國立台北商業大學

資訊管理系

112．資訊系統專案設計

**系統手冊**

**組別：第112510組**

**題目：iCrawler**

**指導老師：陳信宏**

**組長：10856040鍾皓年**

**組員：10856010郭宗翰 10856030彭鈺程**

**10856031彭鈺達 10856032陳永祥**

**中華民國112年5月10日**

目錄

[第一章 前言 1](#_Toc134191273)

[1-1背景介紹 1](#_Toc134191274)

[1-2動機 1](#_Toc134191275)

[1-3系統目的與目標 1](#_Toc134191276)

[1-4預期成果 1](#_Toc134191277)

[第二章 營運計畫 2](#_Toc134191278)

[2-1 可行性分析 2](#_Toc134191279)

[2-2 商業模式－Business model 3](#_Toc134191280)

[2-3 市場分析－STP 3](#_Toc134191281)

[2-4競爭力分析－SWOT 4](#_Toc134191282)

[第三章 系統規格 5](#_Toc134191283)

[3-1 系統架構 5](#_Toc134191284)

[3-2 系統軟、硬體需求與技術平台 5](#_Toc134191285)

[3-3 開發標準與使用工具 6](#_Toc134191286)

[第四章 專案時程與組織分工 7](#_Toc134191287)

[4-1 專案時程 7](#_Toc134191288)

[4-2 專案組織與分工 8](#_Toc134191289)

[第五章 需求模型 9](#_Toc134191290)

[5-1 功能分解圖（Functional decomposition diagram） 9](#_Toc134191291)

[5-2 需求清單： 10](#_Toc134191292)

[第六章 程序模型 11](#_Toc134191293)

[6-1 資料流程圖(Data flow diagram)。 11](#_Toc134191294)

[6-2 程序規格書(Process specification)。 14](#_Toc134191295)

表目錄

[第二章 營運計畫 2](#_Toc134581901)

[表 3-3-1開發標準與使用工具 4](#_Toc134581902)

[第三章 系統規格 5](#_Toc134581903)

[表 3-2-1系統軟、硬體需求 5](#_Toc134581904)

[表 3-3-1開發標準與使用工具 6](#_Toc134581905)

[第四章 專案時程與組織分工 7](#_Toc134581906)

[表 4-1-1、專案時程 7](#_Toc134581907)

[表4-2-1、專案組織與分工 8](#_Toc134581908)

[第六章 程序模型 11](#_Toc134581910)

[表 6-2-1、程序規格書 委託系統 14](#_Toc134581911)

[表 6-2-2、程序規格書 委託區 14](#_Toc134581912)

[表 6-2-3、程序規格書 分潤 14](#_Toc134581913)

[表 6-2-4、程序規格書 會員系統 15](#_Toc134581914)

[表 6-2-5、程序規格書 客服系統 15](#_Toc134581915)

圖目錄

[第五章 需求模型 9](#_Toc134580817)

[圖 5-1、功能分解圖 9](#_Toc134580818)

[第六章 程序模型 11](#_Toc134580819)

[圖 6-1-1、系統環境圖 11](#_Toc134580820)

[圖 6-1-2、圖0 iCrawler 12](#_Toc134580821)

[圖 6-1-3、圖1委託系統 13](#_Toc134580822)

