

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

iCRAWLER



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
1-1 背景介紹	1
1-2 動機	1
1-3 系統目的與目標	1
1-4 預期成果	1
第二章 營運計畫	2
2-1 可行性分析	2
2-2 商業模式－Business model.....	4
2-3 市場分析－STP	5
2-4 競爭力分析－SWOT.....	6
第三章 系統規格	7
3-1 系統架構	7
3-2 系統軟、硬體需求與技術平台	7
3-3 開發標準與使用工具	8
第四章 專案時程與組織分工	9
4-1 專案時程	9
4-2 專案組織與分工	10
第五章 需求模型	12

5-1	功能分解圖 (Functional decomposition diagram)	12
5-2	需求清單：	13
第六章	程序模型	14
6-1	資料流程圖(Data flow diagram) °	14
6-2	程序規格書(Process specification) °	17

表目錄

第二章 營運計畫	2
表 3-3-1 開發標準與使用工具.....	6
第三章 系統規格	7
表 3-2-1 系統軟、硬體需求.....	7
表 3-3-1 開發標準與使用工具.....	8
第四章 專案時程與組織分工	9
表 4-1-1、專案時程.....	9
表 4-2-1、專案組織與分工.....	10
第六章 程序模型	14
表 6-2-1、程序規格書.....	17
表 6-2-2、程序規格書.....	17
表 6-2-3、程序規格書.....	17
表 6-2-4、程序規格書.....	18
表 6-2-5、程序規格書.....	18
表 6-2-6、程序規格書.....	18
表 6-2-7、程序規格書.....	19
表 6-2-8、程序規格書.....	19

圖目錄

第五章 需求模型	12
圖 5-1、功能分解圖	12
第六章 程序模型	14
圖 6-1-1、系統環境圖	14
圖 6-1-2、圖 0	15
圖 6-1-3、圖 1	16

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、AI、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，因此需要一種自動化的方式來收集、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。如果是用人工的方式蒐集資料，搜集上萬筆（或者更多）的資料可不是一天就能完成的，而爬蟲卻可以自動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

主要利用爬蟲去更快地抓取資料，讓爬蟲程式自動化，利用 AI 技術，能自動分析網頁架構

1-4 預期成果

透過培養機器人，讓 USER 在任何網站都可以不受任何限制的進行爬蟲，透過自動化分析，得出想要的結論

第二章 營運計畫

2-1 可行性分析

市場可行性：在市場方面，網頁爬蟲是一個非常有應用價值的技術，能夠幫助使用者自動化地瀏覽和擷取網頁上的資訊。因此，將這項技術應用於提供使用者自由選取網站並爬取資料的網頁介面，有著許多應用場景和市場潛力。首先，對於需要搜集大量網頁資訊的公司或組織而言，這個專題提供了一個自動化的方式，能夠減少人力和時間成本。例如，市場調查公司需要搜集大量網站上的商品資訊，以做出更準確的市場分析報告；或是網路安全公司需要監控特定網站上的漏洞和風險，以提高網路安全性等。此外，對於個人使用者而言，這個專題也提供了一個便捷的工具，讓使用者可以自由選取網站並爬取想要的資訊。例如，一些學術研究人員需要搜集網站上的文獻和資料，以便進行進一步的研究；或是一些愛好者需要搜集網站上的影片和圖片，以滿足自己的興趣和需求。總結來說，這個專題在市場上有著廣泛的應用價值，能夠滿足不同用戶的需求，且有著不斷擴大的應用範圍和市場潛力。

財務可行性：財務可行性是這個專題實施的一個重要方面。考慮到實施這個專題所需的資源和成本，以及預期的營收和利潤，以下是一些財務可行性方面的考慮：首先，對於實施這個專題所需的資源和成本，主要包括人力、硬體、軟體和營運費用等方面。這些成本都需要事先估算和預算，以便進一步確定這個專題的財務可行性。其次，對於預期的營收和利潤，主要來自於使用者的付費訂閱和廣告收入等方面。使用者可以按月或按年付費，以使用網頁介面提供的爬蟲服務。此外，也可以通過與廣告商合作，通過網頁介面上的廣告投放獲得收入。這些收入預測需要根據市場調查和市場需求來進行預測和評估。最後，在考慮財務可行性時，需要考慮到競爭對手和市場

