

致理技術學院

資訊管理系 專題期末報告

Show your life

S
h
o
w

y
o
u
r

l
i
f
e

學生：29710519 游濬璋
29710516 莊舒晴
29710518 陳俊維
29710513 吳昆逸
29710536 曾俊傑
29710538 楊明勳

指導老師：林曉雯 老師

中華民國 98 年 12 月

致理技術學院

資訊管理系 專題期末報告

Show your life

學生：29710519 游濬璋
29710516 莊舒晴
29710518 陳俊維
29710513 吳昆逸
29710536 曾俊傑
29710538 楊明勳

本成果報告書經審查及口試合格特此證明。

指導老師：_____

中華民國 98 年 12 月

授 權 書

實務專題研究授權書

本授權書所授權之實務專題研究為 游濬璋、吳昆逸、莊舒晴、陳俊維、曾俊傑、楊明勳 共 6 人，在致理技術學院資訊管理系 98 學年度第 1 學期完成資管實務專題。

實務專題名稱：Show your life

同意 不同意

本組同學共 6 人，皆同意著作財產權之論文全文資料，授予教育部指定送繳之圖書館及本人畢業學校圖書館，為學術研究之目的以各種方法重製，或為上述目的再授權他人以各種方法重製，

不限地域與時間，惟每人以一份為限。

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。上述同意與不同意之欄位若未鉤選，該組同學皆同意視同授權。

指導教授姓名：

專題生簽名：

學號:29710519

(親筆正楷)

學號:29710513

學號:29710516

學號:29710518

學號:29710536

學號:29710538

中華民國 98 年 12 月

致 謝

林曉雯老師，她是我們的專題老師，從專題開始到現在，已經經過很長一段時間，我們要感謝林曉雯這段時間的帶領與指導，從一開始什麼都是零的狀況下，不厭其煩的和我們溝通與討論，灌輸我們許多我們原本不懂的觀念，讓我們慢慢了解與接受，在這段過程中，需要許多的時間與精神，老師不但沒有閒過累，還是用盡心力的教導我們，在這整段過程中，即使我們的問題再多，林曉雯老師還是會不厭其煩的聽我們發問，也給我們適當的建議，並且領導我們該如何去解決我們的問題，這也讓我們更有力量繼續做下去，我們非常謝謝曉雯老師。

曹主任、劉永麟老師，是在我們期初發表的時候的兩位評審老師，在我們期初發表的時候，兩位評審老師對我們的指教，更是讓我們成長更多，讓我們釐清我們的方向，是否有偏了方向，還是不夠好的地方，雖然很多的評批指導都非常的犀利，但是也讓我們更快了解我們錯的地方，謝謝兩位評審老師對我們的批評指導與建議。

蔡嘉偉同學，我們真的要好好感謝他，他自己也有自己的專題要努力，而對於我們請教他的問題，他也用很多時間和精神來幫助我們，也沒要求任何回報，真的是一位很好的同學，這麼努力的幫我們，真的非常謝謝他。

游濬璋同學，我們的組長，其實要領導一個團隊，是需要學習很多東西的，可能在過程中，小組成員可能有大大小小不同的爭執與問題，但是他還是帶領我們一起到現在，總是在最需要的時候，幫我們一把，不管每個人負責的部分是什麼，只要遇到問題，他都會共同參予一起解決，雖然過程中爭執不斷，也沒特別對他說過什麼感謝的話，在此真的要謝謝他，這麼用心努力的帶著我們一起走到現在，今後也要一起再度努力下去，組長，謝謝你。

整組的組員，有了這些人，才會形成這個團隊，在學校這個階段中，正是我們該學習如何與團隊一起學習努力的過程，在這段時間中，大家有問題一起克服解決，就算有再多的意見不相同，理念無法溝通，到最後大家還是會團結在一起努力，不分日夜的一起在努力這份專題，在整個暑假也投入了許多心力，大家也因此學會了互助合作，不願放棄的精神，謝謝各位組員的配合與努力。

當然還有許許多多幫助我們的老師、學長姐、同學們，感謝你們這麼樣的幫助我們，就算只是一些芝麻小事，或是回答我們簡單的問題，在我們請教你們問題時，總是得到你們的熱心回答與幫助，即使是我們的問題太過於簡單而得到嚴厲的教誨，對我們來說是一種磨練；或者是你們不厭其煩面帶微笑的教導我們，幫助我們一起解決問題，我們深深的感恩與感謝。重要的不在於結果，而是在這樣的過程中，我們學習到了什麼，得到更多知識，也得到更多與人相處的種種，團隊的合作精神更為重要，過程是美好的，即使有再多的不開心，都是學習與成長，大家要一起加油努力到最後。這份專題的製作，得到這麼多人的幫助，我們深深的感謝你們，無窮境的感謝想對你們說，總之謝謝你們大家的幫忙，有你們才有今天的我們，謝謝大家。

摘要

現今在網路社會的人與人之間關係薄弱，尤其讓一個想進入新職場環境的人來說，縮短這距離較為困難，若是能讓畢業的任何一位致理的學生，認識已經在職場上的學長、姐，獲得職場上的協助，減少菜鳥進入新職場有被欺負或是有任何問題，都能提出問題藉由此專題來獲得，提升從原本不認識的人，到能有共同的話題，甚至一起出去，一起解決在外來的困難。

也因為如此，本組將決定以 ” Show Your Life ” 來做為我們專題的題目，一方面是改變多數使用者不需再去 Google 他的底，藉由 Show Your Life 替您自動整理資料，分析及篩選資料，最後再透過此專題幫您自動填入資訊。

關鍵字:人工輸入、自動化輸入、分類、手動輸入、Data Mining、分析資料、篩選資料。

目 錄

授 權 書.....	i
誌 謝.....	ii
摘 要.....	iii
目 錄.....	iv
圖 目 錄.....	vi
表 目 錄.....	vii
第一章 緒論.....	1
第一節 研究背景.....	1
第二節 研究動機.....	1
第三節 研究目的.....	2
第四節 研究範圍與限制.....	2
第五節 操作性定義.....	3
第六節 章節結構.....	4
第二章 文獻探討.....	5
第一節 相關研究應用技術與理論.....	5
第二節 Google API 的理論.....	10
第三節 資料探勘(Data Mining)理論與應用.....	12
第三章 系統研究方法.....	17
第一節 研究流程.....	17
第二節 研究方法.....	18
第三節 SWOT 方法.....	22
第四章 預期研究成果.....	23
第一節 系統功能.....	23
第二節 系統特色.....	23
第三節 使用對象.....	24
第四節 使用環境.....	24
第五節 開發工具.....	24
第五章 結論.....	26
第一節 預期研究效益.....	26
壹. 半自動輸入的方式.....	26
貳. 利用 Agent 及 Data Mining，讓搜尋更準確.....	26
參. 高度移植性.....	26
第二節 預期研究限制.....	26
壹. 發展設備不完備.....	26
貳. 發展時間不夠多.....	26
第六章 分工執掌和進度表.....	27
第一節 分工執掌.....	27
第二節 進度表.....	28

參考文獻.....	29
附錄一 Show your life「使用電腦生活習慣」問卷.....	35

圖目錄

圖 1. 行動型代理人之基本作業圖.....	17
圖 2. 行動代理人流程圖.....	18
圖 3. Agent 模型圖.....	19
圖 4. Google 搜尋網頁流程圖.....	19
圖 5. 研究流程圖.....	17
圖 6. 大多數學生每天使用電腦的平均時數圖.....	18
圖 7. 使用電腦比例圖.....	19
圖 8. YAHOO 及 GOOGLE 的比例圖.....	19
圖 9. 搜尋資料的完整度.....	20
圖 10. 資料輸入方式比例圖.....	20
圖 11. 自動輸入的接受度.....	21
圖 12. 自動輸入與微調的接受度.....	21
圖 13. 填寫問卷年齡層.....	22

表目錄

表 1 Google 主要關鍵字運算子.....	11
表 2 SWOT.....	22
表 3 分工執掌.....	27

第一章 緒論

在本章中我們將介紹研究背景、研究動機、研究目的、研究範圍、操作性定義、章節結構。

第一節 研究背景

現今在網路社會的人與人之間關係薄弱，尤其讓一個想進入新職場環境的人來說，縮短這距離較為困難，若是能讓畢業的任何一位致理的學生，認識已經在職場上的學長、姐，獲得職場上的協助，減少菜鳥進入新職場有被欺負或是有任何問題，都能提出問題藉由此專題來獲得，提升從原本不認識的人，到能有共同的話題，甚至一起出去，一起解決在外來的困難，如此一來使用者也將大大的提高使用的意願。而若是本校的學長、姐身為主管，有找技術人員的需求，也能藉由這裡尋求資深的校友，甚至選出值得且適合培育的人才，也能使學弟、妹在以後工作的路上能夠更順暢，也提升本校的素質，更因為提供了這樣的平台，讓校友們拉近彼此之間的距離。

