表商品維持原樣且存在，若不一樣則代表價錢更改過，直接更新資料庫資料。
4. 判斷商品還在不在的方式是用一個變數 count，只要資料庫 old table 的資料有在這次爬蟲中被比對到，count 就設為 1，否則維持 0，在爬蟲結束之後會把 count = 0 的資料刪除，代表此商品已不再網頁上，這樣可以避免訂閱者需求找到已被刪除的頁面。

b. 爬蟲過程（露天）

1. parse() 會從 urls 陣列中取出欲爬取的網址，並用 for 迴圈去一一輸入瀏覽器。
2. 因為利用 selenium 模擬登入，所以會打開瀏覽器到接收到的網址，再使用 BeautifulSoup 套件去解析爬到的 html，方便我們去擷取需要的資訊。
3. 露天的瀏覽方式為換頁式，意即網址最後的數字為頁碼，所以只要將那個數字設定好範圍，利用迴圈下去跑，即可自己設定要爬的頁數。

c. 爬蟲過程（樂天）

1. parse() 會從 urls 陣列中取出欲爬取的網址，並用 for 迴圈去一一輸入瀏覽器。

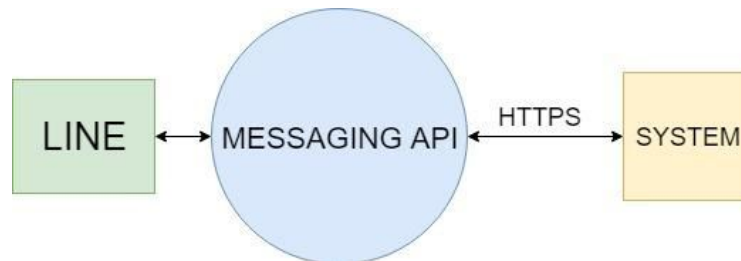
2. 因為利用 selenium 模擬登入，所以會打開瀏覽器到接收到的網址，再使用 BeautifulSoup 套件去解析爬到的 html，方便我們去擷取需要的資訊。
3. 樂天的瀏覽方式為換頁式，爬蟲資料爬到最後的頁數，因為頁數可能會增加或減少，所以有特別寫動態更換最後的頁數，防止錯誤發生。

d.爬蟲過程（PChome）

1. 使用 requests.get() 來抓下網頁的資料，再用json函數庫來將資料解析。
2. 將需要的數值從解析後的資料中一個個取出。
3. 雖然PChome24是使用滾輪式列表來顯示，但是網址卻是用頁碼來處理，所以只要將頁碼改成下一頁就能繼續重複抓取資料。