第一章 前言

1-1背景介紹

在現在這個大數據時代，科技的進步日新月異，不管是5G、大數據等科技都在不斷的成熟。而網路上的資訊量非常龐大，因此需要一種方式來收集、分析和整理這些資料。

1-2動機

隨著互聯網的迅速發展，人們需要大量的資訊來滿足日常生活和工作需求。爬蟲卻可以從網路上獲取大量的資訊，可以有效地提高人們的資訊搜尋效率和準確性。

1-3系統目的與目標

一個爬蟲的軟體，讓一些對資料有需求，且沒有擁有程式語言能力的人能夠使用。

1-4預期成果

方便沒有程式語言能力的人，去做爬蟲功能。

第二章 營運計畫

2-1 可行性分析

市場可行性：提供了一個半自動化的爬蟲工具，讓不懂程式語言的人能夠輕鬆擷取網頁內容。

目標受眾：市場研究專業人士、數據分析師、學生等。

競爭分析：現在的半自動爬蟲的市場無法做重複的爬蟲。

市場增長潛力：隨著更多人需要獲取網絡數據的情況下。大數據和數據分析領域直線成長。

財務可行性：開發這個程式所需的成本，包括軟體開發、維護、伺服器和託管成本，以及可能的市場營銷費用。

預期收入：通過訂閱費、授權費用或廣告方式產生收入。

風險評估：有可能違反智慧財產權。

2-2 商業模式－Business model

基於使用量的模式：計費方式根據客戶的使用量，例如每月爬取的頁面數、爬取的數據量或操作次數。這適用於需要大量爬取的用戶，並根據他們的實際需求來定價。

2-3 市場分析－STP

S： 地域區分：爬蟲可以面向不同國家和地區的市場，全世界都可以用。

統計變數：根據不同收入水平，必須有電腦和網路才能使用。

生活型態：常接觸到電腦、會使用電腦來查詢資料 。

T： 集中行銷：針對特定的利基市場，例如:針對有資料需求的人。

P： 功能利益面：爬蟲網頁的功能利益包括自動化網路數據收集、處理與分析，提高工作效率，減少人力成本，提供即時資訊。

適用情境面：爬蟲網頁只適用於特定行業，例如金融、電子商務、新聞媒體等，可以應用於大量數據的收集、分析、處理等工作。

文化象徵面：爬蟲可以被看作是一個現代化和高效率的工具，這符合現代社會對於科技與效率的追求。

2-4競爭力分析－SWOT

▼表 3-3-1開發標準與使用工具

|  |  |  |
| --- | --- | --- |
| 內部因素  外部因素 | 優勢 (Strengths) | 劣勢 (Weaknesses) |
| 時間節省：協助用戶自動化重複性任務，節省了時間和精力。數據提取：有助於用戶輕鬆從網站上提取數據，進行分析或報告。  可定制性：具有彈性的設置選項，使用戶能夠定制其爬蟲操作以滿足特定需求。 | 網站變化：網站結構的變化可能需要不斷調整和更新爬蟲，使其保持有效。  可能受限於網站規則：部分網站可能設置了反爬蟲機制，導致無法正常收集資料。 |
| 機會 (Opportunities) | SO  (Strengths-Opportunities) | WO  (Weaknesses-Opportunities) |
| 市場需求：非技術人員和企業在數據自動化方面的需求不斷增加，爬蟲工具可以滿足這種需求。數據分析：數據爬蟲可以幫助企業更好地理解市場趨勢、競爭情況和客戶需求。擴大功能：將工具擴展到不同平台，例如移動應用，以滿足更廣泛的用戶需求。 | 基於可定制性和時間節省的優勢，開發更多針對不同行業和用途的爬蟲模板，以滿足市場需求。  利用市場需求增加的機會，擴展銷售，以吸引更多的非技術人員和企業客戶。 | 鑒於網站結構的不斷變化是一個劣勢，建立自動化的網站結構變化檢測功能，以減少調整和更新的需要。  可以與網站所有者的合作，以了解其反爬蟲機制，並提供解決方案以緩解這一劣勢。 |
| 威脅 (Threats)： | ST  (Strengths-Threats) | WT  (Weaknesses-Threats) |
| 競爭：有其他類似的爬蟲工具。  法律風險：由於網絡爬蟲可能觸碰到法律問題，需要嚴謹的規範。 | 基於數據提取和可定制性的優勢，提供教育服務，以幫助用戶更好地應對法律風險和網站規則的挑戰。 | 通過投資於研究和開發，提高軟體的效能和適應性，以應對競爭對手的爬蟲工具。 |

第三章 系統規格

3-1 系統架構



3-2 系統軟、硬體需求與技術平台

▼表 3-2-1系統軟、硬體需求

|  |  |  |
| --- | --- | --- |
| 軟、硬體需求 | | |
| 作業系統版本 | 最低系統需求 | 建議系統需求 |
| Windows 7 以上版本 | Windows 10 以上版本 |
| 處理器磁碟可用空間 | 雙核心以上 | 四核心以上 |
| 1 GB以上可用空間 | 2 GB以上可用空間 |
| RAM、網路 | 4 GB以上可用記憶體 | 8 GB以上可用記憶體 |
| 4G行動網路、Wi-Fi | |

3-3 開發標準與使用工具

▼表 3-3-1開發標準與使用工具

|  |  |
| --- | --- |
| 系統開發環境 | |
| 作業系統 | Win10、Win11 |
| 開發平台 | Python |
| 應用程式 | Visual Studio Code、 |
| 程式開發工具 | |
| 前端 | HTML、CSS、JavaScript |
| 後端 | Python |
| 文件美工工具 | |
| 文件 | Microsoft word |
| 簡報 | Microsoft PowerPoint、Canva |
| 圖樣 |  |
| 專案管理平台 | |
| 專案管理 | GitHub、GitKraken |
| 檔案存放 | MySQL、GitHub |