趨勢等方面。確定專題在市場上的差異化優勢，以便應對市場競爭和變化，進一步確定財務可行性。綜合考慮以上方面，可以進一步進行財務模型的建立和評估，以確定這個專題的財務可行性。

2-2 商業模式－Business model

商業模式：訂閱服務模式：用戶訂閱服務，按照會員等級提供相對應的服務。這個商業模式可以通過提供不同的訂閱方案，讓用戶根據自己的需求選擇適合自己的方案，購買會員的用戶可以獲得更高速的下載速度與服務，提供給需要客製化服務的店家客製化服務。

付費數據模式：提供用戶爬取數據的服務，收取一定的費用。這個商業模式可以通過對不同的數據及每個網站所爬取資料的難易度進行定價，讓用戶根據自己的需求選擇要爬取哪些數據，以及爬取的數量。

廣告模式：通過向用戶推送廣告，獲取收益。這個商業模式可以通過分析用戶的爬蟲需求，向用戶展示相關廣告，從中獲取收益。

諮詢服務模式：為需要特定數據的客戶提供定制化的諮詢服務，協助客戶分析數據並提供專業的建議。這個商業模式可以通過對用戶需求的了解，向客戶提供定制化的諮詢服務。

資料庫模式：建立一個完整的數據庫，讓用戶可以進行搜索和查詢，並且提供一些高級搜索和查詢功能。這個商業模式可以通過向用戶收取訪問費用或訂閱費用，或者通過賣出一些特定的數據報告來賺取收益。

聯營模式：與其他相關公司進行聯營，將自己的數據作為他們的產品之一，以增加收益。例如，與 SEO 公司合作，將自己的數據作為他們 SEO 服務的一部分銷售。

數據分析工具模式：除了提供數據之外，還提供一些工具和分析功能，讓用戶可以更好地理解和分析自己的數據。這個商業模式可以通過向用戶收取訪問費用或訂閱費用，或通過銷售一些高級分析報告來賺取收益。

2-3 市場分析—STP

S： 地域區分：爬蟲機器人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，如大型企業、大型研究機構等。 市場專業化：提供一種爬蟲機器人產品，針對不同的目標市場。 無差異行銷：針對所有的市場提供相同的爬蟲機器人產品。

P： 功能利益面：爬蟲機器人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，並且能夠檢測網路上的變化，提供即時資訊。 價格比較面：在爬蟲機器人市場中，價格比較是非常重要的因素，因為許多競爭對手也提供類似的產品。因此，要考慮定價策略與價格彈性，確保價格合理且具有競爭力。 適用情境面：爬蟲機器人適用於各種行業，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，也可以為企業提供即時的市場情報和競爭情報。 文化象徵面：爬蟲機器人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。此外，爬蟲機器人還有助於促進數據的共享和流通，這與現代社會中重視信息共享和開放的價值觀相符。 模仿競爭者：在爬蟲機器人市場中，模仿競爭者的行為是常見的，但同時也需要保持創新和差異化，以吸引更多的消費者和擴大市場佔有率。