3.2.4 LINE Chatbot基礎架構

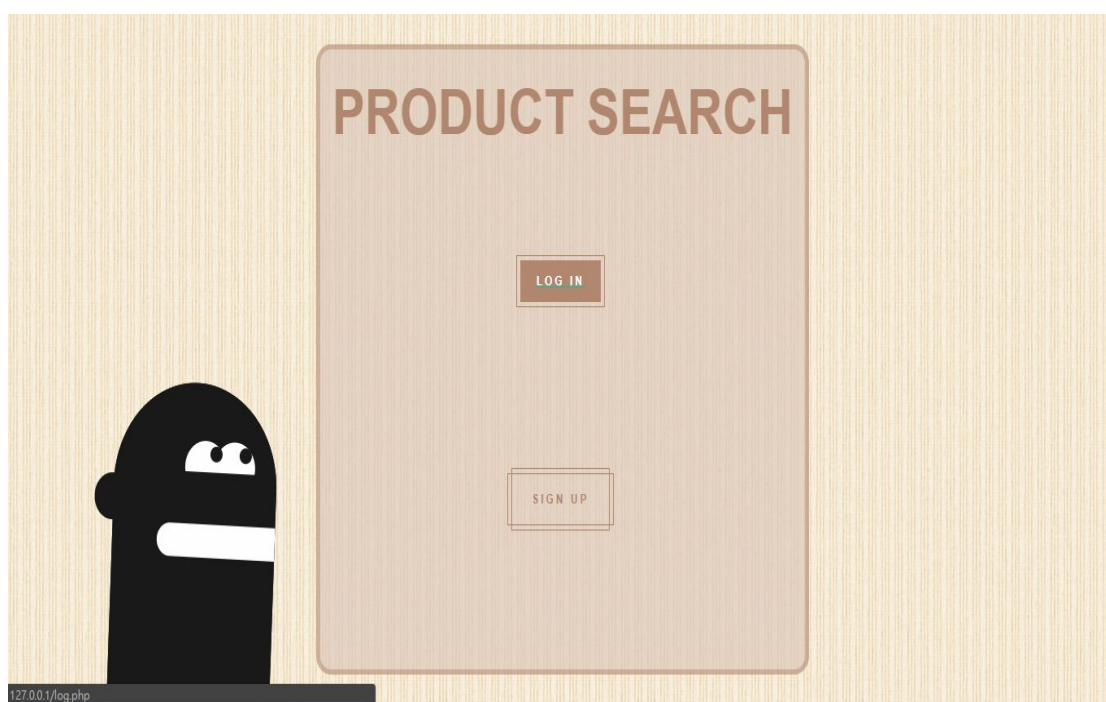
LINE Chatbot是LINE 提供的Messaging API（LINE公司的伺服器），它可以讓開發人員建立一個可接收並回覆訊息給好友的應用程式，也就是俗稱LINE BOT聊天機器人。LINE BOT的基本處理訊息的架構為下圖，Messaging API負責接收客戶端的訊息，並把接收的訊息傳給開發者的網站伺服器進行處理，伺服器端處理完後會回覆給Messaging API，並由它轉發相對應的訊息給客戶端。



第四章 結果與討論

4.1 前端畫面

4.1.1 首頁畫面



4.1.2 會員註冊畫面



會員註冊

帳號:

密碼:

確認密碼:

信箱:

手機:

註冊

[回到首頁](#)

4.1.3 登入畫面



登入

帳號

密碼

登入

[回到首頁](#)

4.1.4 商品搜尋

LET'S SEARCH!

商品類型

平板

商品名稱

商品價格

搜尋

登出

4.1.5 歷史結果

登錄

登出

以下為您的搜尋紀錄 若有符合條件之商品 將立即通知您!!

商品名稱	商品價錢	搜尋日期
asus	20000	2018-08-01 14:08:32
蘋果	12345	2018-08-13 20:42:32
華碩	2500	2018-08-17 12:01:54
聯想	21000	2018-08-24 09:24:36
華碩	132	2018-09-18 10:50:42
華碩	32000	2018-09-25 14:04:40

4.2 Line ChatBot 介面

4.2.1 基本操作選單:

進入Chatbot後底下點選「操作指令集」，目前有六個功能可以使用，分別是訂閱、查看、刪除、指令、登入和登出。



4.2.2 指令:

點選「指令」後，Chatbot會告知此程式的操作指令說明。



4.2.3 登入:

點選「指令」後，必須填入網站註冊帳戶的帳號和密碼，才能使用Chatbot的功能。Chatbot設定為一個帳戶只能在一個裝置登入使用。



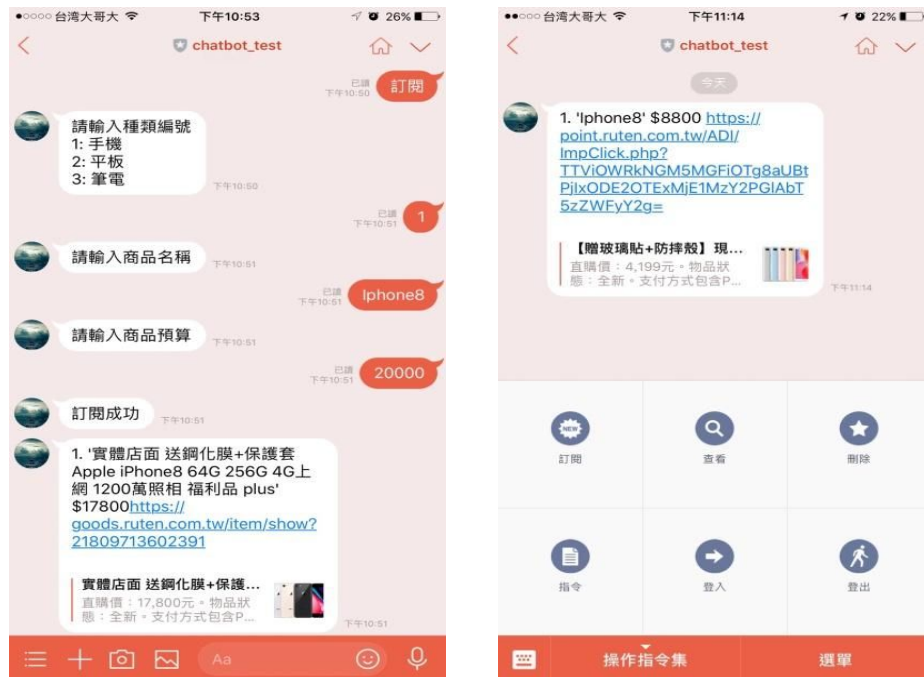
4.2.4 登出:

點選「登出」後，可以在別的裝置重新登入或登入另一個帳戶。



4.2.5 訂閱:

點選「訂閱」後，輸入相對應的文字即可完成訂閱商品功能。當首次完成訂閱功能，如果當下有符合需求的商品，Chatbot會立即回覆；如果沒有的話，Chatbot會直到搜尋到符合商品才回覆訊息。



4.2.6 查看:

點選「查看」後，Chatbot會告知目前訂閱商品的列表。



4.2.7 刪除:

點選「刪除」後，輸入相對應的刪除編號即可完成刪除訂閱的商品。



4.3 寄信範例



參考文獻

[1] 學習WEB開發:

https://developer.mozilla.org/zh-CN/docs/Learn/CSS/Introduction_to_CSS
W3school : <https://www.w3schools.com/>

[2] PHP:

<https://ithelp.ithome.com.tw/articles/10191950>

[3] PHP教學:

<http://emn178.pixnet.net/blog/post/78014066-php%E6%95%99%E5%AD%B8---%E5%9F%BA%E6%9C%AC%E8%AA%9E%E6%B3%95%28syntax%29>

[4] 六角學院:

<https://www.hexschool.com/>

[5] Messaging API

<https://developers.line.me/en/docs/messaging-api/overview/>

[6] Line Message API with NodeJS

<http://yonaichin.logdown.com/posts/1256421-line-message-api-with-nodejs>

[7] LineBot+Python, 輕鬆建立聊天機器人

<https://yaoandy107.github.io/line-bot-tutorial/>

[8] 只要有心, 人人都可以做卡米狗

<https://ithelp.ithome.com.tw/users/20107309/ironman/1253>

[9] 網路爬蟲實戰教學

<https://www.youtube.com/watch?v=ceUhb2-gYOU&list=PLohb4k71XnPaQRTvKW4Uii1oq-JPGpwWF>

[10] re---Regular expression operation

<https://docs.python.org/3/library/re.html>

[11] Python 正則表達式

<http://www.runoob.com/python/python-reg-expressions.html>

[12] python smtp lib

https://self.jxtsai.info/2016/09/python_22.html

[13] selenium 莫煩

<https://morvanzhou.github.io/tutorials/data-manipulation/scraping/5-01-selenium/>

[14] Mysql 遠端連線

<https://www.phpini.com/mysql/mysql-enable-remote-connect>

[15] Python pandas

<https://jerryneet.io/python-pandas-get-data/>

[16] Python sqlite

<http://www.runoob.com/sqlite/sqlite-python.html>

[17] python bs4

<https://blog.gtwang.org/programming/python-beautiful-soup-module-scrape-web-pages-tutorial/>

[18] requests網站

<http://docs.python-requests.org/en/master/>

[19] Python 爬蟲實作

<http://ivanjo39191.pixnet.net/blog/post/70548675-python-%E7%88%AC%E8%9F%>

B2%E7%B7%B4%E7%BF%92%E7%B4%80%E9%8C%842%28%E4%B8%80%29

[20] Python網路爬蟲

<https://www.youtube.com/watch?v=YHc6KvLsv20>

[21] sqlite 語法教學

<http://work.oknow.org/2014/02/python-sqlite.html>