第四章 專案時程與組織分工

4-1 專案時程

▼表 4-1-1、專案時程

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **年** | **111年** | | | | | | | | | | |
| **月** | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 系統發想 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 工具學習 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 前端設計 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 前端開發 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 資料庫設計 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 資料庫建置 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 後端設計 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 後端開發 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 系統整合 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 系統測試 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| 手冊製作 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Logo製作 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| PPT製作 |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

4-2 專案組織與分工

▼表4-2-1、專案組織與分工

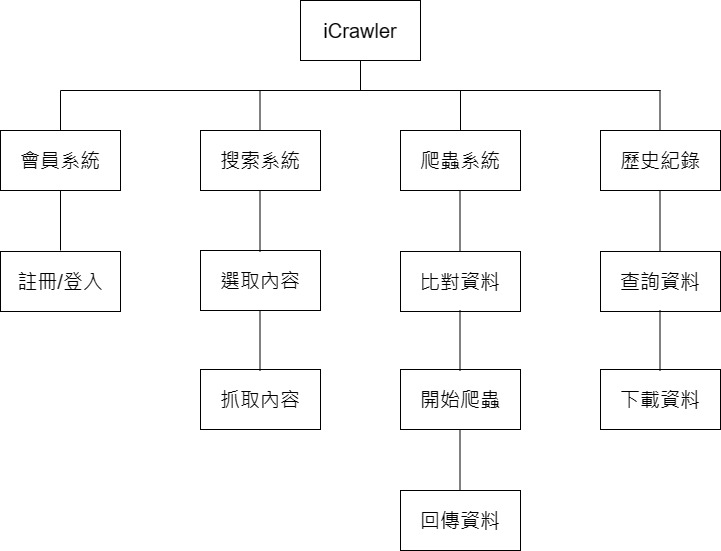
●主要負責人 〇次要負責人

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 項目/組員 | | 10856041  鍾皓年 | 10856010  郭宗翰 | 10856030  彭鈺程 | 10856031  彭鈺達 | 10856032  陳永祥 |
| 後端開發 | selenium |  | ● |  | 〇 |  |
| 腳本比對 |  |  |  |  |  |
| 伺服器架設 |  | 〇 |  | ● |  |
| webengine | ● |  |  |  |  |
| 腳本 |  | 〇 |  | ● |  |
| Xpath |  |  |  |  |  |
| 會員系統 | ● |  | 〇 | 〇 |  |
| 前端開發 | Pyqt5 | 〇 |  | ● |  | 〇 |
|  | 〇 | ● |  |  |  |
|  | ● |  |  |  |  |
| 美術設計 | UI/ UX | 〇 |  |  |  | ● |
| Web/APP介面設計 |  |  |  |  | ● |
| 色彩設計 |  | 〇 | 〇 |  | ● |
| Logo設計 |  |  | 〇 |  | ● |
| 素材設計 |  |  |  | 〇 | ● |
| 文件撰寫 | 統整 | ● |  |  |  | 〇 |
| 第1章 前言 |  | ● |  |  |  |
| 第2章 營運計畫 |  |  | 〇 | ● |  |
| 第3章 系統規格 | 〇 |  | ● |  |  |
| 第4章專題時程與組織分工 |  | 〇 | ● | 〇 |  |
| 第5章 需求模型 |  |  | 〇 | ● |  |
| 第6章 程序模型或設計模型 |  | ● |  | 〇 |  |
| 第7章 資料模型或實作模型 |  |  | ● |  |  |
| 第8章 資料庫設計 |  | ● |  |  |  |
| 第9章 程式 | ● |  |  |  |  |
| 第10章 測試模型 | 〇 |  |  | ● |  |
| 第11章 操作手冊 |  | 〇 |  |  | ● |
| 第12章 使用手冊 | 〇 |  |  |  | ● |
| 報告 | 簡報製作 |  | 〇 | 〇 |  | ● |
| 影片製作 |  |  |  | 〇 | ● |
| 海報製作 |  |  |  |  | ● |

第五章 需求模型

5-1 功能分解圖（Functional decomposition diagram）

▼圖 5-1、功能分解圖



5-2 需求清單：

**功能需求**

會員系統

1.1 使用者可以註冊和登入帳號及密碼

搜索系統

2.1 選取反白內容進行抓取

爬蟲系統

3.1 讀取抓取的資料，根據標籤進行比對內容

3.2根據比對內容去爬蟲

3.3回傳爬蟲資料

歷史紀錄

4.1 將使用者爬取的資料儲存在資料庫，以供使用者查看和分析

**非功能需求**

合法性

5.1 遵守相關的法規和法律，並提供提示，以防止非法或不當的使用

第六章 程序模型

6-1 資料流程圖(Data flow diagram)。

▼圖 6-1-1、系統環境圖



▼圖 6-1-2、圖0 iCrawler

▼圖 6-1-3、圖1 委託系統



6-2 程序規格書(Process specification)。

▼表 6-2-1、程序規格書 會員帳號系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 1 | **功能名稱** | 創建帳號密碼 |
| **操作說明** | 新增帳號、密碼、使用者名稱 | | |
| **輸入值** | 帳號、密碼、使用者名稱 | | |
| **輸出值** | 帳號、密碼、使用者名稱 | | |