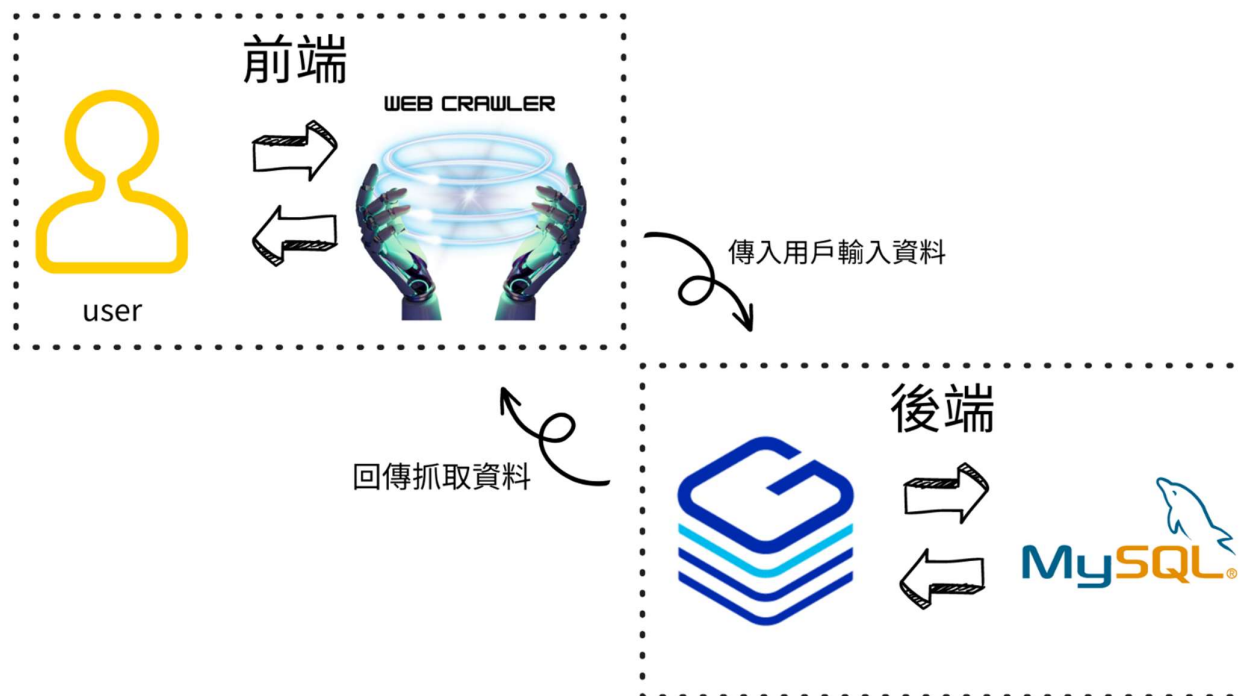
2-4 競爭力分析－SWOT

▼表 3-3-1 開發標準與使用工具

<div>內部因素</div> <div>外部因素</div>	優勢 (Strengths)	劣勢 (Weaknesses)
	<p>自動化程度高：爬蟲機器人可以自動收集資料，節省人力成本且提高工作效率。 精確性高：爬蟲機器人可以快速且準確地收集資料，避免人為錯誤。 彈性大：爬蟲機器人可以針對不同網站進行自定義設置，擁有較高的彈性度。</p>	<p>需要技術支持：開發和維護爬蟲機器人需要具備較高的技術水平，這也增加了使用成本。 可能受限於網站規則：部分網站可能設置了反爬蟲機制，導致爬蟲機器人無法正常收集資料。 安全風險：爬蟲機器人可能會被惡意使用，導致資料洩漏等安全問題。</p>
機會 (Opportunities)	SO (Strengths-Opportunities)	WO (Weaknesses-Opportunities)
<p>巨大市場：隨著大數據和人工智慧的發展，對於自動化資料收集的需求逐漸增加，為爬蟲機器人帶來了市場機會。 新興行業的需求：部分新興行業如社交媒體監測、競爭情報收集等，需要大量的資料收集和分析，爬蟲機器人可以為這些行業提供解決方案。 整合其他技術：爬蟲機器人可以與其他技術如自然語言處理、機器學習等進行整合，為使用者提供更全面的解決方案。</p>	<p>基於爬蟲機器人的自動化和精確性高的優勢，可以針對新興行業的需求，如社交媒體監測、競爭情報收集等，提供解決方案，擴大市場佔有率。</p>	<p>基於爬蟲機器人的可定制化和彈性大的優勢，可以與其他技術如自然語言處理、機器學習等進行整合，為使用者提供更全面的解決方案，增加附加價值。</p>
威脅 (Threats)：	ST (Strengths-Threats)	WT (Weaknesses-Threats)
<p>法律風險：部分網站可能設置了反爬蟲機制，這可能會涉及到法律風險。 技術風險：爬蟲機器人的技術不斷更新</p>	<p>基於爬蟲機器人的自動化和精確性高的優勢，可以針對市場競爭激烈的威脅，提供更優秀的產品和服務，增強競爭力。</p>	<p>基於爬蟲機器人的技術需要不斷更新的劣勢，可以通過合作或收購其他技術公司，補充技術優勢，增強產品的創新能力。</p>

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
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	112 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 ○次要負責人

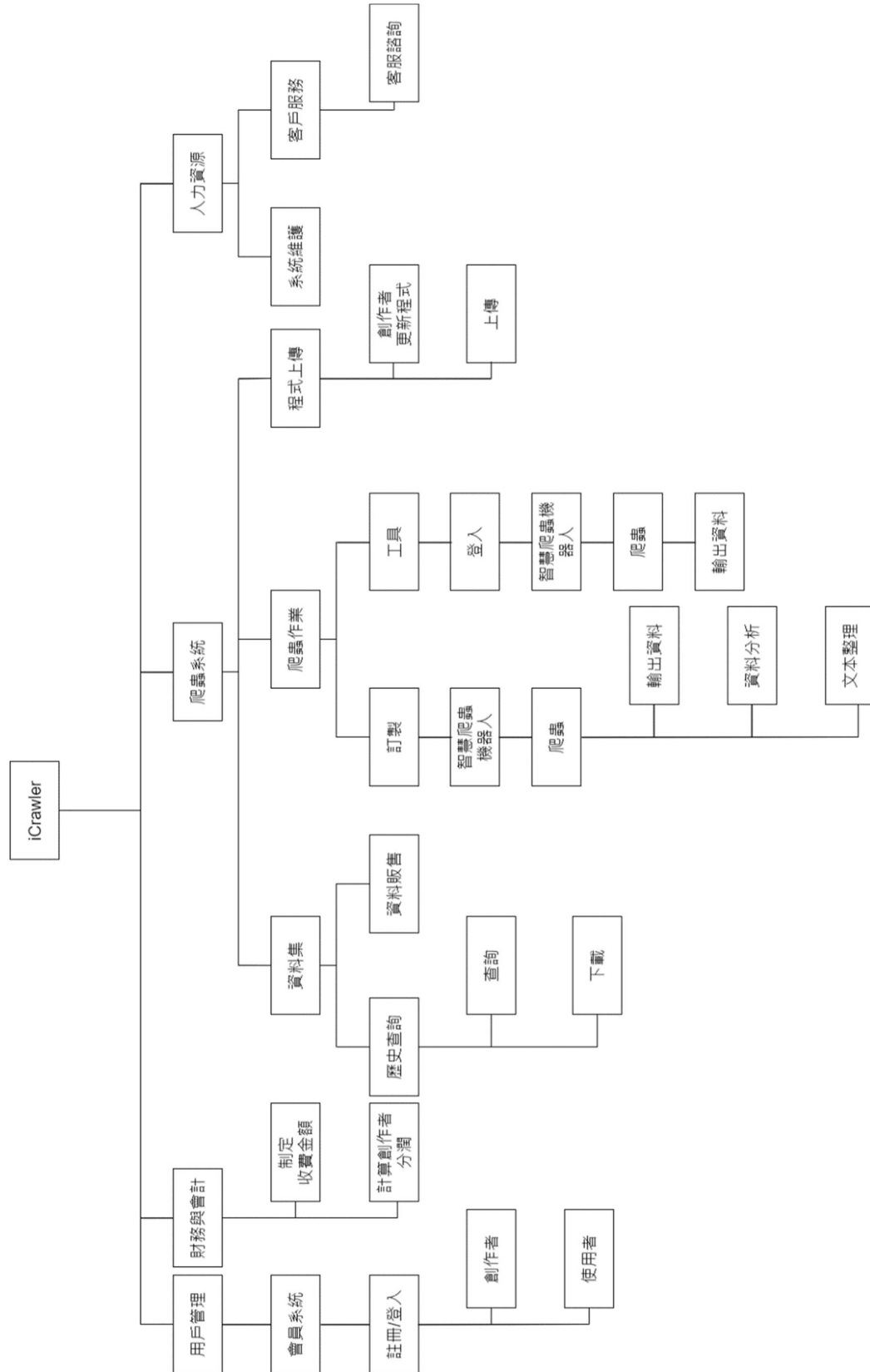
項目/組員		10856041 鍾皓年	10856010 郭宗翰	10856030 彭鈺程	10856031 彭鈺達	10856032 陳永祥
後端開發	資料庫建置		●		○	
	伺服器架設	●	○	○		
	Flask 架構	●		○		
	爬蟲功能	●	○			
	會員系統		○	○	●	
前端開發	HTML 設計	○		●		
	JavaScript 設計	●	○			
	jQuery 設計	●				
	CSS 設計	○		●	○	
美術設計	UI/ UX	○				●
	Web/APP 介面設計	●				○
	色彩設計		○			●
	Logo 設計			○		●
	素材設計				○	●
文件撰寫	統整	●				
	第 1 章 前言		●			
	第 2 章 營運計畫			○	●	
	第 3 章 系統規格	○		●		
	第 4 章 專題時程與組織分工	●		○	○	
	第 5 章 需求模型			○	●	
	第 6 章 程序模型		●		○	
	第 7 章 資料模型			●		
	第 8 章 資料庫設計		●			
	第 9 章 程式	●	○		○	
	第 10 章 測試模型			○	●	
	第 11 章 操作手冊	●			○	○
	第 12 章 使用手冊	●	○			○
報告	簡報製作		○			●
	影片製作	○			○	●

	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：

需求清單：功能需求

- 1.1 支持用戶自主選擇爬取網站和版面
- 1.2 支持用戶自主設定爬取的時間間隔和範圍
- 1.3 支持對爬取的數據進行過濾、清洗、轉換等操作
- 1.4 提供結構化的數據格式，便於用戶進行數據分析和應用
- 1.5 支持對爬取過程進行監控和管理
- 1.6 支持用戶自主設定爬取的代理 IP 和其他配置

非功能需求

- 2.1 可靠性：爬取過程中出現錯誤時，需要能夠自動恢復或提示用戶處理
- 2.2 效率：需要支持高效的爬取速度，並能夠應對高併發的情況
- 2.3 安全性：需要對用戶數據進行加密和保護，防止數據外洩或損壞
- 2.4 可擴展性：需要能夠方便地擴展爬取的網站和功能

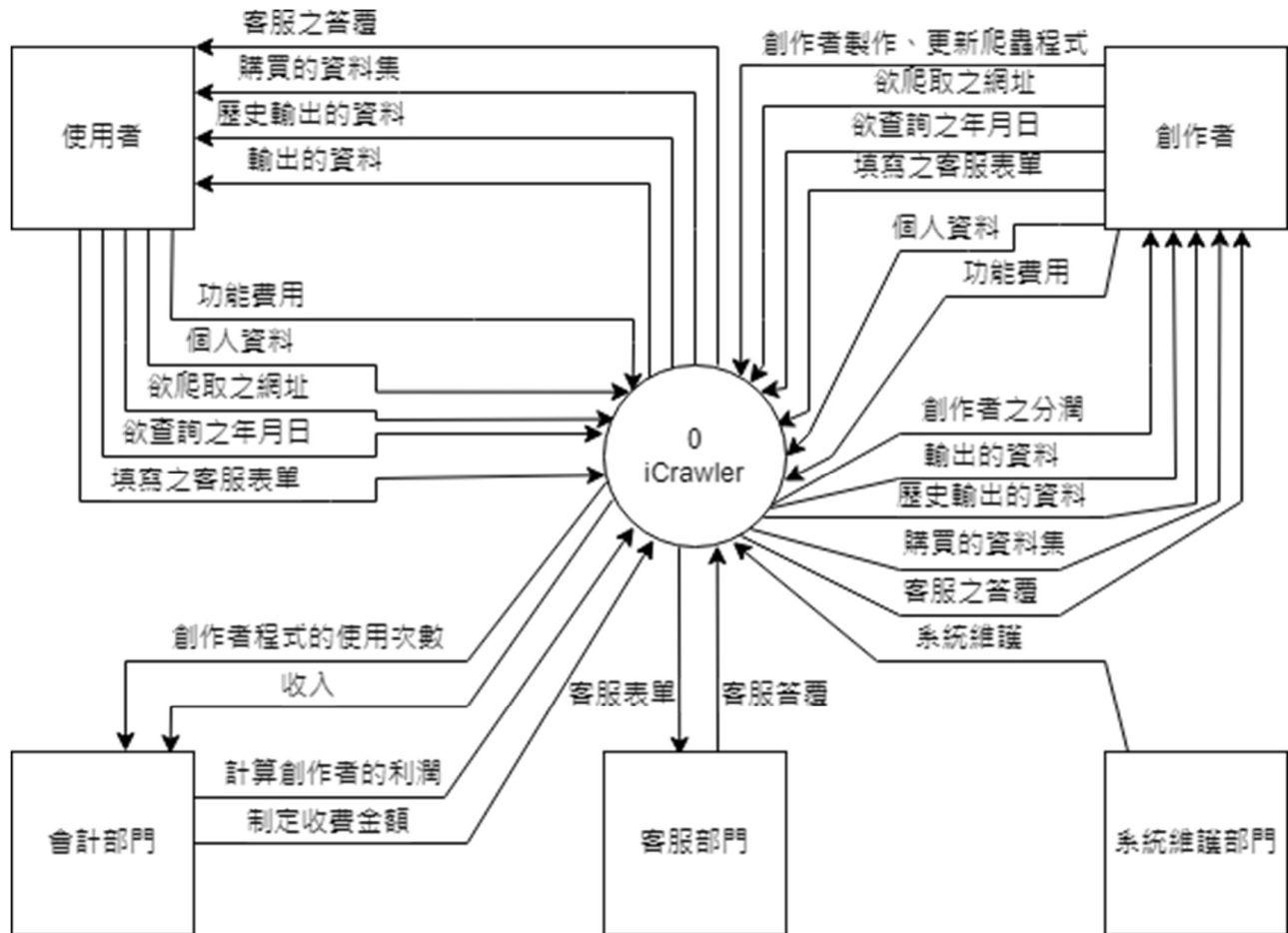
限制條件

- 3.1 法律法規：必須遵守當地和國家的相關法律法規，如保護個人隱私等
- 3.2 技術限制：需要適應不同網站的技術架構和防盜採取的技術手段

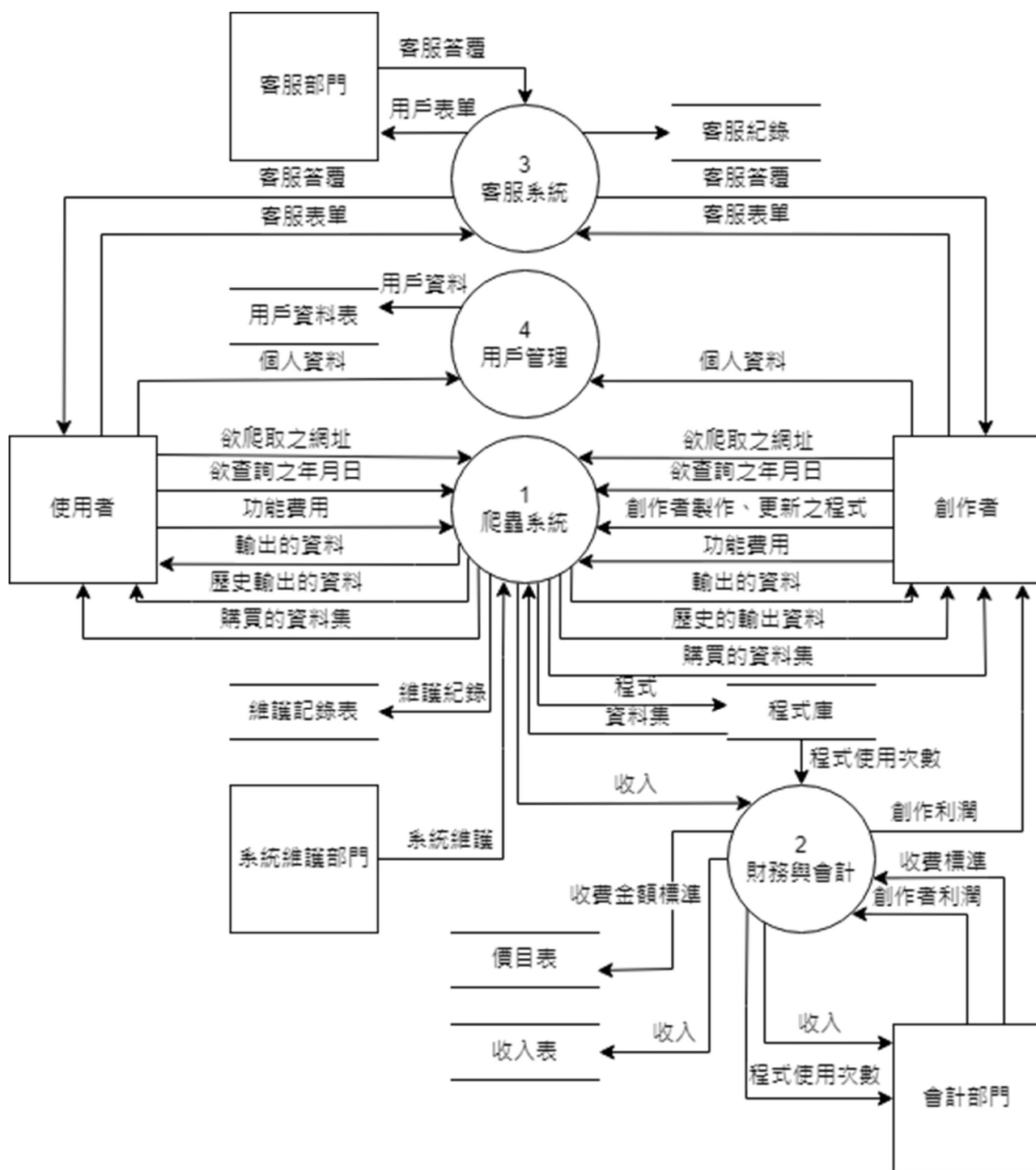
第六章 程序模型

6-1 資料流程圖(Data flow diagram)。

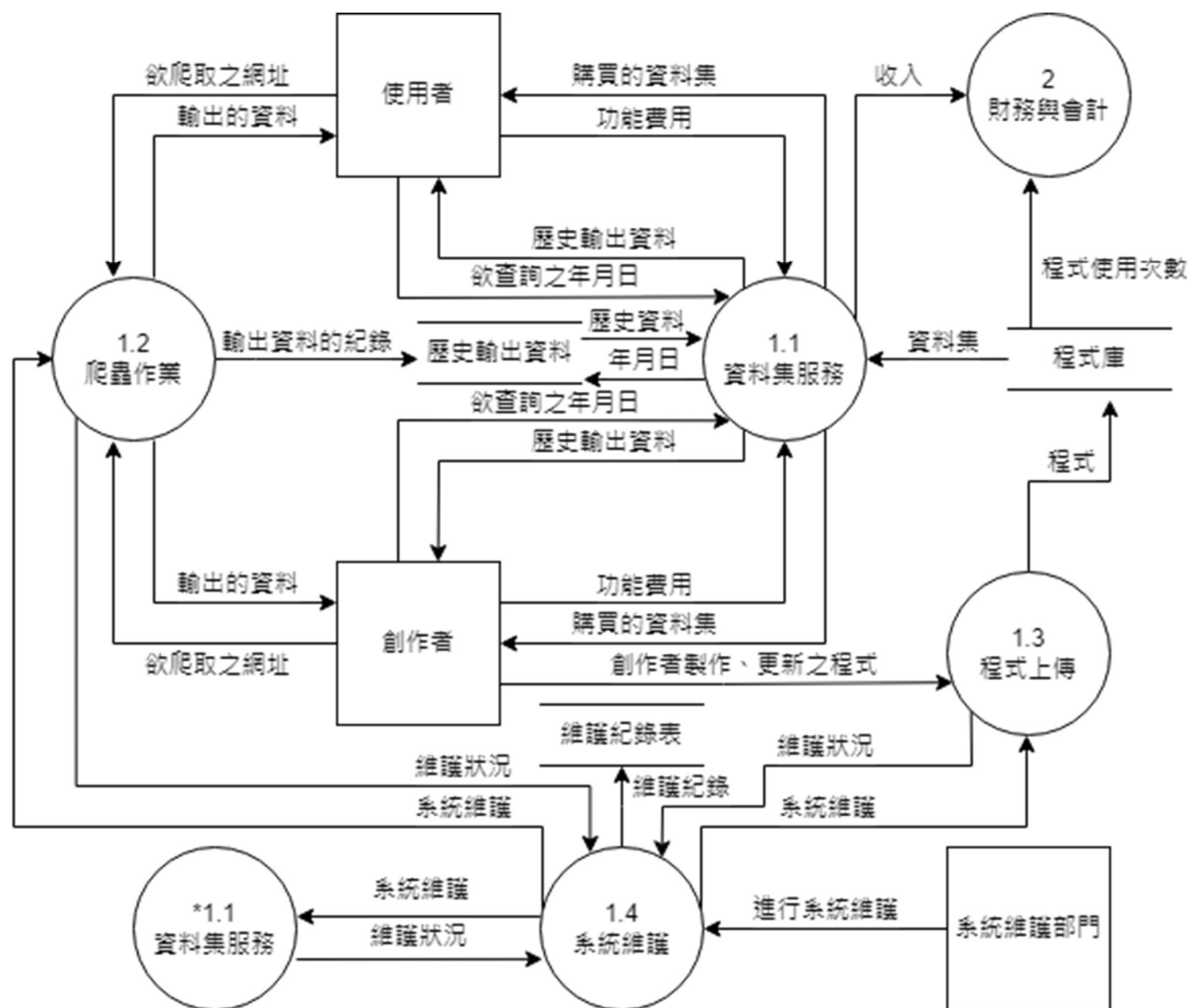
▼圖 6-1-1、系統環境圖



▼圖 6-1-2、圖 0



▼圖 6-1-3、圖 1



6-2 程序規格書(Process specification)。

▼表 6-2-1、程序規格書

編號	1	功能名稱	爬蟲系統
操作說明	進行爬蟲		
輸入值	欲爬取之網址、日期、費用、程式		
輸出值	資料		

▼表 6-2-2、程序規格書

編號	1.1	功能名稱	資料集
操作說明	存資料、查詢、販售		
輸入值	資料、日期、金額		
輸出值	資料		

▼表 6-2-3、程序規格書

編號	1.2	功能名稱	爬蟲作業
操作說明	爬取資料		
輸入值	欲爬取之網址		
輸出值	資料		

▼表 6-2-4、程序規格書

編號	1.3	功能名稱	程式上傳
操作說明	上傳程式		
輸入值	程式		
輸出值	上傳成功、失敗		

▼表 6-2-5、程序規格書

編號	1.4	功能名稱	系統維護
操作說明	維護系統		
輸入值	欲維護之程式、網頁		
輸出值	維護狀態		

▼表 6-2-6、程序規格書

編號	2	功能名稱	財務與會計
操作說明	訂製價目、計算利潤		
輸入值	收入		
輸出值	統整後的收入		

▼表 6-2-7、程序規格書

編號	3	功能名稱	客服系統
操作說明	服務客戶		
輸入值	客服表單		
輸出值	客服答覆		

▼表 6-2-8、程序規格書

編號	4	功能名稱	用戶管理
操作說明	管理用戶		
輸入值	用戶資料		
輸出值	更新的用戶資料		

附錄

評審建議事項	修正情形