在本研究之中我們提出了三個系統開發研究動機如下：

第二節 研究動機

一、人與人之間關係薄弱。舉例來說，一個想進入新職場的學生，因為網路上的資訊不足，缺乏他需要的相關資料，而造成新鮮人的不確定與不敢嘗試而退縮。但很有可能企業內部的人員有本校的畢業生，但他們在校內卻沒有接觸的機會，而達不到互相幫助及提拔的效用。

二、在外部資料搜索有著資料資訊的不確定性，而降低使用的效率。先透過使用者提供的資訊，利用一套分類規則將資訊導向為整理會員的相關資料，造就了這個動機。

三、現行個人資料的輸入仍然使用傳統的人工建置，造成使用者因為填入資料過於繁瑣，而降低使用系統的意願，但依照會員的相關資料所使用到的實用性頗高。

第三節 研究目的

目的一：

運用此平台，提高校友之間的緊密性，達到人際關係與互助的效用，提升學生面對未來的挑戰性，也提升學校未來的素質，在企業上有更卓越的表現。

目的二：

減少新鮮人面對陌生的社會職場的恐懼，來提升勇於嘗試的心態。幫助企業上為主管的資深校友，透過學校這個溫馨的地方，找尋相關技能技術人才。

目的三：

運用資料探勘技術，將校友們做更詳細的分類分群，以便達到搜尋人才的目標。

本專案訂定研究範圍如下：

第四節 研究範圍與限制

本專案以研究動機與目的訂定研究範圍與限制如下：

一、研究範圍

臺灣網路資訊業已蓬勃發展到飽和程度，台灣現階段資料來源分為搜尋引擎及 PTT，以搜尋引擎為比較大宗，因為使用方便，但是現在的搜尋引擎已經無法滿足網路大眾的需求，所以網路業者紛紛做各種搜尋方式的技術研發，例如 Google Map、Google API，行動型代理人等技術來提升網路大眾使用率。因此本研究將運用資料探勘技術來發展搜尋及整合的技術。

二、限制

本研究再進行時可能產生以下幾點限制:

- (1)本研究目的對內部資料制定搜尋功能，但會自動更改內部資訊。
- (2)本研究目的只對有內部資料庫使用權的相關人士。

第五節 操作性定義

壹、Agent 的技術

- 代理人必須擁有更便利以及多樣性的溝通方式，作為人類與它溝通的介面。
 - 語音輸入
 - 影像辨識
- 更友善的操作介面來服務一般使用者而非專業電腦人士。
 - Ex:小幫手, MS-Agent
- 多個代理人之間也可能彼此傳達有”意義”的訊息，來尋求其他代理人的協助。

貳、Google API 的技術

Google搜索數據API使用戶端和網路應用程式可以搜索函式定義和示例代碼的公共原始程式碼。

- 以下是您可以使用Google搜索數據API執行的操作：
 - 為您喜歡的IDE創建一個外掛程式，用於在代碼搜索索引中搜索相關的示例代碼。
 - 創建使用 Google 搜索查找函式定義和示例代碼的用戶端應用程式

參、資料探勘(Data Mining)的技術

分類按照分析對象的屬性分門別類加以定義，建立類組(class)。

推理 根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。

預測根據對象屬性之過去觀察值來推估該屬性未來之值。

關聯分組從所有物件決定那些相關物件應該放在一起。

同質分組將異質母體中區隔為較具同質性之群組(clusters)。

第六節 章節結構

在此篇文件中總共分為六章與最後的參考文獻，分別為第一章：序論、第二章：文獻探討、第三章：系統研究方法、第四章：預期研究成果、第五章：結論與最後的第六章：分工執掌和進度表。

第二章 文獻探討

第一節 相關研究應用技術與理論

本系統需要透過內部資料庫以及外部網路的資料擷取技術，因此針對這兩個方向，我們去深入的探討了 AGENT、GOOGLE API 以及 WEB SPIDER 等技術。將 AGENT 應用在內部的蒐集資料，而另兩種技術針對外部的資料擷取做研究。然而最後獲取的資訊必須進行整合以及正確的取得所需資訊，因此還介紹了 DATA MINING 的相關探討。以下將針對上述四種技術進行深入的文獻研究與探討。

壹、Agent 代理人的技術與應用

何謂行動型代理人？

行動型代理人是一種軟體程式主要是處理例行性、反覆性和資訊搜尋的工作幫助使用者克服大量資訊的處理過程，直接分析出多個替代方案，提供使用者制定決策之參考，例如：「郵件代理人」。

- 使用者可以設定郵件規則
- 由代理人對大量的電子郵件做分類
- 使用者可按郵件的優先順序來處理郵件內容
- 使用者也可封鎖寄件者，拒絕接收某來源處所發送之信件

行動型代理人的特性

- 自主性

一般程式只能直接處理指令的動作，而行動型代理人則可判斷環境的變化，不需要其他人事物的介入即可自動進行事件的處理。

- 時間持續性

一般程式在處理完畢後即終止程式的執行與待命，直到再度被起始。而行動型代理人可由使用者設定目標或工作，代理人會針對目標持續檢視事件並在適當的時機執行工作。

- 反應性

行動型代理人是依據環境的變化來反應執行的動作，其規則模式就像 ” 若 A 發生，則執行 B ” ，其中 “ A “就是行動型代理人持續檢視的事件對象目標導

向。

行動型代理人可由使用者設定特定目標及使用者需求，代理人會進一步決定如何滿足使用者的需求，必要時代理人也會修改目標或建立自己的目標

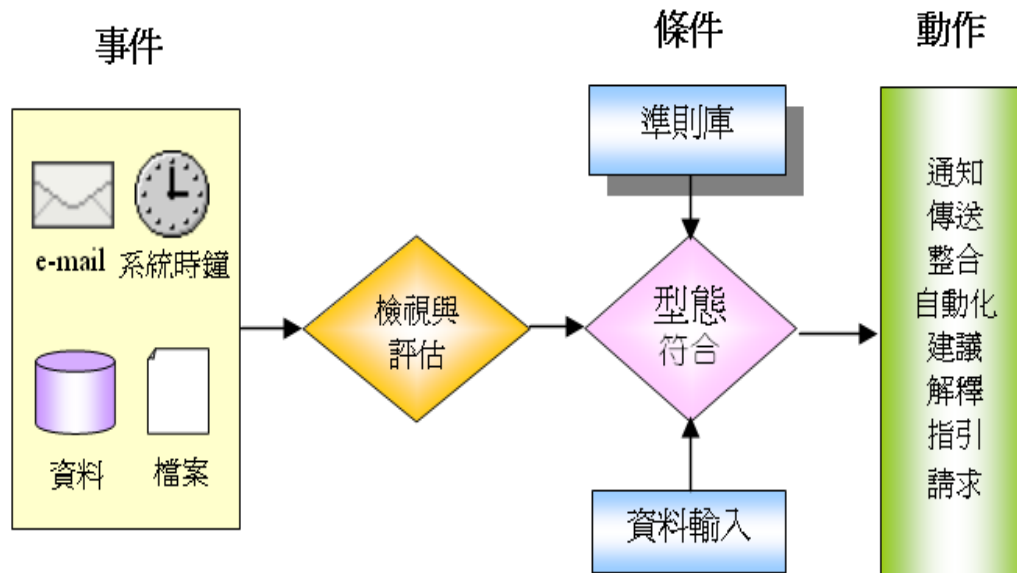


圖 1.行動型代理人之基本作業圖

行動型代理人(intelligent agents)是近十年來的新興資訊技術與概念。網際網路興起，資訊的供給與需求大幅而快速地增加，上網搜尋資料，已成為現代人習以為常的活動之一。行動型代理人主要是處理例行性、反覆性和資訊搜尋的工作幫助使用者克服大量資訊的處理過程，直接分析出多個替代方案，提供使用者制定決策之參考。

行動型代理人之類型

搜尋、過濾型行動代理人

- 使用者在龐雜的資訊中搜尋並過濾出使用者想要的資訊，並進一步提供選擇的方案給使用者參考。

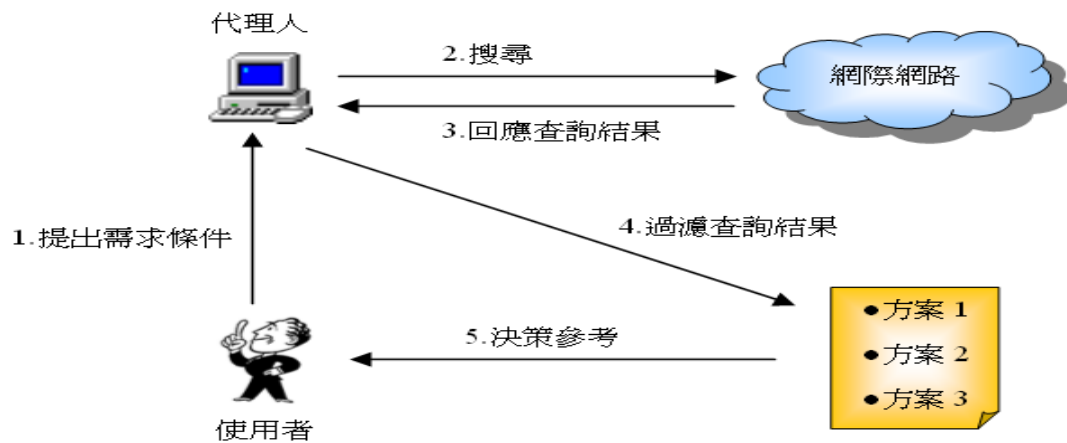


圖 2.行動代理人流程圖

Agent的理論及發展史

- 代理人的概念可以追溯到 1977 年 Hewitt 的演員模型。
 - 具有自主性,互動性及平行處理的物件。
 - 這些物件具有其自身之狀態及通訊能力。

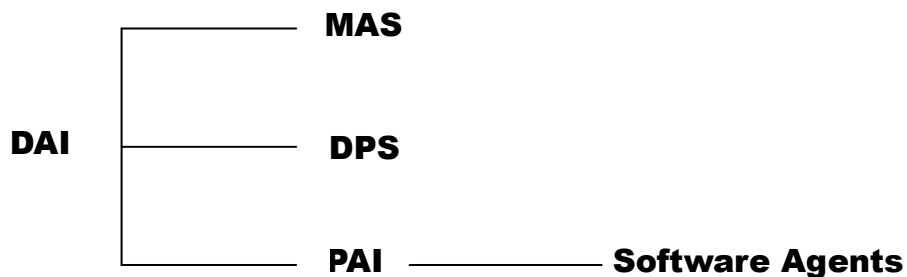


圖 3.Agent 模型圖

- 由多重代理人系統(MAS)衍生而來。
- MAS 由分散式人工行動(DAI),分散式問題解決(DPS) 及平行式人工行動(PAI)衍生而來。
- 不論優點或缺點皆繼承自 DAI 和 AI。
- John Sculley 在 1987 “知識導航” 影片中顯示了使用者們和代理人之間的關係。
- Being 由於代理人是先有理論接著才進入實作階段,而非一般的由上而下模式, 這個領域先後經歷了數次不是頗成功的產品。
- 從 1990 年開始,越來越多不同的代理人出現。
 - 網際網路,搜尋引擎,……等等。
- 越來越多的事情皆運用代理人機制。

Agent 的技術

- 代理人必須擁有更便利以及多樣性的溝通方式,作為人類與它溝通的介面。
 - 語音輸入
 - 影像辨識
- 更友善的操作介面來服務一般使用者而非專業電腦人士。
 - Ex:小幫手, MS-Agent
- 多個代理人之間也可能彼此傳達有”意義”的訊息,來尋求其他代理人的協助。

Agent 的應用未來發展

1.安全性

在安全性上,我們所關心的是要如何地去防範不只是有些行動代理人會惡意地去攻擊主機。相對地,還包括非法主機會去攻擊行動代理人等事情。如果從利用行動代理人作為工具的電子商務角度來看,我們所要擔心的是代理人所保存的資料或電子貨幣,有可能被惡意的主機所竊取或竄改之虞,所以,最近就行動代理人及 run time system 的認證機制也被許多學者作深入研究。

2.行動代理人之位置

當傳送 message 給在其他移動的代理人，有時候必須從代理人所在主機回傳 message 給原傳送端主機，因此，我們有必要知道行動代理人目前所在之位置之機制，而目前有被提出討論之方式有 3 種，其方式如下：

(1)在每次移動時，在目前之主機上註記下一個欲移動到之主機的位置，使用者則利用 trace 方式來找尋最新的位置。

(2)在每次移動時，利用 message 告知目前的位址給特定的主機。

(3)將從特定主機中所定期尋問到目前位置之 message 傳送給多數個主機。

3. design pattern

在物件導向設計裡，雖然我們利用了些物件和類別來開發，但是其組合方式可分成數十種類的 pattern；另外，就行動代理人而言，已經有代理人程式的 design pattern 以及代理人移動路徑的 pattern 等各種相關議題研究被提出。

4.藉由程式變換使狀態得以保持在 Java 行動代理人中，只要沒有變更 Java 虛擬機器，就很難將 method 的區域變數和程式計數器予以資料化。因此，更新代理人程式將區域變數和程式計數區之資料當作 instance 變數予以保持的方式也有被提出來討論研究。

5. CORBA3.0 物件傳遞

CORBA (Common Object Request Broker Architecture) 為 OMG 於 1991 年提出之物件導向分散式工作環境規格。CORBA 規定對一個物件請求仲介者 (Object Request Broker) 之標準介面作了詳細的規定。根據 OMG 的定義，一個物件請求仲介者，為一可提供分散式環境上各物件透明化的請求服務與回應接收功能的應用程式建構工具；而在 CORBA 新版本 (CORBA3.0) 的相關規範則導入一個新的傳送模式，就是把物件的複製當作是一個遠端 method 呼叫的參數來傳遞。同理，藉由把行動代理人也當作是一個遠端 method 呼叫的參數，那麼就可以很容易地完成包含執行狀態的傳送了。

Agent 扮演內部資料庫供給及搜尋的部份，而 Google API 用中扮演著搜尋國際網路上的資訊角色，其為 GOOGLE 公司目前搜尋引擎提供大量的 API 函式提供一般使用者使用，在這當中 Google 將 spider 搜尋到的網頁，建置了 Index 增加搜尋的速度，利用深度搜尋 (Depth search) 直接對 URL 裡的內容進行關鍵字分析。

第二節 Google API 的理論

壹 搜尋引擎

搜尋引擎的運作方式可分為三類：關鍵字搜尋、目錄搜尋以及匯總式搜尋 [4,5,6,9]。關鍵字搜尋引擎，又稱全文搜尋，是利用Crawler 或Robot 定期到網際網路上收集網頁存到資料庫中，然後對資料庫建立索引以提升搜尋網頁的速度。此外，如何提升網頁搜尋的準確度也是研究的重點[1,5]。此種的搜尋引擎在目標網頁有所改動的情況下，都能得到這些網頁的變動資訊，進而改變庫存網頁的內容與標題。此類搜尋引擎較為常見的有AltaVista[10]、Openfind[11]、Google[12]等。目錄搜尋，又稱主題目錄搜尋，是由人工將網站分類成階層的架構儲存[2,4]。主要的分類是透過關鍵字、連結與網站登錄者的想法進行分類。使用者可以依據所需要的類別進行搜尋，若是不清楚搜尋的方向也可利用關鍵字輸入的方式進行搜尋。此類搜尋引擎較著名的有Yahoo[13]、PChome[14]、Excite[15]等。匯總式搜尋引擎，又稱整合搜尋引擎，本身並不做搜尋，也不必建立索引。當使用者進行查詢時，匯總式搜尋引擎會將查詢的內容交給其他搜尋引擎去進行搜尋。當搜尋結束後，匯總式搜尋引擎將結果過濾、刪除重複，加以整理後，再傳回給使用者瀏覽[11]。近來搜尋引擎為了提升再利用性，常結合Web Service 以提供更多樣化的功能[3,7]。本研究便是利用匯總式搜尋引擎的概念結合Google 提供的Google API，來達成搜尋英文例句的目標。

貳 Google API

Google 提供了Web Service 的機制供需要搜尋服務的程式開發者使用[7,8]。Google API 提供了搜尋網頁、取得庫存網頁以及進行拼字檢查的功能。在使用Google API 的功能必須先透過e-mail 申請註冊碼，而一個註冊碼在一天中可對Google 呼叫1000 次(包含搜尋網頁與取得庫存網頁的次數)。每一次最多所能回傳10 筆搜尋的結果。

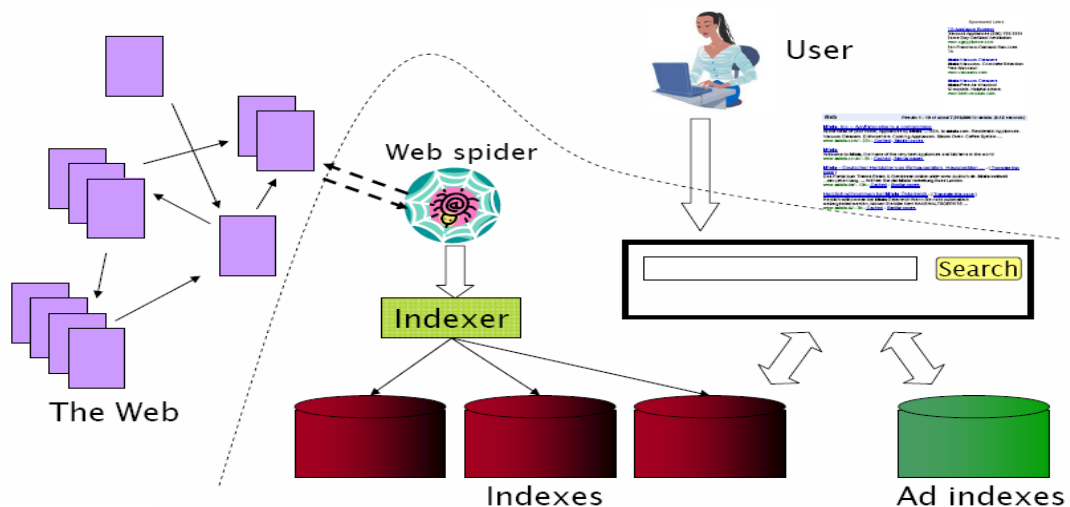


圖 4 Google 搜尋網頁流程圖

參 Google API的技術

在Google 中，使用者設定的查詢條件包含「搜尋關鍵字限定」與「搜尋來源限定」兩部分。其中「搜尋關鍵字限定」主要由使用者所要搜尋的關鍵字以及作用在關鍵字上的運算子（請參表1）所構成。「搜尋來源限定」是用來限制搜尋資料的範圍（請參表2），例如網域或來源檔案型態的限定。

Google 主要的關鍵字運算子

表 1.Google 主要關鍵字運算子

關鍵字運算子	範例說明
分隔關鍵字的空格	Q1 Q2：搜尋同時包含關鍵字Q1和Q2的網頁。
OR	Q1 OR Q2：搜尋包含關鍵字Q1或Q2的網頁。
“ ”	“Q1 Q2”：將引號內容視為一個的關鍵字來查詢。
+ - () *	強制關鍵字一定要列入查詢條件中。由於Google 會自動將出現頻率過高的關鍵字忽略掉，因此若使用者輸入的關鍵字屬於高出現頻率的字，使用此運算子可避免Google 將此關鍵字忽略掉。
+	強制關鍵字不出現在查詢結果之內。
()	群組查詢：可將共同的關鍵字獨立出來，放在括弧的外面。例如A (B OR C) 。
*	代表任意一個英文單字或一個中文的詞。注意：「資料」或「資料庫」都算是一個中文的詞。

搜尋來源限定範例	說明
filetype : pdf	限定搜尋來源為 PDF 檔案。
Site : www.stanford.edu	限定搜尋來源為特定的網域。可將搜尋的範圍限
Site : www.stanford.edu	定在特定的網域之內，例如學術 (edu)、政府或機構(org) ，或是商業 (com)網站。
cache : www.edu.tw	限定搜尋來源為 Google 事先至該網址抓回的網頁。
allintext :	限定關鍵字必須在網頁的內文出現

Agent 代理人在此研究應用當中扮演者搜尋內部資料的角色，讓其可以在不同的主機當中自由的擷取資料庫，及運用 Google API 扮演著搜尋網際網路上的資訊角色，其為 GOOGLE 公司目前搜尋引擎提供大量的 API 函式提供一般使用者使用，結合這兩項技術的特性，再配合線在資料整合技術中的資料探勘，將內外部所搜尋龐大的資料群體，作有效的自動分類、比對、預測等技術，來發掘出有用的的資訊。

肆、Google API的技術與應用

Google 公司提供搜尋索引的API，可讓任何想運用數據來完成IDE的外掛程式，透過網路應用程式定義搜索函式公共原始程式碼。Google 主要提供在代碼搜索中的API和相關的示例碼操作。Google 也提供函式和代碼置入在用戶端的應用程式。

第三節 資料探勘(Data Mining)理論與應用

所謂的資料探勘(Data Mining)就是指從資料庫中挖掘出潛在、明確、而且非常有用資訊的過程。可解釋為資料庫之知識發掘 (Knowledge Discovery in Databases, 簡稱KDD)。也就是說可以從一個大型資料庫裡頭所儲存的大量資料當中萃取出有用知識，這個大型資料庫有可能是線上作業的資料庫 (on-line database)，也有可能是資料倉儲 (data warehouse)。

壹、資料探勘(Data Mining)的理論

資訊系統中所儲存的各種交易資料，會隨著時間的累積，資料量也隨之增加，這些資料往往隱藏著許多特徵或關係等未知而有用的資訊，使用傳統的資料查詢方式，並無法發掘它們的存在，透過資料探勘技術，將可以有效解決此一問題。資料探勘(Data Mining)是從大量資料中擷取或挖掘出知識過程，所探勘出來的知識可以用來支援決策分析等活動。資料探勘被定義為：「針對大量的資料，應用自動或半自動的方式進行分析，以找出有意義的關係或法則」。廣義而言，資料探勘指的是：從儲存於資料庫、資料倉儲或其他資訊貯藏所中發掘出使用者感興趣的知識之過程，或是針對使用者所提出的問題，自龐大的資料庫中萃取出有用資訊、資料樣式與趨勢的過程。

資料庫知識發掘(Knowledge Discovery in Database；KDD) 經常與資料探勘混為一談，根據學者對 KDD 所作定義為：「資料庫知識發掘是從資料中找出合理、新奇及潛在效益的一個非瑣碎(Nontrivial) 的萃取過程，其目的在瞭解資料的型樣(Patterns)」。資料庫知識發掘的整個過程包含七個步驟，分別為資料清理(Data Cleaning)、資料整合(Data Integration)、資料選擇(Data Selection)、資料轉換(Data Transformation)、資料探勘(Data Mining)、樣式評估(Pattern Evaluation)、知識呈現(Knowledge Presentation)，而資料探勘只是其中的一個過程。

貳、資料探勘(Data Mining)的技術

分類按照分析對象的屬性分門別類加以定義，建立類組(class)。

推理 根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。

預測根據對象屬性之過去觀察值來推估該屬性未來之值。

關聯分組從所有物件決定那些相關物件應該放在一起。

同質分組將異質母體中區隔為較具同質性之群組(clusters)。

資料探勘一般可分為七個步驟：(1)Data cleaning (消除雜亂或不一致的資料)、(2)Data integration (將多種資料來源組合在一起)、(3)Data selection (從資料庫中搜尋分析與任務相關的資料)、(4)Data transformation (將資料轉換或統

一成適合探勘的形式)、(5)Data mining (用適當的演算法來對資料做分析)、(6)Pattern evaluation (根據某種interestingness measures, 來辨識並表達真正有趣的模式)、(7)Knowledge presentation (使用視覺化和知識表達技術, 來向使用者提供探勘出的知識)。如超市交易, 透過資料探勘對於銷售記錄的分析, 可以得知消費者在買一樣食物時會跟著買那樣飲料, 使得市場的銷售人員可以對於物品的擺設以及促銷方案, 能夠有更多的資訊以供參考。又例如銀行可以對於過去的金融交易做資料探勘, 可以從中得到日後市場走向的資訊, 進而幫助客戶迴避風險, 做出更有效率的投資。

有一點很重要的是, 沒有一種 Data Mining 的工具可以應付所有的要求。對於某一種問題, 資料本身的特性會影響你所選用的工具。所以你可能需要用到許多不同的工具以及技術從資料中找到最佳的模式。Classification 模式是最常使用的模式, 所以在這裡我們就來介紹建立這種模式的一些常見的方法。

Classification 通常會牽涉到兩種統計方法: Logistic Regression 以及 Discriminant Analysis。然而因為 Data Mining 已漸普遍, 所以 Neural Nets 以及 Decision Tree 也漸漸受到採用。雖然這些統計方法本身都十分複雜, 但使用者並不會牽涉到這些繁雜的統計。

Neural Nets 使用許多參數 (每個參數代表 Net 上的一個 Node) 來建立一個模式, 這個模式接受一組輸入值來預測出一個連續值或分類值。每一個節點 (Node) 都是一個函數, 這個函數是使用輸入該節點的相鄰節點值的加權總和 (Weighted Sum) 做運算。

在建立一個模式的過程中, 我們要用一些資料來 '餵' 給這個網路, '訓練' 它來找到一組能夠產生最佳輸出結果的加權值 (Weights)。有一種最常用的 '訓練法' 稱為 Back-Propagation, 它是把輸出結果與一個已知的正確結果相比。每次相比之後就產生另一組調整過的 Weights, 然後再產生一個新的輸出值再與該已知值相比。這個過程經過反覆的執行後, 這個 Neural Net 就被 '訓練' 得能夠相當正確的做預測了。

可是 Neural Net 有兩個問題。首先, Neural Net 最受質疑的是它的 '曖昧不明' 的特性, 也就是它做的預測所根據的因素並不明確。第二, Neural Net 對測試資料可以做相當正確的預測, 但是對真實資料預測的準確性則較差。但是現在已經有一些新的技術可以改正這個缺點。Decision Tree 則是利用一系

列的規則來得到一個類別或數值。例如，你想把申請貸款的人歸類成 '風險高' 與 '風險低' 兩種，Fig. 2 顯示了一個可以解決這個問題的簡單的 Decision Tree。有了這個 Decision Tree，銀行的放款人員就可以審查申請人的條件，決定該人是屬於高風險或低風險群。例如 '收入高於 40000' 而且 '高負債' 的人會被歸為高風險之類，而 '收入低於 40000' 而且 '工作超過 5 年' 則會被歸為低風險之類。Decision Tree 現在相當普遍，因為它所做的預測相當正確，而且又比 Neural Net 容易瞭解。Decision Tree 與 Neural Net 也可以用來做 Regression，某些種類的 Neural Net 甚至可以用來做 Clustering。IBM 的 Intelligent Miner 可支援 Decision Tree 以及 Neural Net。

參、資料探勘(Data Mining)的應用未來發展

Data Mining 對每個公司來說都是一種重要的策略性的的計畫，而將之列為高度機密，所以要調查各家公司到底用 Data Mining 來做什麼樣的事其實相當不容易。根據 Two Crows Corp. 最近的調查顯示，Data Mining 主要的三個應用方式 - 如我們所預期的 - 都在市場推廣方面，分別是：Customer Profiling、Targeted Marketing、以及 Market-Basket Analysis。

在 Customer Profiling 方面，我們希望找出客戶的一些共同的特徵，希望能藉此預測哪些人可能成為我們的客戶，以幫助行銷人員找到正確的行銷對象。Data Mining 可以從現有客戶資料中找出他們的特徵，再利用這些特徵到潛在客戶資料庫裡去篩選出可能成為我們客戶的名單，作為行銷人員推銷的對象。行銷人員就可以只針對這些名單寄發廣告資料，以降低成本，也提高行銷的成功率。

Market-Basket Analysis 主要是用來幫助零售業者瞭解客戶的消費行為，譬如哪些產品客戶會一起購買，或是客戶在買了某一樣產品之後，在多久之內會買另一樣產品等等。利用 Data Mining，零售業者可以更有效的決定進貨量或庫存量，或是在店裡要如何擺設貨品，同時也可以用來評估店裡的促銷活動的成效。

客戶關係的管理是 Data Mining 的另一個常見的應用方式。我們可以由一些原本是我們的客戶，後來卻轉而成為我們競爭對手的客戶群中，分析他們的特徵，再根據這些特徵到現有客戶資料中找出有可能轉向的客戶，然後公司必須設計一些方法將他們留住，因為畢竟找一個新客戶的成本要比留住一個原有

客戶的成本要高出許多。

近來電話公司、信用卡公司、保險公司、股票交易商、以及政府單位對於詐欺行為的偵測 (Fraud Detection) 都很有興趣，這些行業每年因為詐欺行為而造成的損失都非常可觀。 Data Mining 可以找出可能的詐欺交易，減少損失。財務金融業可以利用 Data Mining 來分析市場動向，並預測個別公司的營運以及股價走向。 Data Mining 的另一個獨特的用法是在醫療業，用來預測手術、用藥、診斷、或是流程控制的效率。

第三章 系統研究方法

在本章中我們將介紹研究流程、研究方法、SWOT 分析。

第一節 研究流程

本系統之研究流程分為系統規劃、系統分析與設計、程式設計、系統測試及系統維護五大部分。首先確定系統之題目及方向後，進行資料蒐集與設計問卷，完成後，進行版面設計、資料庫建置、程式設計三部份，最後再對系統做最後的測試及維護動作。

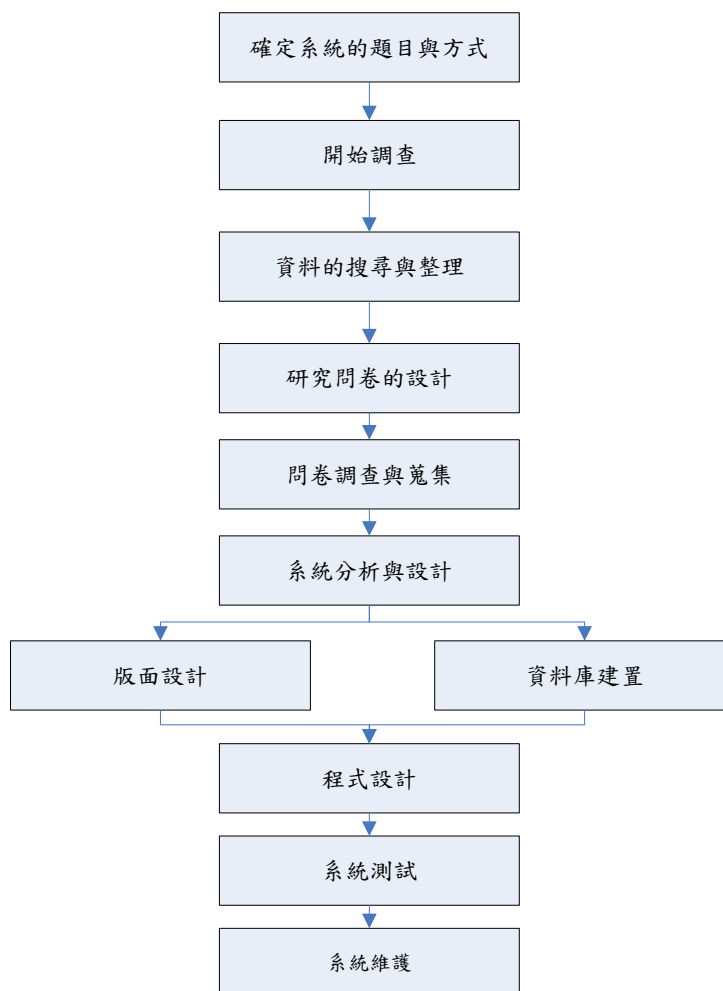


圖 5 研究流程圖

第二節 研究方法

針對本系統在現行社會的應用是否有使用以及存在的價值，我們設計了一份問卷(附件一)，此問卷的設計方向大致上針對現行的搜索引擎滿意度、個人資料的填入方式、可以接受的資訊輸入方式等方面做一個問卷的探討，以下將對於問卷的回收後的效益做一個簡介。

下面數據是呈現在現今社會中，大多數學生每天使用電腦的平均時數，大多數使用時數約在四小時以上。

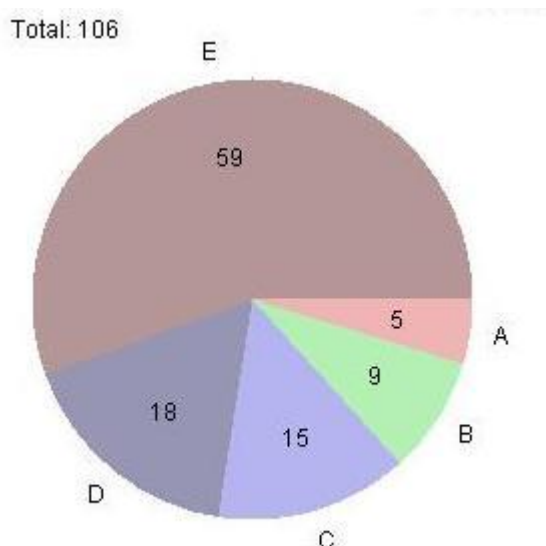


圖 6 大多數學生每天使用電腦的平均時數圖

那麼在這麼長時間使用電腦的狀況下，大家使用電腦的目的是為何呢?根據我們的問卷整理出，約有半數的人在使用電腦的目的是[瀏覽網頁]。

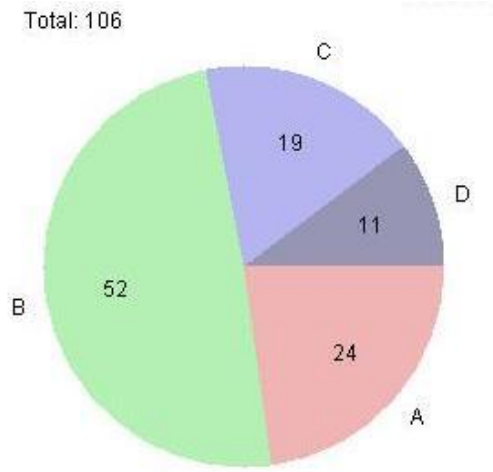


圖 7 使用電腦比例圖

而現今的搜尋引擎眾多，但大多數人依舊習慣使用 YAHOO 奇摩搜尋引擎作為蒐集資料的主要入口網站，在這方面 YAHOO 足足占上 66.9% 的高比例。以下數據即可看出目前使用 YAHOO 及 GOOGLE 的比例圖。

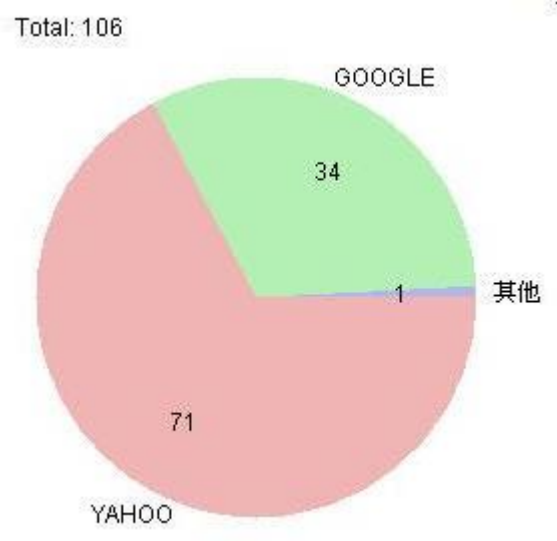


圖 8 YAHOO 及 GOOGLE 的比例圖

現今有這麼多的搜尋引擎，在搜尋資料的完整性是如何呢?怎麼確保我們的資料是否足夠、是否完整呢?下面數據是我們調查目前使用搜尋引擎來搜尋資料的完整性之落點。

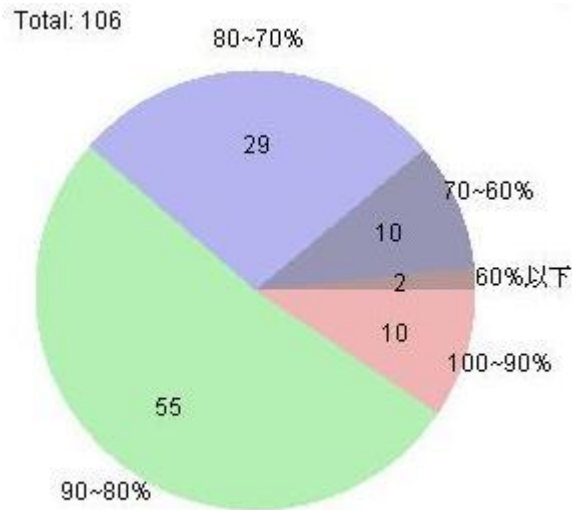


圖 9 搜尋資料的完整度

以下數據是呈現在目前大多數學生對於個人資料輸入部分，較能接受的方式是如何？有 64% 的人認為[半自動輸入]的方式較能接受，又能增加填寫資料的速度，又不會被精靈給設定住條件，這樣半自動輸入的方式是最理想化的。

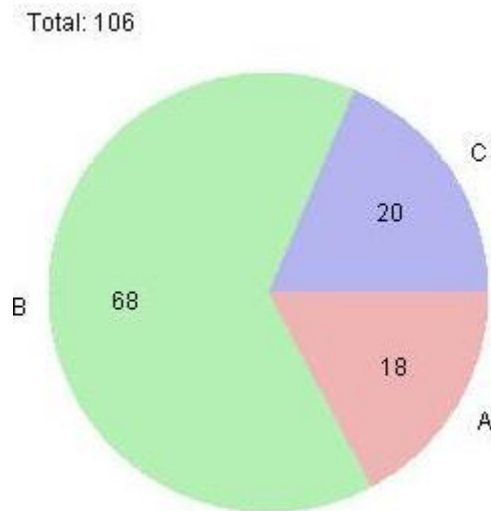


圖 10 資料輸入方式比例圖

那麼現在如果有一套自動輸入的系統，您是否願意採用呢？在市場上，這樣自動輸入的系統接受度高嗎？據我們的調查，有 86% 的使用者是願意使用的，代表這樣的系統在市場上的接受度很高，是值得開發的。



圖 11 自動輸入的接受度

如現今開發一套全新系統，資料獲取方式是全自動搜索及輸入，只需要最後人工進行些微調整，這樣的系統在市場上的接受度經調查後得知，願意使用的比例是相當高的。

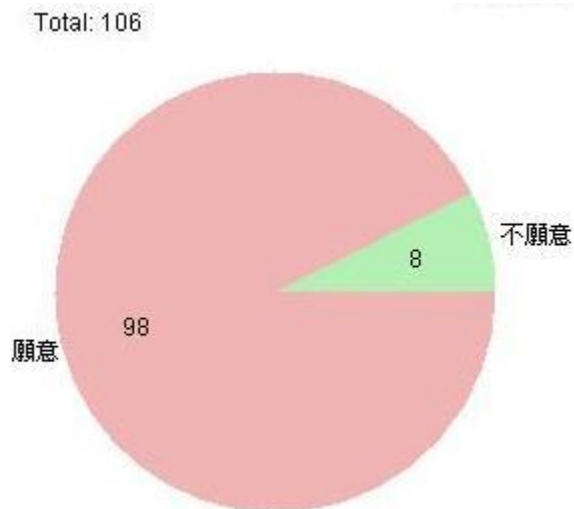


圖 12 自動輸入與微調的接受度

此份問卷目前經由數百人撰寫，撰寫此份問卷的年齡層大多為學生族群，以下是撰寫此問卷的年齡層比例圖。

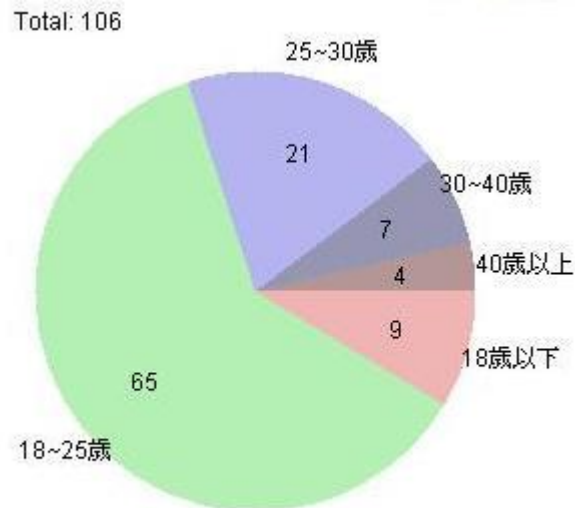


圖 13 填寫問卷年齡層

第三節 SWOT 方法

本系統的最終目的在於資料整合的應用，將資訊自動整合與搜尋，其面對現存資訊的搜尋方式較為新穎，針對本系統的優勢、劣勢、機會以及威脅我們列出以下之 SOWT 分析。分析著我們的立足點以及可能面臨的威脅及劣勢。

表 2 SWOT

Strength：優勢	Weakness：劣勢
<ul style="list-style-type: none"> ● 取代傳統文字方塊的輸入方式 ● 增加使用者使用的意願 ● 使用者不必了解複雜的技術，便可得知準確資訊 ● 使用負載平衡架構，不容易讓伺服器增加負擔 	<ul style="list-style-type: none"> ● 資料庫大量蒐尋時的時間，所花費的時間過多 ● 搜尋的資料量龐大，使用者確認資料所需花費的時間增加
Opportunity：機會	Threat：威脅
<ul style="list-style-type: none"> ● 使用 Agent 搭配 Data Mining，讓固定族群使用 ● 使用技術創新，不容易被其他技術取代 ● 資料精準，且資料眾多 ● 顛覆以往的搜尋引擎 ● 顛覆以往使用者的習慣 	<ul style="list-style-type: none"> ● 若 Google API 改版以後，程式有可能需要更動(有交\$無累，使用者付費) ● Deep Web 概念引發搜尋技術改革所帶來的衝擊

第四章 預期研究成果

在本章中我們將介紹系統功能、系統設計、系統特色、使用對象、使用環境、開發工具、網站架構圖。

第一節 系統功能

- 藉由 Agent 網路代理人方式進行內部資料收集建立索引，以達成收尋無時差。
- 利用 Data Mining 提供之演算法進行資料分群分組決策樹，以達建立標籤收尋 Index 目的。
- 透過 Web2.0 技術讓使用者做最後認證採 SSL3.0 〈Security Socket Layer〉。

第二節 系統特色

- 擬人化(Personality)：代理程式藉由良好的設計，必須表現與人類的行為相同，以利與人類溝通、互動。
- 持續性(temporal continuity)：代理程式必須是依個不斷執行的程序(a continuously running process)，不能僅對某依輸入(Input)產生回應之輸入(Output)後，就結束執行。
- 自主性(Autonomy)：代理程式可以依照某種情況、條件、獨立自主的執行某些工作。在必要的時候，得以持續不斷的執行某項特定任務。
- 溝通能力(communication ability)：代理程式必須具有跟其他代理程式溝通的能力，以及與使用者溝通的能力，藉以獲取資訊並且幫助任務順利達成。
- 調適力(Adaptability)：透過紀錄過去與使用者互動的經驗，代理程式要能夠自我改變，以適應使用者的偏好。此外，代理程式也必須能隨著環境的不同改變自己。

- 機動性(Mobility)：代理程式必須能夠在不同類型的系統、機器、平台中穿梭，以利代理程式擷取資訊。

第三節 使用對象

本校校友及學生，且經由授權後的特定族群作為規範。

第四節 使用環境

Microsoft Windows Server 2003

第五節 開發工具

- 本組以 Java 語言做為基礎程式設計架構，配合其圖型使用介面 (Graphic User Interface ; GUI) 一種圖形呈現的使用者介面，程式使用者可以在使用介面上轉換成滑鼠操作，圖示、視窗、按鈕和對話方塊等圖形元件的圖形使用介面。Java 程式語言支援標準 GUI 元件的 AWT 和 Swing 元件，能夠在不同的作業系統，建立相同的外觀和操作方式的視窗應用程式。
- 使用 Agent 也就是所謂的代理人來做內部資料獲得，最後進行進階的篩選和分析。
- XML 是一種「中介標籤語言」(meta-markup language)，可提供描述結構化資料的格式，這將有助於文件內容的宣告，並符合跨平台的搜尋作業。
- 運用 Microsoft SQL Server 2005 資料庫系統做為後端的資訊整合
- Microsoft Visual Studio 2005 用於改善開發程序，並使得工作產生重大突破性且更容易符合需求。
- 使用 Internet Information Services(IIS) 網頁應用程式伺服器來做管理，它是微軟的 IIS 與 Windows NT Server 作業系統來緊密的整合在一起。
- IIS 6.0 與 Windows Server™ 2003 為網頁應用程式伺服器的管理、可用性、可靠性、安全性、效能及延展性，帶來許多新功能。同時，IIS 6.0 不但增強網頁應用程式的開發過程，還可

改善多語系的支援。IIS 6.0 與 Windows Server™ 2003 同時建置時，能夠提供最可靠、最有效率的完善連線，及整合式網頁伺服器解決方案。

第六節 系統平台架構

系統平台分為三大部分，第一部份為內部的資料庫行動代理人，第二部份為資料探勘的發掘。

本系統平台的各項系統模組之功能分別說明如下：

- (1)建立動態整理：有效搜尋及整合內部資料庫比對後的資料。
- (2)建立內部搜尋的 Data Mining：利用內部資料條件篩選後的資料。
- (3)建立內部驗證：比對內部資料正確性。
- (4)建置便利性輸入：利用內建的選擇，為使用者將資料填入資料庫。
- (5)運用內部資料庫進行資料探勘，發掘有用參數。

第五章 結論

在本章中我們將介紹預期研究效益、預期研究限制。

第一節 預期研究效益

壹. 半自動輸入的方式

大幅減少使用者輸入資料的時間，只需進行些微的調整來達到所需的目的。

貳. 利用 Agent 及 Data Mining，讓搜尋更準確

透過資料擷取，再透過正確的分類來達到高度的資訊正確率。

參. 高度移植性

可以讓這套系統可以移植到任何需要資料填入的系統當中。

第二節 預期研究限制

壹. 發展設備不完備

目前我們的資源僅限於學校提供的資源，導致無法提供足夠的設備讓我們做更進一步的開發。

貳. 發展時間不夠多

由於此系統開發的時間非常有限，無法取得更多的相關資訊，來加以探討及更改。

第六章 分工執掌和進度表

在本章中我們將介紹分工執掌、進度規劃

第一節 分工執掌

表 3 分工執掌

系統規劃	工作人員
訂定主題內容	全員
需求分析	全員
資料蒐集	
Google API	游濬瑋、吳昆逸
AJAX	游濬瑋、吳昆逸
Agent	陳俊維、楊明勳
Data Mining	莊舒晴、曾俊傑
雛形建置	全員
系統分析	
系統分析	全員
規劃架構	全員
製作企劃書	全員
系統設計	
系統設計	
Google API	游濬瑋、吳昆逸
AJAX	游濬瑋、吳昆逸
Agent	陳俊維、楊明勳
Data Mining	莊舒晴、曾俊傑
系統實作	
整合資料結構、相片管理	游濬瑋
資料庫的建置、個人名片	吳昆逸
註冊頁面	莊舒晴
留言版	陳俊維
美術編輯版面藝術、基本資料	曾俊傑
活動公告事項	楊明勳

第二節 進度表

識別碼	工作名稱	開始	完成	期間	2009											
					四月	五月	六月	七月	八月	九月	十月	十一月	十二月			
1	訂定目標	2009/3/19	2009/4/2	2.2w												
2	需求分析	2009/4/3	2009/4/16	2w												
3	資料蒐集	2009/4/3	2009/4/30	4w												
4	雛形建置	2009/5/1	2009/5/8	1.2w												
5	系統分析	2009/5/11	2009/7/1	7.6w												
6	規劃架構	2009/5/15	2009/5/29	2.2w												
7	製作企劃書	2009/4/17	2009/6/30	10.6w												
8	系統設計	2009/6/1	2009/6/15	2.2w												
9	系統開發	2009/7/1	2009/11/30	21.8w												

參考文獻

- [1] B. O. Szuprowicz, "Search engine technologies for the World Wide Web and Internet," 1997, USA, Computer Technology Research Corp.
- [2] Bo Shu and Subhash C. Kak, "A Neural Network-based Intelligent Metasearch Engine," Information Sciences, 1999, vol. 120, pp.1-11.
- [4] J. C. Reynar and A. Ratnaparkhi. "A Maximum Entropy Approach to Identifying Sentence Boundaries," In Proceedings of the 5th Conference on Applied Natural Language Processing, 1997, pp. 16-19.
- [5] Mei Kobayashi and Koichi Takeda, "Information retrieval on the web," ACM Computing Surveys, 2000, vol. 32, no. 2, pp. 144-173.
- [6] Robert E. Filman and Sangam Pant, "Search the internet," IEEE Internet Computing, 1998, vol. 2, pp. 21-23.
- [7] Sharon Tsai, "A Writing Assistant Using Language Models Derived from the Web", Language processing and computational linguistics, 2003
- [8] Vankat N. Gudivada, V. Raghavan, William I. Grosky, and Rajesh Kananagottu, "Information retrieval on the world wide web," IEEE Internet Computing, 1997, vol.1, pp. 58-68.
- [9] Google Inc , "Google Web APIs Reference", 2002, Mueller John , "Mining Google web services :building applications with the Google API", San Francisco Sybex, 2004
- [10] 陳志榮, Metasearch engine 中資料庫分類之關鍵詞維護機制, 元智大學資訊管理學研究所碩士論文, 2002
- [11] 林東清(2008)資訊管理 e化企業的核心競爭能力三版
- [12] altavista <http://world.altavista.com/>
- [13] Openfind <http://www.openfind.com.tw/>
- [14] Google <http://www.google.com.tw>
- [15] yahoo <http://tw.yahoo.com>

- [16] PChome <http://www.pchome.com.tw>
- [17] Excite <http://www.excite.com/>
- [18] 探宇科技 <http://www.uniminer.com/>
- [19] 張云濤、鞏玲，資料探技術與原理 2007版
- [20] **UniMiner** 探宇科技股份有限公司 <http://www.uniminer.com/center01.htm>
- [21] Cabena, P., Hadjinian P. and Stadler, R. (1997), Discovering Data Mining From Concept to Implementation, Prentice-Hall Inc.
- [22] Berry, M. J. A. and Linoff, G. (1997), Data Mining Technique for Marketing, Sale, and Customer Support, Wiley Computer.
- [23] Han, Jiawei, and Kamber, M. (2001), Data Mining : Concepts and Techniques, Mogan Kaufmann Publishers.
- [24] Han, J., Pei, J., and Yin, Y. (2000), “Mining Frequent Patterns without Candidate Generation” , In Proceedings of the ACM-SIGMOD International Conference on Management of Data (SIGMOD’ 00), pp. 1-12.
- [25] Pei, J., Han, J., Lu, H., Nishio, S., Tang, S. and Yang, D. (2001), “H-Mine: Hyper-Structure Mining of Frequent Patterns in Large Databases” , ICDM 2001 , pp. 441-448.
- [26] Witten, I.H. and Frank, E. (2005), Data Mining Practical Machine Learning Tools Elsevier.
- [27] Intel Pentium 4 Processor ,
http://www.intel.com/pentium4/index.htm?iid=intelhome+roll_Pentium4& .
- [28] BBC News | HEALTH | Screensaver aids cancer fight ,
http://news.bbc.co.uk/hi/english/health/newsid_1255000/1255804.stm .
- [29] SETI@home: Search for Extraterrestrial Intelligence at Home ,
<http://setiathome.ssl.berkeley.edu/> .
- [30] Intel philanthropic peer-to-peer program , <http://www.intel.com/cure/> .
- [31] 楊朝棟、張宏守，「在 Linux 上建構與應用叢集式平行電腦」，零客情報，第 12 期，84~103 頁，November 2000 .

- [32] Parallel Processing Concepts ,
<http://www.tc.cornell.edu/services/edu/topics/ParProgCons/more.asp> 。
- [33] 李新夏, 「Turbolinux Cluster Server 6.0」, 凌客誌, 第 3 期, 45 頁, January 2001 。
- [34] 張西亞、蔡佳璋, 「NCHC PC Cluster 簡介」,
http://hpcserv2.nchc.gov.tw/project/lecture_notes/NCHC_PC_Cluster_簡介.pdf。
- [35] The Beowulf Project , <http://www.beowulf.org> 。
- [36] Bernd O. Christiansen, et al., “Javelin: Internet-Based Parallel Computing Using Java”.
- [37] Ying-Lung Wang, “Design and Implementation of a Parallel Virtual Machine with Distributed Shared Memory by Using Java”, Master thesis, Department of Computer Science, National Tsing Hua University, June 2000.
- [38] MPI — The Message Passing Interface Standard ,
<http://www-unix.mcs.anl.gov/mpi/> 。
- [39] Message Passing Interface Forum, ”MPI: a Message-Passing Interface Standard”, Tech. Rep. CS-94-230, Department of Computer Science, University of Tennessee, Knoxville, TN, 1994.
- [40] PVM: Parallel Virtual Machine , <http://www.csm.ornl.gov/pvm> 。
- [41] A. Geist, et al., PVM: Parallel Virtual Machine — A Users’ Guide and Tutorial for Networked Parallel Computing, MIT Press, 1994.
- [42] [Fitzmaurice, G., “Form-counteracted Workflow Automation Using an Agent Framework”, Brown university, 1991.
- [43] P. Maes, “Agents that Reduce Work and Information Overload”, Communications of the ACM, Vol.37, No.7, pp. 31-40, ACM Press, July 1994.
- [44] 邱顯正, 「網際網路上產品資訊擷取代理人雛形之設計與建置」, 元智大學, 資訊研究所碩士論文, 民國 89 年。
- [45] 數位之牆 , <http://www.digitalwall.com>
- [46] Kalakota, R., and Whinston, A. B., Frontiers of Electronic Commerce, Reading,

- Mass., Addison-Wesley, 1996.
- [47] Bloch, M., Pigneur, Y., and Segev, A., "On the Road of Electronic Commerce — a Business Value Framework, Gaining Competitive Advantage and Some Research Issues," an extended version of a publication in the Proceedings of the Ninth International EDI-IOS Conference, Bled, Slovenia, Jun 1996.
- [48] Kalakota, R., and Whinston, A. B., *Electronic Commerce : A Manager's Guide*, Reading, Mass., Addison-Wesley, 1997.
- [49] Applegate, L. M., Holsapple, C. W., Kalakota, R., Radermacher, F. J., and Whinton, A. B., "Electronic Commerce: Building Blocks of New Business Opportunity," *Journal of Organization Computer Electronic Commerce*, 1996.
- [50] Engel, J. F., Blackwell, R. D., and Miniard, P. W., *Consumer Behavior*, 6th Edition, Dryden Press, Orlando, 1990.
- [51] 鄭又豪，「電子式產業交易平台的形成與發展」，中央大學資訊管理研
- [52] 究所碩士論文，民國87年6月。
- [53] Kambil, A., "Doing Business in the Wired World," *IEEE Computer*, Vol.3, No.5, May 1997, pp.56-61.
- [54] 林夏萍，「台灣地區網際網路消費者購物資訊搜尋行為之初探」，中正
- [55] 大學電訊傳播研究所碩士論文，民國87年6月。
- [56] P. Maes. "Agents that Reduce Work and Information Overload." *Communications of the ACM*, Vol.37, No.7, pp. 31- 40, 146 ACM Press, July 1994.
- [57] Kalakota, R., and Whinston, A. B., *Frontiers of Electronic Commerce*, Reading, Mass., Addison-Wesley, 1996.
- [58] Kambil, A., "Doing Business in the Wired World," *IEEE Computer*, Vol.3, No.5, pp.56-61, May 1997.
- [59] Cowie, J. and Lehnert W., "Information Extraction", *Communication of ACM*, Vol 39, No1, pp.80-91, January 1996.
- [60] Kushmerik, N, "Gleaning the Web", *IEEE Intelligent Systems*, Volume: 14 2 ,

Page(s): 20 -22, 1999.

- [61] San, M., "Intelligent agents on the Internet and Web", TENCON '98. IEEE Region 10 International Conference on Global Connectivity in Engery, Computer, Communication and Control, Vol.1, pp.97-102.
- [62] Faloutsos, C. and Oard D., "A survey of Information Retrieval and Filtering Methods", University of Maryland College Park, CS-TR-3514, 1995. Also available at <http://www.enee.umd.edu/medlab/filter/papers/survey.ps>
- [63] Nie J., Briscois M., and X. Ren, "Template-Based Information Mining from HTML." In: Proceedings of AAAI-97, Providence, USA, 1996.
- [64] 莊水龍，「助理型軟體資訊擷取技術：樹狀結構 HTML 文件樣板之自動產生法」，國立臺灣大學資訊工程學研究所，1998。
- [65] Chun-Nan Hsu, "Initial results on wrapping semistructured web pages with finite-state transducers and contextual rules", In Proceedings of AAAI-98 Workshop on AI and Information Integration, Technical Report WS-98-01, Menlo Park, CA, 1998.
- [66] Yanlei Diao, Hongjun Lu, Songting Chen, Zengping Tain, "Toward Learning Based Web Query Processing", VLDB 2000, 317-328, 2000.
- [67] Wolfgang May, "An integrated Architecture for Exploring, Wrapping, Mediating and Restructuring Information from the Web", Database Conference, 2000. ADC 2000. Proceedings. 11th Australasian, Page(s): 82 -89, 2000.
- [68] Ling Liu, Calton Pu, Wei Han, "XWRAP an XML-enabled Wrapper Construction System for Web Information Sources", Data Engineering, 2000. Proceedings. 16th International Conference, Page(s): 611 -621, 2000.
- [69] 陳光華，電子文件自動處理之研究，學生書局，台北，1999。
- [70] 曾元顯，"關鍵詞自動擷取技術之探討"，中國圖書館學會會訊，5 卷 3 期 (106)，1997。
- [71] Salton and McGill, Introduction to Modern Information Retrieval, McGraw Hill Book Co, 1989.

- [72] Lee H. C, Dagli C. H., Ercal F., and Ozbayoglu A. M., "SimNet: A Parallel Neuro-Fuzzy Paradigm for Data Clustering", OAI Neural Networks Symposium and Workshop(OAINN '95), Athens, Ohio, USA, 1995.
- [73] 許加文、李錫捷，「網際網路搜尋過濾系統[一個「關鍵頁」超搜尋行動型代理引擎]」，第九屆國際資訊管理學術研討會，中壢，台灣，1998。
- [74] 謝超煒，「網際網路資訊擷取過濾系統-中文關鍵頁超搜尋代理人」，私立元智大學資訊管理研究所，碩士論文，2000。
- [75] S. Aggarwal, F. Hung, Weiyi Meng, "WIRE-a WWW-based information retrieval and extraction system", IEEE Database and Expert Systems Applications, Page(s): 887 -892, 1998.
- [76] 黃宏軒，「全球資訊網資訊整合可程式化包覆程式的設計與實作」，國立臺灣大學資訊工程學研究所，碩士論文，1999。
- [77] 陳昱彰，「個人化網際網路電子購物代理人」，國立交通大學資訊工程學系，碩士論文，1997。
- [78] 簡逸音，「個人化技術在網際網路購物代理人之研究」，中華大學電機工程學系碩士班，碩士論文，1999。
- [79] 顏逸品，「網際網路半結構化資料之蒐集與整合研究」，國立中央大學資訊管理研究所，碩士論文，1999。
- [80] 黃國倫，「網際網路購物代理人之設計與建構」，元智大學資訊管理研究所，碩士論文，2001。
- [81] 陳伯文，「代理人架構下分散式平行運算平台之設計與建構」，元智大學 資訊管理研究所，碩士論文，2001。

附錄一 Show your life 「使用電腦生活習慣」問卷

致理技術學院資訊管理系畢業專題研究問卷	
1. 請問目前平均使用電腦時數？	<input type="radio"/> 1 小時以下 <input type="radio"/> 1~2 小時 <input type="radio"/> 2~3 小時 <input type="radio"/> 3 小時以上
2. 請問使用電腦的時間大多在做何種事務？	<input type="radio"/> 蒐集資料 <input type="radio"/> 玩遊戲 <input type="radio"/> 瀏覽網頁 <input type="radio"/> 購物
3. 請問目前最常使用之搜尋引擎網站為何？	<input type="radio"/> YAHOO <input type="radio"/> GOOGLE <input type="radio"/> 其他_____
4. 請就上述搜尋引擎網站的完整性大約在哪個落點？	<input type="radio"/> 90~100% <input type="radio"/> 80~90% <input type="radio"/> 70~80% <input type="radio"/> 60~70% <input type="radio"/> 60% 以下
5. 請問目前個人資料輸入方式何者較可以接受？	<input type="radio"/> 人工輸入 <input type="radio"/> 系統協助輸入 <input type="radio"/> 系統自動輸入 如果選擇系統自動輸入請至第 4 題。
6. 承上題，如系統預設為自動輸入，是否願意採用？	<input type="radio"/> 願意 <input type="radio"/> 不願意 如果願意請至第 5 題，不願意請說明理由。 理由：_____。
7. 承上題，如採用，是否願意將系統的錯誤採人工方式做些微調整？	<input type="radio"/> 願意 <input type="radio"/> 不願意 如果不願意請說明理由。 理由：_____。
問卷填寫人：_____。	
年齡： <input type="radio"/> 18 歲以下 <input type="radio"/> 18~30 歲 <input type="radio"/> 30~40 歲 <input type="radio"/> 40~50 歲 <input type="radio"/> 50 歲以上	
性別： <input type="radio"/> 男性 <input type="radio"/> 女性	
非常感謝你的配合，你的一份問卷讓我們的研究目標可以更為確定。	
敬祝	
身體健康！	
致理技術學院資訊管理系上	