▼表 6-2-2、程序規格書 登入、登出系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 1.1 | **功能名稱** | 會員登入、登出 |
| **操作說明** | 登入、登出帳號 | | |
| **輸入值** | 帳號、密碼 | | |
| **輸出值** | 帳號、密碼 | | |

▼表 6-2-3、程序規格書 搜索系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 2 | **功能名稱** | 搜索系統 |
| **操作說明** | 多次欲爬蟲的內容 | | |
| **輸入值** | 反白內容 | | |
| **輸出值** | 反白內容的xpath | | |

▼表 6-2-4、程序規格書 爬蟲系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 3 | **功能名稱** | 爬蟲系統 |
| **操作說明** | 透過得知反內容的標籤進行比對後，進行爬蟲 | | |
| **輸入值** | 反白內容 | | |
| **輸出值** | 爬蟲結果 | | |

▼表 6-2-5、程序規格書 比對系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 3.1 | **功能名稱** | 標籤比對 |
| **操作說明** | 讀取抓取的資料，根據標籤進行比對 | | |
| **輸入值** | 反白內容的標籤 | | |
| **輸出值** | 反白內容的xpath | | |

▼表 6-2-6、程序規格書 爬蟲系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 3.2 | **功能名稱** | 爬蟲 |
| **操作說明** | 讀取比對結果，進行爬蟲 | | |
| **輸入值** | 反白內容的xpath | | |
| **輸出值** | 爬蟲結果 | | |

▼表 6-2-7、程序規格書 回傳爬蟲系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 3.3 | **功能名稱** | 回傳爬蟲資料 |
| **操作說明** | 把爬好的資料傳入資料庫 | | |
| **輸入值** | 爬蟲資料 | | |
| **輸出值** | 爬蟲資料 | | |

▼表 6-2-8、程序規格書 回傳爬蟲系統

|  |  |  |  |
| --- | --- | --- | --- |
| **編號** | 3.3.1 | **功能名稱** | 查看爬蟲資料 |
| **操作說明** | 查看存放在資料庫的爬蟲資料 | | |
| **輸入值** |  | | |
| **輸出值** | 爬蟲資料 | | |

第七章 資料模型

7-1 實體關聯圖(Entity relationship diagram)。



7-2 資料字典(Data dictionary)。

▼表 7-2-1、資料字典

|  |  |  |
| --- | --- | --- |
| **資料表編號** | **資料表名稱** | **資料表名稱(中文)** |
| **T01** | **scraping** | **爬蟲** |
| **T02** | **user** | **使用者** |

第八章 資料庫設計

8-1 資料庫關聯表(Constraints)。

8-2 表格及其Meta data

▼表 8-2-1、T01 Meta data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T01 scraping 爬蟲資料 | | | | |
| 欄位名稱 | 欄位中文名稱 | 資料型態 | 資料長度 | 主索引 |
| user\_id | 使用者ID | INT |  | Y |
| scrap\_time | 爬取時間 | DATE |  |  |
| scrap\_data | 爬蟲資料 | LONGTEXT |  |  |
| url | 網址 | VARCHAR | 45 |  |

▼表 8-2-2、T02 Meta data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| T02 user 使用者資料 | | | | |
| 欄位名稱 | 欄位中文名稱 | 資料型態 | 資料長度 | 主索引 |
| user\_id | 使用者ID | INT |  | Y |
| account | 帳號 | VARCHAR | 15 |  |
| password | 密碼 | VARCHAR | 15 |  |
| username | 使用者名稱 | VARCHAR | 25 |  |

第九章 程式

9-1 軟體架構與程式清單。

第十章 測試模型

10-1 測試計畫：

第十一章 操作手冊

第十二章 使用手冊

第十三章 感想

10856040鍾皓年:

10856010郭宗翰:

10856030彭鈺程:

10856031彭鈺達:這次的專題非常有挑戰性，學習新的知識，配合團隊中其他成員的時間，還要一起花時間討論，確實能增進我的團隊合作，並讓我的程式設計更進一步。

第十四章 參考資料

https://www.youtube.com/watch?v=KBLQ7GJLlQE&t=453s

**附錄**

|  |  |
| --- | --- |
| **評審建議事項** | **修正情形** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |