

國立交通大學

多媒體工程研究所

碩士論文

個人知識重整 - 以記憶模型與標籤技術為基礎

Reorganizing Personal Knowledge  
Based on Memory Models and Tagging Technology

研究生：林依文

指導教授：孫春在 教授

中華民國九十六年六月

個人知識重整 - 以記憶模型與標籤技術為基礎

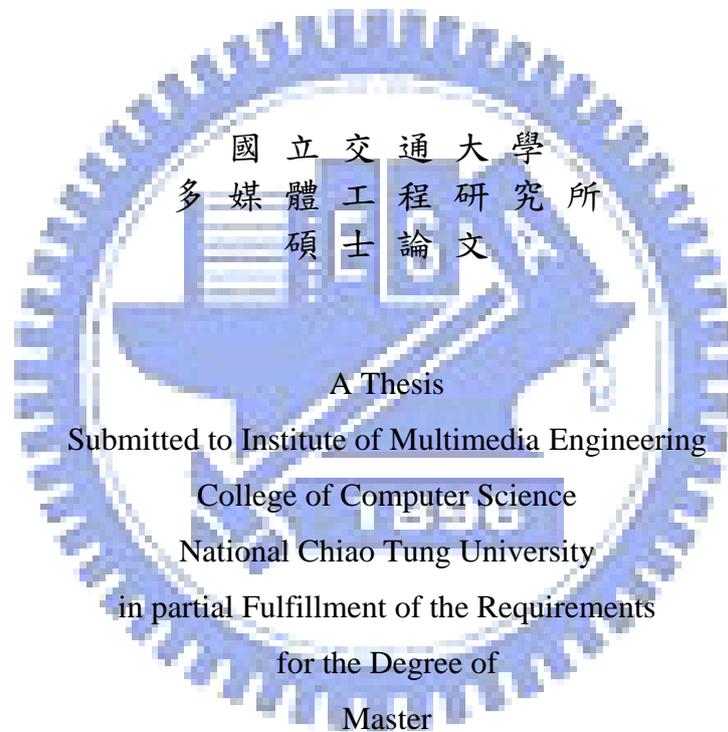
Reorganizing Personal Knowledge  
Based on Memory Models and Tagging Technology

研究生：林依文

Student：I-Wen Lin

指導教授：孫春在

Advisor：Chuen-Tsai Sun



in

Computer Science

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

# 個人知識重整 - 以記憶模型與標籤技術為基礎

學生：林依文

指導教授：孫春在 教授

國立交通大學多媒體工程所

## 中文摘要

資料夾為自從電腦開始使用以來，就一直擔任著整理資訊的工具，但是資料夾本身有其限制及缺點，例如：樹狀結構所產生的問題、階層優先度問題等，導致有時候無法有效地尋找到所需的資訊，造成人們的困擾。現有的檔案搜尋系統，如 Google Desktop Search，限於關鍵字與資訊內容可能不太相關的問題，也無法完全解決這個問題。由於標籤網路為平面式網路結構無階層性問題，以及是由使用者對資訊的認知所下的標籤，所以可以避免檔案搜尋系統關鍵字相關度的問題，所以我們將採取此結構來做為解決方案。

本論文以在網際網路上發展迅速的標籤(Tag)分類法，針對個人組織整理其所擁有資訊的方面，來設計出一個個人知識重整系統，以認知心理學的記憶模型和記憶歷程的理論為基礎，期望能改善資料夾和檔案搜尋系統現有的問題，以及進而能幫助人類做記憶中知識的重整。針對高中三年級的學生，藉由採用個人知識重整系統和傳統資料夾分類二種方式來做實驗，第一階段請二班各別使用二種方式來整理圖檔，隔週之後，第二階段請學生使用當初建立結構的系統來找尋指定的目標圖片，於建立與搜尋的過程中以電腦程式記錄下使用者的動作。比較二者搜尋到目標檔案的效率、各自結構建立的風格、二種結構的搜尋風格差異，以及個人與整體間是否存在某些認知的共識。

研究結果發現使用個人知識重整系統比資料夾在搜尋目標檔案時，時間上較為快速，步驟數較多。使用者大部分為圖片加上 2~3 個標籤。個人所建立的標籤網路，在某些部分會和團體共同認可的結構相同。

關鍵字： 標籤、認知心理學

# **Reorganizing Personal Knowledge Based on Memory Models and Tagging Technology**

Student: I-Wen Lin

Advisors: Dr. Chuen-Tsai Sun

Institute of Multimedia Engineering  
National Chiao Tung University

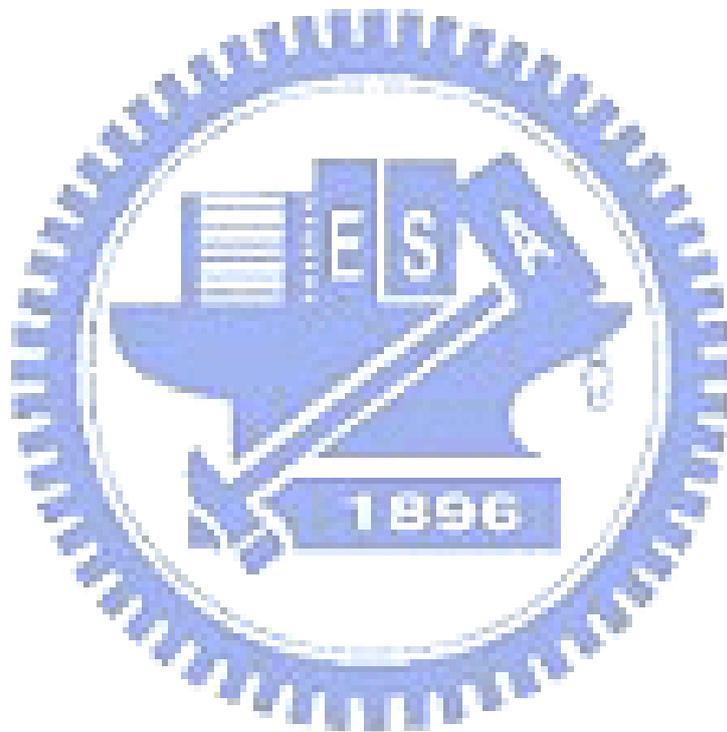
## **Abstract**

Since the first version of Window operation system was developed in 19XX, a categorization approach “folder” have served as an important information tool to organize data. However, the hierarchical structure of folder has its limitations and weaknesses for organization data, e.g. problem of tree structure and hierarchical priority...etc. The drawback of folder categorization may make people difficult to effectively search for demanded information. Besides, the existing file searching system like Google Desktop Search limited by keyword may not reflect the context of content, so it needs a more effective and efficient approaches to help people search in organized knowledge structure even unorganized data pool. In this study, I focus on re-organizing knowledge by tagging approach to help data search and organization. Because tagging network is structured as plane network, it has no hierarchical problem. And due to tagging procedure is established based on user’s cognition, users can avoid content relation problem resulted from keyword search.

This thesis uses tagging technology to design a personal knowledge re-organizing system which is based on memory models and memory recording processes of cognitive psychology. The proposed tagging system aims to improve original files searching system based on folder structure, and thus can help people record, store, and re-organize knowledge in user’s personal biological memory. Several experiments were run in senior high school for evaluating the proposed tagging system by comparing the traditional hierarchical folder system. Two phases experiments including pretest and posttest show that using tagging system to re-organize personal knowledge is more efficient, effective, and intuitive than traditional approach. Moreover, by visualizing how user records and

organizes his/her knowledge, the tagging network structure organized by different user can show user's personal memory style which can be viewed as a useful tools to observe how people organize their personal knowledge.

Keyword: Tag, Cognitive Psychology



## 誌謝

本論文能夠順利地完成，首先要感謝我的指導教授孫春在教授，獨特的思考方式帶領我們用不同的角度去切入問題，以其豐富的經驗與知識和對於新領域的接納，讓學生可以有各式各樣的研究，也才能有本論文的出世。其次，感謝口試委員張智星教授、胡毓志教授以及陳穎平教授針對本論文之指正與建議，使本論文得以更加完整。

我也萬分感謝研究室的學長姊們：吉隆學長指導我研究的方向，在一次又一次的討論中，確定了論文的骨架。佩嵐學姊給予我關於教學實驗的意見，讓對於教學實驗較陌生的我能規劃出不失的實驗方法。建勳學長全力輔助幫忙系統的實作，能成功實作出系統都要感謝學長對於理論和實作關係的指點。右敏學姊和朝淵學長幫忙協調實驗班級狀況，讓實驗能夠順利完成。

另外還要感謝實驗室的同學們：宗元在論文理論上提供的觀點，正宏在口試前協助投影片的修正，岳暄、文翊…等一起努力的伙伴們，還有學弟妹們的關心，謝謝你們！

最後，要感謝的就是我的家人，感謝爸爸、媽媽民主開放的作風，讓我選擇有興趣的領域學習，一直在背後默默地、溫暖地支持著我，煩惱時陪我一同擔憂，沒有你們我也許走不到這裡，謝謝！

林依文 謹於

交通大學

中華民國九十六年六月

# 目錄

中文摘要 .....	i
Abstract.....	ii
誌謝 .....	iv
目錄 .....	v
圖目錄 .....	vii
表目錄 .....	ix
一、 緒論 .....	1
1.1 研究動機 .....	1
1.2 研究問題 .....	3
1.3 研究重要性 .....	4
二、 文獻探討 .....	5
2.1 背景探討 .....	5
2.2 標籤(Tag)技術 .....	6
2.3 記憶 .....	10
2.3.1 記憶模型 .....	10
2.3.2 記憶歷程 .....	12
三、 系統架構與設計 .....	15
3.1 核心概念 .....	15
3.2 系統概念架構 .....	16
3.3 實作內容 .....	16
3.3.1 使用者介面 .....	17
3.3.2 資料庫設計 .....	19
3.3.3 資訊內容的選擇 .....	19
3.3.4 系統功能計算 .....	20
四、 實驗方法與結果分析 .....	21
4.1 研究對象 .....	21
4.2 研究工具 .....	21
4.2.1 記錄結構建立及搜尋的過程 .....	21
4.2.2 背景調查問卷 .....	21

4.3	實驗流程 .....	21
4.3.1	搜尋任務說明 .....	23
4.4	實驗結果與分析 .....	23
4.4.1	系統使用狀況 .....	23
4.4.2	搜尋效率 .....	24
4.4.3	結構建立風格 .....	26
4.4.4	搜尋風格 .....	33
4.4.5	個人與群體差異 .....	34
五、	結論 .....	38
	參考文獻 .....	39
附錄 A.	實驗中搜尋任務的詳細說明 .....	41
附錄 B.	系統介面的操作方法 .....	43



## 圖目錄

圖 2.1.1	個人網站架構範例.....	6
圖 2.2.1	二分網路(bipartite network).....	7
圖 2.2.2	資訊內容(Content) 和 標籤(Tag) 的二分網路 .....	8
圖 2.2.3	書目管理網站 CiteULike .....	9
圖 2.3.1	記憶的三段儲存模型.....	10
圖 2.3.2	工作記憶的示意圖.....	11
圖 3.2.1	三段儲存模型與記憶歷程的基本操作.....	16
圖 3.2.2	概念架構.....	16
圖 3.3.1	實作架構圖.....	17
圖 3.3.2	系統介面.....	18
圖 3.3.3	資料庫架構圖.....	19
圖 4.4.1	每張圖片平均擁有標籤數的分佈圖，橫軸代表每張圖片擁有的標籤數，縱軸代表擁有特定標籤數的圖片平均有幾張，結果的高峰點顯示一張圖片有 3 個標籤的情形最多.....	24
圖 4.4.2	搜尋目標圖片所需的時間圖.....	25
圖 4.4.3	搜尋目標圖片所需的步數圖.....	26
圖 4.4.4	資料夾結構.....	27
圖 4.4.5	連結度高的標籤網路，此為某一受測者所建立的標籤網路，大部分的節點皆連結在一起，代表對於不同圖片間有相同的標籤，一張圖片含有多個標籤.....	28
圖 4.4.6	單點散佈的標籤網路，此為某一受測者所建立的標籤網路，大部分的節點之間無相連，代表此受測者大部分圖片只下一個標籤.....	29
圖 4.4.7	區塊分隔的標籤網路，此為某一受測者所建立的標籤網路，節點相連成一個個的區塊，區塊與區塊間無相連，代表此受測者不同圖片間很少下相同的標籤	
	30	
圖 4.4.8	使用比率與使用範圍的分佈圖.....	31
圖 4.4.9	特定圖片.....	32
圖 4.4.10	特定圖片的資料夾名稱使用統計.....	32
圖 4.4.11	特定圖片的標籤名稱使用統計.....	33
圖 4.4.12	資料夾搜尋範例.....	33

圖 4.4.13	搜尋標籤範例－步驟 1 .....	34
圖 4.4.14	搜尋標籤範例－步驟 2 .....	34
圖 4.4.15	群體共識標籤網路.....	35
圖 4.4.16	共通標籤在個人標籤網路上的位置(範例 1).....	36
圖 4.4.17	共通標籤在個人標籤網路上的位置(範例 2).....	37



## 表目錄

表 2.3.1	三個不同的處理層次.....	11
表 4.4.1	搜尋時間效率比較表.....	25
表 4.4.2	搜尋步數比較表.....	26



# 一、緒論

## 1.1 研究動機

自從二〇〇五年 Time O'Reilly 提出「What Is Web2.0」之後，Web2.0 的觀念，提供了使用者一個重新審視與檢索線上資料或連結的新方法，從原本單純由服務提供者主動提供知識且使用者被動地接受與搜尋知識，轉變服務提供者提供平台，讓使用者主動地在網路上發表、分享、蒐集、與整理知識。Web2.0 的服務種類繁多，因此，一大批應用在線上各種多媒體與知識的專業網站應運而生，包含部落格 (blog)、社會性書籤 (Social Bookmarking)、線上圖片共享網站 (Photo sharing website) 等。以網路書籤 (或稱為社會性書籤) 而言，它是一種超連結的收藏和分享的服務，使用者可以通過它來收集、分類感興趣的網路資訊，例如新聞、圖片、影片、網頁等。同時，也能方便地與其他人分享自己的個人收藏，並從其他使用者收藏中進行資訊的採集。代表性的網站為 del.icio.us<sup>1</sup>、reddit<sup>2</sup>，中文則有 HEMiDEMi 黑米共享書籤<sup>3</sup>...等。網路相簿或稱為照片共享網站提供使用者發佈並與其他使用者分享數位相片的平台，此服務也可以提供作為網誌圖片的存放空間，所以受到許多部落格作者的喜愛。代表的網站如 Flickr<sup>4</sup>與 Google Picasa Web Album<sup>5</sup>等。上述的網站有一個共同的特色，就是其提供使用者以標籤 (Tag) 作為分類的工具之一。

自 Internet 興起以來，為了方便使用者能夠連結與記錄網站的網址，而不需背誦長長的 URL 網址，以 IE 與 FireFox 兩大瀏覽器而言，均提供書籤(Bookmark)的功能讓使用者能夠記錄、保存、或整理網址，當使用者下次想走訪該網站，只需要找到相對應的書籤就可連到網站。在微軟的瀏覽器 Internet Explorer 裡，此機制被稱呼為我的最愛 (My Favorite) 在使用者工具列中提供使用者快速存取。隨著 Web2.0 的發展，網路使用者在網路上增加內容機會大增，網路上可獲取的資訊、知識日益增多，使用者所保存的「我的最愛」或「書籤」亦隨著日益增多。使用者不僅要面對網路上龐大的資訊量，尚要面對整理重組自己所保存或感興趣的「我的最愛」，以方便日後找尋。目前現行的瀏覽器中 (包含 FireFox, IE)，均以樹狀階層化的方式在歸類書籤。此種方法為傳統生物分

<sup>1</sup> Del.icio.us: <http://del.icio.us/>

<sup>2</sup> reddit: <http://www.reddit.com/>

<sup>3</sup> 黑米共享書籤: <http://www.hemidemi.com/>

<sup>4</sup> Flickr: <http://www.flickr.com/>

<sup>5</sup> Google Image: <http://picasaweb.google.com/>

類學（界、門、綱、目、科、屬、種）、傳統的圖書館圖書分類、以及目前作業系統中的檔案分類系統（包含微軟視窗作業系統系列、蘋果電腦的 MacOS、與 Linux 等）所採用，在此階層化的架構中，使用上使用者會從最上層的分類開始往下找細的分類，直到找到所要的書籤為止，是一種 Top-Down 的方式，上層的分類多半具有一般化的特性，而下層分類則具有特殊化的特性。一般來說，為了使書籤避免繁多雜亂，通常只會加入一個網站最上層的首頁當作書籤，但當你想找某一篇特定文章而不知出自於哪個網站的時候，反而會難以找尋。因為網路關鍵字搜尋技術的發達，此時會乾脆放棄使用自己建立整理的書籤，使用搜尋引擎來直接找，這樣就失去了當初花費時間整理的意義。

直到後來網路上社會性書籤服務興起，使得使用者可以在收藏連結的時候，寫下一些備註，以便供自己或其他使用者日後參考或檢索，最有特色的一點就是可以對連結非階層式地加上使用者所慣用的關鍵字以區辨一個連結的內容描述（Content Description）或概念（Concept），一般的做法是給連結加上不同的標籤（Tag）。標籤（Tag）是一個被關連到或被指派到某個資訊的關鍵字，資訊可以是一張圖片、一篇文章或是一段影片。標籤通常是非很正式的且隨個人的觀感而被它的創造者或使用者所定義，所以它可以很有彈性地被使用。沒有人規定它該怎麼使用，所以每個人可以針對他對這項資訊的了解，來為資訊貼標籤，並作為整理資訊的工具。這裡所提到的「社會性」書籤則特別著重其可分享在不同使用間的特長，意即不同使用者之間，彼此可分享他們相同標籤的連結，這種社會性的分類機制被稱為分眾分類法（Folksonomy）。此種分類方式為一平面化的網路架構，沒有階層之分，由於標籤是使用者慣用的自行定義，雖然不夠嚴謹，但使用上顯得方便、靈活。分眾分類法為資訊標標籤，是為了想要讓資訊的本體更容易、更快地被找到，從某個資訊開始看到誰下了這個標籤，他還下了那些其他的標籤，用這種方法去找尋其他使用者傾向使用哪些標籤，並對其他內容下有意義的標籤。

社會性標籤似乎成功地從分享與自我整理的角度解決了使用者在整理網路上鉅量資訊的問題。然而，就目前使用者所賴以上網、收集資訊、生產資訊的個人電腦的作業系統而言，有如上述書籤的歷史一般，自從人類使用電腦以來，一直都依靠著階層式檔案系統，以視覺化資料夾來整理檔案。因為階層化的歸類方式，具有一般性與特殊性的考量，最上層的分類過度一般化（General），而最下層的分類則容易過度特殊化（Specific），成為階層化的儲存方法具有如同歸類網路上最愛連結時的缺點。以找尋一篇論文為例，例如以社會網路分析找尋生物基因網路中的基調，該篇論文橫跨了資訊科學與社會學的社會網路分析和分子生物學等二種領域，使用者在歸類時若把它放入其中一種領域的資料夾中，那麼在另一種領域的資料夾裡就看不見它的存在，不論是放入哪個資料夾，都會導致資訊的喪失，以及限制了日後找尋的方式。使用者可能考慮複製兩份檔案在兩個

資料夾中，然而則會有檔案重複的問題；或者當使用者對文字類型檔案加以註解，則必須對兩份檔案加以同步化。不僅文字檔案，其他多媒體資料的分類，如以圖片為例，亦會遇到如此的問題。例如給定一張 Wii Sport 裡在玩網球遊戲的螢幕畫面，有的人會把它放到網球的資料夾，有的人會放到遊戲的資料夾，假設今天放到網球的資料夾中，之後想找遊戲的圖片就找不到了。反之亦然，所以整理時的思考方式會限制找尋時的思考方式。

近年來，除了在 Windows 本身所提供的檔案搜尋功能，或像是 Mac OS X 內建的 Spotlight 搜尋功能，以及 Google Desktop Search 以強大的搜尋引擎幫助個人在自己的電腦中搜尋資料，微軟最新推出的作業系統 Windows Vista 裡也有內建相似的功能。讓使用者可以不用辛苦的整理檔案，只要打幾個關鍵字即可搜尋到自己所需要的資訊。在網路上盛行的搜尋引擎，如今也逆傳回個人電腦，不過如同使用網路上的搜尋引擎來搜索東西一樣，使用者在下關鍵字以找尋檔案的時候，針對所要找的資訊，關鍵字難以定義與回想，並且所找尋到的結果，呈現方法不具有結構性和關聯性，系統在將與該關鍵字依照相似度排列後，即以表列 (List) 的方法呈現，使用者所見會受到系統預設相似度搜尋演算法所影響，不盡然為使用者所需要的結果。在前述搜尋論文的例子中，若使用 Google Desktop Search 之類的搜尋引擎，可利用論文標題或內文所擁有的關鍵字來做搜尋，但是關鍵字找到的論文未必就是所想找的，跟論文內容本身和關鍵字的相關度有關。再以前述圖片搜尋為例，搜尋引擎目前對圖片的尋找比文字類的檔案困難，因為圖片中不存在關鍵字無法搜尋，只能依靠檔名的搜尋，若檔名都是一堆流水號的話，可能就很難找到了。基於上述的情況，我們渴望有其他的方式來幫助人們整理檔案，以彌補上述二種方法的不足。

為了要幫助人們整理及找尋電腦裡堆積的檔案，我們需要先了解人類記憶的過程。就像在儲存檔案時可以想像成在記憶一件事物，大腦裡保存著所有記憶，到要用到時再去回想起來，也就是平時電腦裡儲存著檔案，需要時再將檔案尋找出來。所以我們先研究認知心理學(cognitive psychology)，它關心的是人們對訊息如何知覺、學習、記憶和思考，從裡面挑出記憶模型<sup>6</sup>的部分理論，做為系統的主要核心概念。以認知心理學為基礎貼近人們的思考模式，以期能真正改善以及幫助人們做資料的尋找和再利用。

## 1.2 研究問題

基於上述的動機，本研究欲探討的問題分別如下：

---

<sup>6</sup> 詳細於 2.3.1 節做說明

- 一、以認知心理學中的記憶模型為基礎的標籤(Tag)系統，是否比原本的資料夾結構，更能幫助人們找尋、提取曾經儲存過的檔案？
- 二、使用資料夾與標籤系統在找尋所需檔案時，各自擁有哪些特色？
- 三、以資料夾與標籤系統建立出來的架構，各自擁有何風格？以及是否存在某種關連，對其用不同系統的結果會造成何種影響。

### 1.3 研究重要性

電腦做為人類社會越見重要而普遍的資訊處理與儲存媒介。如何快速有效的管理、搜尋、索引電腦中的大量個人資訊，成為電腦系統做為人類重要外部記憶媒介以及解決日常問題的常用工具時所需要面對的共通重要課題。這種需求反應在各種新型人機介面(Human Computer Interface)的設計上。本篇論文將目前在現階段 Web2.0 服務中大量被使用的標籤系統設計引入於個人電腦的環境，並且與傳統個人知識重整時常見的資料夾系統進行比較。

本論文的重要性在於以下幾點：

- 引進認知心理學的理論來解釋，標籤為何能比資料夾更有效地整理及提取資訊。並且比較探討兩者在實際應用時的不同點。
- 設計結合記憶模型、記憶歷程和標籤網路結構為基礎設計個人化環境時適用的系統，並藉由實驗來驗證標籤與認知結合的系統能幫助人們做記憶的反芻與提取。

## 二、 文獻探討

### 2.1 背景探討

人們使用階層式架構，除了之前提過的檔案資料夾系統外，其實從很早以前就開始，直到今日仍廣泛的使用中。以下舉幾個例子說明。

- 一、圖書館的藏書分類：像是整理中文書的「中國圖書分類法」，此分類法目前在台灣用得相當普遍，無論公共圖書館、大學圖書館、中小學圖書館，甚至專門圖書館，採用此法整理圖書資料者比比皆是。它分為 000 總類、100 哲學類、200 宗教類、300 自然科學類、400 應用科學類、500 社會科學類、600-700 史地類、800 語文類、900 美術類，以下各種類再繼續細分，是一個很標準的樹狀結構。
- 二、生物學：地球上的生物都依據一套物種分類規則來進行分類，此規則中訂出生物分類的七大層級，從高到低依序為界(Kingdom)、門(Phylum)、綱(Class)、目(Order)、科(Family)、屬(Genus)、種(Species)，階層越高代表所包含的生物種類越多，較低的階層包含的種類就較少，但彼此之間的構造特徵卻較相似。而這七大層級之間的關係簡單來說為，界擁有許多門，門擁有許多綱，綱擁有許多目，目擁有許多科，科擁有許多屬，屬擁有許多種，種是最基本單位。是一種由上到下的階層式結構。但不是所有的物種都可以順利的分類，像是鴨嘴獸的分類，由於牠是哺乳動物又會產卵，曾造成生物學家的爭論，最後終於在哺乳綱、原獸亞綱、單孔目找到一個位置，成為哺乳動物中唯一一個不是胎生的動物。
- 三、網站架構：在建立設計一個網站時，需要事先做好規劃，以吸引人氣和留住瀏覽的人潮。對於網頁的配置，為了能讓瀏覽的人有順暢的瀏覽經驗，以及減少到達瀏覽者想看頁面的連結數，所需的連結層數(次數)愈多將會影響瀏覽者的觀看的意願，所以一般而言，會先規劃個網站地圖，圖 2.1.1 為個人網站的設

計架構：

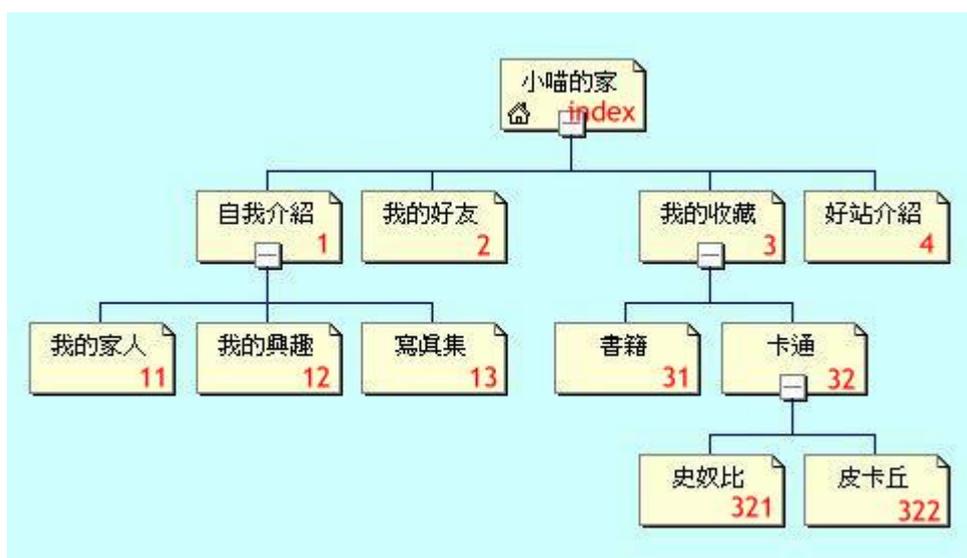


圖 2.1.1 個人網站架構範例

可以看得出它本質上仍是一個樹狀階層式架構。

## 2.2 標籤(Tag)技術

人類在看到一篇文章時，可以很快地將其歸類在所屬的領域裡，例如：經濟、政治或運動。Florian Wolf, Tomaso Poggio 和 Pawan Sinha 提到說其實只需幾個字詞，以這些字詞來為文章分類，就可以近似原本文章分類的精確度[1]，也就是我們可以把文章濃縮成幾個字詞。因為標籤本身也就是字詞的一種，如果把字詞替換成標籤，就代表著對文章下幾個標籤後，我們就可以用這些標籤來對文章做分類。

而透過標籤技術，要如何去組織資訊內容呢？在下標籤的同時，對於每個標籤會對應到一個或多個資訊內容上，每個資訊內容也會對應到一個或多個標籤上，彼此間無任何方向性，我們可以利用這些關連來做進一步的利用，以下將做更詳細的說明。

從六個人的小世界(SIX DEGREES)此書中的理論可知，標籤和資訊內容本身為所謂的關聯網路(Affiliation Network)，當二個標籤共同標示於同一個檔案上，就被認為是「相關連的」。關聯網路永遠由二種節點所組成，最好的表示方法為二分網路(bipartite network)，所有節點可被分為二群，同一群內的節點彼此不相連。如圖 2.2.1 的上方，數字編號的為一群，英文字母編號的一群。若將二分圖展開來，拿掉其中一群的節點，將可得到剩下那群節點的直接連結圖。如圖 2.2.1 的下方，拿掉數字的節點展開後，可得到字母的連結圖。

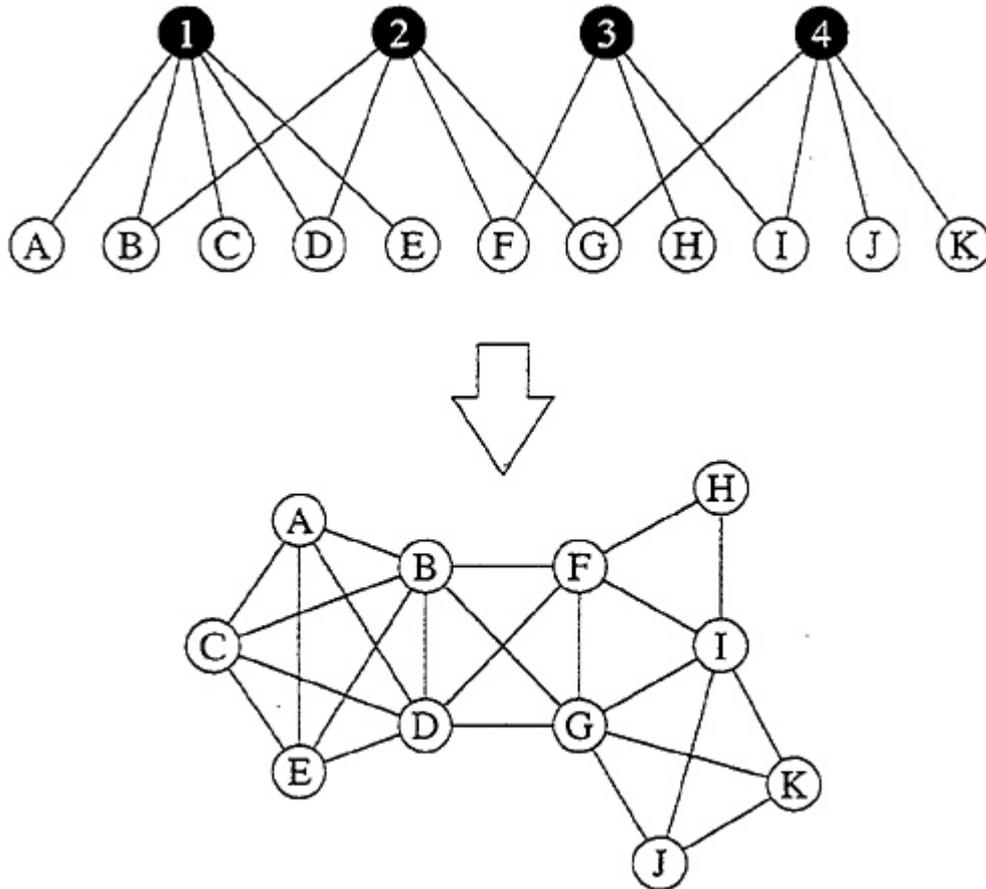


圖 2.2.1 二分網路(bipartite network)

若將資訊內容(Content) 和標籤(Tag)繪製成二分網路，就會像錯誤! 找不到參照來源。一般，可被分為二個單元網路，一個是資訊內容的網路(Contents Network)，而另一個則為標籤網路(Tags Network)。

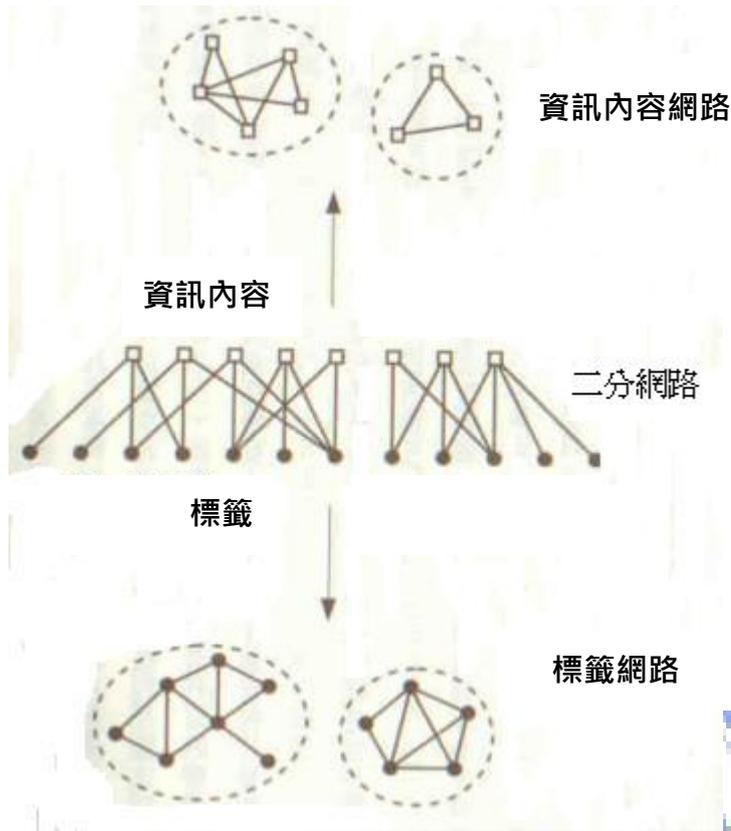


圖 2.2.2 資訊內容(Content) 和 標籤(Tag) 的二分網路

藉由資訊內容的網路，我們可以用資訊找資訊，像是今天拿到一篇論文，可由此網路找出與它相關的論文。藉由標籤網路，可以用標籤找相關標籤下的資訊，從一種領域找到另一種可能相關的領域，例如假設在「遊戲」這個標籤下，有一張 Wii Sport 打網球的圖，圖片擁有二個標籤：遊戲、網球，那麼由標籤網路可找到「網球」這個相關的標籤，進而去尋找其他打真實網球的圖。

CiteULike<sup>7</sup>是一個可整理書目引文資料的網站，如圖 2.2.3。它延續了幾個 del.icio.us 特色：以個人標籤為主的分類組織、可以瀏覽其他人分享的資料，並且提供了匯出匯入的功能，可以跟研究者的書目管理軟體配合使用。CiteULike 能讓使用者訂閱特定期刊的書目資訊，使用者可以從訂閱資訊中選擇是否要加入自己的書目資料庫。他跟一般書目管理軟體不同的地方是，當初步探索特定主題文獻時，可以透過其他人的書目資訊，類似滾雪球的方式，找出許多相關文獻。在滾雪球的過程中，可以發現引用收錄次數較高的文獻，則這篇文獻相對可能較為重要。另外，可以學習其他收錄者用何種標籤描述這篇文獻，可以幫助檢索者進一步擴張或是限制自己的檢索策略。提供 RSS，可以訂閱

<sup>7</sup> CiteULike: <http://www.citeulike.org/>

特定主題或特定人物的書目資訊，能第一時間取得最新資訊。提供社群功能，可以協同管理或編製特定書目。他將標籤做了一個很好的應用，研究者只要收集記錄自己找到的論文，就可以從此為出發點，得到大量其他人整理的資料。

The screenshot shows the CiteULike website interface. At the top, there is a navigation bar with the CiteULike logo and a search bar. Below the search bar, there are several tabs for different subject areas: Your Library, Computer Science, Biological Science, Social Science, Medicine, Engineering, Economics/Business, Arts/Humanities, Mathematics, Physics, Chemistry, Philosophy, and Earth/Environmental Science. The main content area is titled "Everyone's library" and displays a list of recent papers posted to the site. The interface includes a navigation menu on the left, a search bar at the top right, and a list of active tags on the right side. The main content area displays several article entries with titles, authors, and publication details.

圖 2.2.3 書目管理網站 CiteULike



## 2.3 記憶

### 2.3.1 記憶模型

人類藉由記憶(memory)來保存及提取過去的經驗，以便未來的使用。在認知心理學的研究上，不同的認知心理學家對於記憶有不同的詮釋方式。不同詮釋方法上的差異，主要是因為所用來了解記憶的比喻(metaphor)的不同所導致[2]。隨著新研究的發現，後人會對其比喻提出修正，或是直接提出一個新模型。

傳統且被人廣泛接受的記憶三段儲存模型(three-store model of memory)，是 Richard Atkinson 和 Richard Shiffrin 於 1968 年所提出[3]。他們將記憶視為三個儲存庫：

- (1) 感覺儲存(Sensory store)：儲存非常短暫且數量有限的訊息。
- (2) 短期儲存(Short-term store)：可以將訊息儲存較長的時間，但容量仍有限。
- (3) 長期儲存(Long-term store)：有非常大的容量，能長久儲存訊息，甚至無限期中地儲存。

他們以不同名稱來區別結構和儲存在結構中的訊息，稱結構為儲存庫(store)，稱儲存在結構中的訊息為記憶。後來的認知心理學家稱三段儲存系統為感覺記憶、短期記憶和長期記憶。如同前面提到記憶的比喻，Atkinson 和 Shiffrin 並不認為三個儲存庫擁有獨特的生理結構，而是一種假設性結構(hypothetical constructs)，它指的是本身無法被直接測量或觀察的概念，但是可以幫助我們容易去了解記憶如何運作的心智模型。

Atkinson 和 Shiffrin 後來的研究將記憶的三段儲存模型化為圖 2.3.1 呈現，來說明儲存結構的系統，強調被動的儲存庫來存放記憶[4]。

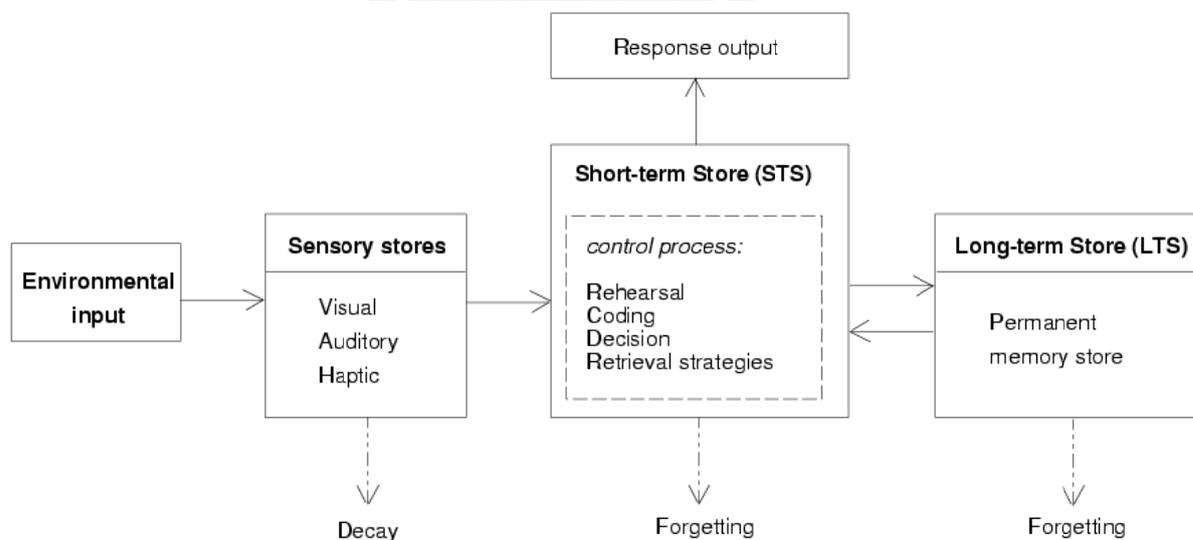


圖 2.3.1 記憶的三段儲存模型

另一種假設，處理層次架構(levels-of-processing framework)是由 Craik 和 Lockhart 所提出的[5]。它與記憶的三段儲存模型不同，他們假設記憶並非由幾個儲存區所組成，而是假設儲存沿著一個連續的向度(即收錄的深度)，而有所變化。此研究假設在理論上有無限數量的處理層次(LOP)，它用來收錄項目，而層次與層次之間並沒有分界線，並強調處理是儲存的關鍵，訊息被儲存的層次大部分取決於它被收錄的方式。Brown 和 Craik 後來的研究顯示，一般而言處理的層次越深，項目會被提取出來的機率也越高[6]。Craik 和 Tulving 以一系列的實驗來支持這個觀點[7]。他們先將一串字給受測者看，然後在每一個字前面加上一個問題，依問題的變化來引發三個不同層次的處理，依淺到深的順序排列，分別為：物理的、聽覺的、語意的，如表 2.3.1 所示。實驗的結果呈現，問題引發的處理層次越深，回憶的表現會比較好。

表 2.3.1 三個不同的處理層次

LEVEL OF PROCESSING	BASIS FOR PROCESSING	EXAMPLE
Physical	Visually apparent features of the letters	Word: TABLE Question: Is the word written in capital letters?
Acoustic	Sound combinations associated with the letters (e.g., rhyming)	Word: CAT Question: Does the word rhyme with "MAT"?
Semantic	Meaning of the word	Word: DAFFODIL Question: Is the word a type of plant?

工作記憶(working memory)被定義為長期記憶的一部分，且包含了短期記憶。它是由一些心理學家從不同角度來看短期記憶與長期記憶[8-11]。工作記憶只保留長期記憶中近期被活化的部分，且在短期記憶儲存內外移動被活化的部分，如圖 2.3.2。其實我們可以把它視為一個加工廠，加工完後的資訊送到長期記憶中，標標籤的動作可比喻成加工的動作。



圖 2.3.2 工作記憶的示意圖

在古典知識論中，就知識結構分成二種類別：

(1) 陳述性知識(declarative knowledge)：可被陳述的事實

(2) 程序性知識(procedural knowledge)：可被執行的程序

由於儲存在電腦裡的資料，皆屬於陳述性知識，所以將以陳述性知識為主軸來設計系統架構。

在 1983 年 Anderson 提出的 ACT-R 的模型中，包含了陳述性知識的網路表徵部分 [12]。他的陳述性網路，像語意網路一樣，概念儲存在網路上各個節點，節點可分為活動和不活動的二種狀態，活動的節點也可以說是被激發的狀態。主要有二個原因會造成節點被激發，一種是直接地由感官或外界刺激來被激發，另一種則是間接地從已激發的節點的活動來使鄰近的節點也被激發。若是當節點不斷地激發鄰近的節點時，活動節點的範圍就會開始拓展開來，直到達到有限的活動節點數，此稱為擴散性激發(spreading activation)，在網路中沿著一組節點而散開。對於系統中的資訊，我們將以網路為其架構，以標籤做為節點，把檔案同時存在二個標籤的關係，畫為二個節點中間的連線，將節點的擴散激發化為標籤和相關(相鄰)標籤的激發。

### 2.3.2 記憶歷程

認知心理學家一般指的記憶主要歷程包含三種基本的運作：

(1) 收錄(encoding)：指如何將一個物理的、感覺的輸入，轉換為可以放入記憶中的一種表徵。

(2) 儲存(storage)：指如何將已收錄的訊息保留在記憶中。

(3) 提取(retrieval)：指如何取得已儲存在記憶中的訊息。

每一個運作分別代表記憶歷程中的一個階段[6, 13, 14]。三種運作通常被視為有順序性的過程：首先先接收訊息，接著將它保留一段時間，後來再取出來使用。John Bransford 和 Marcia Johnson 請他們的實驗受測者閱讀一段文章，說明較為抽象，在嘗試收錄文章時，會發現難以收錄，以致於儲存和提取也變得困難。可是如果給予受測者這段文字一個標題，大部分的人會表現得比較好。語言的標籤幫助我們收錄並記住原本難以理解的文章[15]。「標籤」原本就是一個語言上的字詞，以此實驗的結果，我們可預期加標籤會幫助電腦中檔案的收錄。

由於我們想將訊息長期保存後，還能再次地提取出來，所以接下來將把焦點鎖定在長期記憶的方面。

大部分儲存在長期記憶裡的訊息似乎主要是以語意的方式收錄，也就是說以字的意義來收錄。一些不同類型的實驗結果[16, 17]可以當作證據，證明語意的收錄方式被使用於長期記憶中。例如 Grossman 和 Eagle 的實驗[16]，要求受測者記下 41 個不同的

單字，結束後五分鐘時，請受測者從清單中勾選剛才看過的單字。清單中含有非剛才記憶的單字，其中 9 個混淆字與原本的 41 個字有語意上的關連，還有 9 個混淆字與原本的沒有關連。最後受測者錯認了較多的同義字，對於較無關連的字錯認較少，顯示語意混淆的可能性較大，可推測主要是以語意方式收錄。事實上標籤本身就是一個字，也就是說以標籤來收錄儲存可相當於以語意來收錄。

訊息由短期記憶轉移到長期記憶需要一個不同的歷程，名稱為固化(consolidation)。它的方法是在新訊息和我們已經知道或了解的訊息間建立連結或聯想，藉由將新訊息合併到儲存訊息的既有基模中來建立連結，達成轉移的目的。若將基模比喻成系統內的網路結構，那麼在加標籤的過程中，會依資訊內容和標籤的關係，自動建立連結。

記憶的失敗主要是來自提取失敗而非儲存失敗，由 Endel Tulving 和 Zena Pearlstone 的實驗結果所得[18]。他們的實驗中，受測者會先聽到類別名稱，例如「文具類」，接著會聽到該類別的字彙，如：鉛筆、尺、原子筆、橡皮擦…。之後分成二組，一組採自由回憶，請受測者儘可能回憶出剛才出現過的字彙；另一組採線索回憶，給予受測者類別名，請他回憶該類別的所有字彙。最後的結果得到，平均來說線索回憶較自由回憶好很多，若只讓受測者自由回憶，可能會得到人不能儲存很多字的結論，但此結果證明了記憶的失敗主要是來自提取失敗而非儲存失敗。

Gordon Bower 和他同事(Bower, Clark, Lesgold, & Winzenz)的研究顯示分類如何戲劇性地影響提取[19]。接受階層性呈現的受測者回憶出 65%的字，而接受隨機順序呈現的受測者只回憶出 19%。顯示有結構性的分類記憶，比單純隨機記憶更能幫助未來的提取，與 Tulving 和 Pearlstone 的結果相呼應。所以我們想要藉由新設計出來的系統，比傳統資料夾結構更能幫助人們對記憶的提取。

有時候我們在提取訊息時會出現困難，有記憶過的訊息不一定想得起來。以下說明二個名詞：

- (1) 可獲得性(availability)：訊息有儲存在長期記憶中
- (2) 可及性(accessibility)：我們觸及可獲得訊息的程度

記憶的表現須視記住訊息的可及性而定，希望藉由這個系統能幫助人們使用電腦做為長期記憶區，提高可及性趨近於可獲得性。

2001 年 Nielsen 提到再認(recognition)比回憶(recall)來得好[20]，傳統資料夾系統總是要使用者做回憶出檔案置放的位置，而標籤總是要使用者作再認，不斷地再認現在所看到標籤，看是否有關聯。以使用性上來說，標籤會比較好。

收錄特定性(encoding specificity)由 Tulving 和 Thomson 於 1973 年所提出[21]。收錄特定性主要在說回憶什麼決定收錄什麼，因為之前的研究者各種提取實驗的結果都顯

示，訊息如何被收錄對於如何被提取以及提取得多好有強烈的影響。Watkins 和 Tulving 的實驗[22]漂亮地證實了收錄特定性。他們讓受測者記憶 24 組配對單字，一個是原本要記憶的單字，另一個是和單字有關的刺激字，例如：寒冷—雪，雪作為寒冷的刺激字。接著請受測者圈出曾看到的單字，平均再認出 60% 的單字，然後給予 24 個刺激字，請受測者將記得的單字背出來，結果回憶出 73% 的單字。就一般來說再認通常都會比回憶來得簡單，因為只要說有沒有過這個字就好，比起要記誦出曾出現過的字容易。但這個結果反而是回憶比再認好，刺激字比原來單字本身對單字來說是更好的線索，因為當初是以配對的方式來記憶的，突顯收錄方式深深地影響將來提取的成功。

由於資料夾結構的特性限定了資訊收錄的方式，一次只能被分配到其中一個資料夾底下，意味著只能以一種類別來收錄，最多就是轉換到另一個類別。但是標籤本身的特性就可以為同一個資訊標上許多的標籤，可在不同時間為資訊決定要收錄的類別，累積各種不同的收錄方式。所以在收錄時的多樣性，可幫助提高提取成功的可能。

若個人產生自己的提取線索時，這個線索會比其他人所產生的更強而有力。此結果由 Greenwald 和 Banaji 所發表[23]。Timo Mantyla 的實驗可證實這點[24]。他給予受測者一串目標字的列表，請他們為每個字產生另一個字，也就是線索，然後只給他們線索字的列表，請他們回憶出目標字，結果發現他們幾乎可以無誤地記得 500 和 600 字的列表。所以我們讓使用者自行建立自己的標籤，也預期能比網路上他人建立的標籤來得更提高提取的效果。

## 三、系統架構與設計

### 3.1 核心概念

根據上一章所談到的認知心理學的理論，以及使用標籤來做為主要的工具，來設計一個系統，做為一個新的整理工具，以期能達到知識重整的效果。以下各點分別說明一個概念：

1. 將標「標籤」視為一種收錄的過程。  
由於就收錄本身的意義來說，就是將外界的輸入轉化為記憶的一種表徵，所以我們可以把標籤本身當做一種表徵，那麼標標籤的動作就成為一種收錄的過程。
2. 將有標過標籤的資訊視為長期記憶。  
之前提到長期記憶的訊息主要是以語意的方式收錄，標籤本身就是一個字，當你以標籤來為資訊做收錄時，等於是語意來收錄。在三段記憶的模型當中，經過收錄後的訊息將會儲存在長期記憶中，所以經過標「標籤」的資訊也就成為長期記的一部分。
3. 系統儲存使用標籤網路架構。  
標籤網路可視為陳述性網路，因為同樣是語意所構成的。將標籤視為節點，與資訊內容的節點畫為二分圖(bipartite graph)，隱藏資訊內容的節點，可得到標籤網路。節點與節點中間的連線，即為同時擁有二端節點標籤的資訊。
4. 選擇標籤將提示其相關的標籤，可視為激發過程，並反過來刺激感官。  
當使用一個標籤時，將激發其鄰近的相關標籤並將其呈現出來，讓使用者可從畫面中得知這些相關標籤，近而造成人類的感官刺激，引起腦內陳述性網路節點激發。
5. 頁面中可供處理的檔案視為工作記憶，為長期記憶的一部分。  
讓使用者可對頁面中資訊下標籤的地方為加工區，是工作記憶的一種比喻。
6. 頁面中檔案的順序依建立標籤的先後排序，相當於記憶中近期被活化的部分。  
對於人們來說，最近記憶的東西在回憶上較為清晰，也就是說近期才對它做收錄的動作，以及被活化，會在系統頁面中排在較前面的位置上。

## 3.2 系統概念架構

原本記憶的三段儲存模型再加上記憶歷程的基本操作，我們可以得到圖 3.2.1 的認知模型。

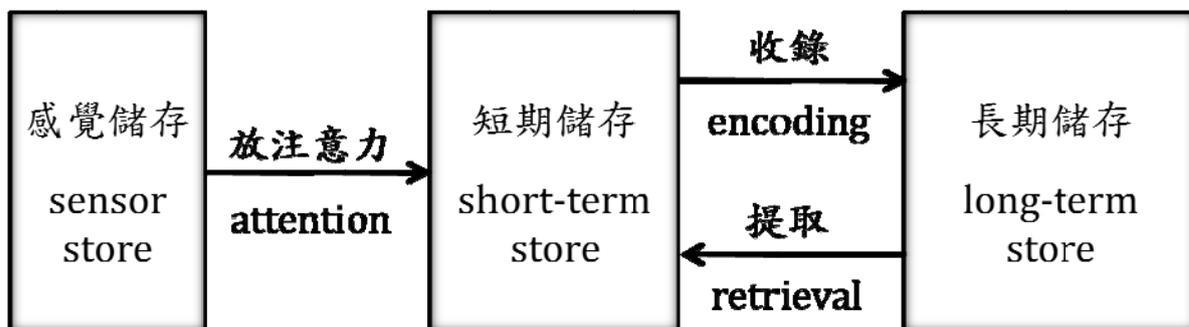


圖 3.2.1 三段儲存模型與記憶歷程的基本操作

根據核心概念來對上述認知模型做整合，可以設計出本論文使用系統的概念架構，如圖 3.2.2。由於感覺儲存依舊是依靠人類的部分，所以不做任何更動，但不屬於系統內的結構。

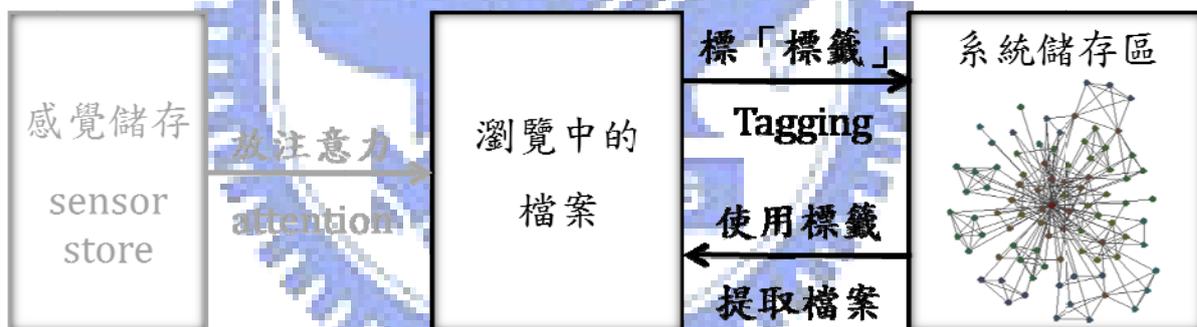


圖 3.2.2 概念架構

## 3.3 實作內容

本研究以 3.2 節的概念架構為基礎，以此來發展實際的系統。建立系統的主要語言為 Python，因為 Python 對於網頁後端的處理功能強大，對於架構整個系統有所幫助。圖 3.3.1 繪出系統實作的架構圖，先於儲存裝置上置放所有實驗用的圖片，以及架設資料庫用以儲存標籤和檔案的資訊。中間的大方塊是系統中心部位，其中一個單位負責圖片的讀取，一個單位處理從資料庫讀出來的資料，建立出檔案及標籤的關連。其中一個最重要的單位，負責計算指定標籤的相關標籤、標籤使用率的排序，和選擇某固定標籤時計算有關檔案的排序，詳細的說明在後面章節會再提到。最後當使用者操作時，我們配合前面計算出來的結果和讀取的圖片作畫面上的呈現，使用者介面以網頁的方式讓人

瀏覽，以及提供簡便的操作過程，完整的使用者介面將在下一節來做說明。

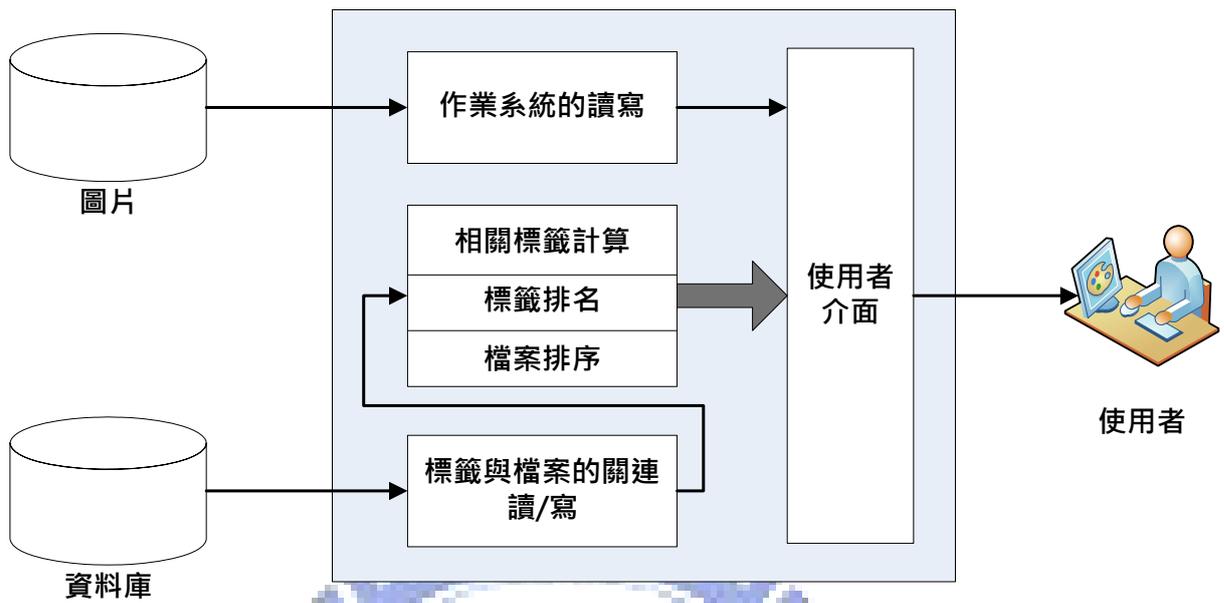


圖 3.3.1 實作架構圖

### 3.3.1 使用者介面

圖 3.3.2 為實作出來的系統介面圖，右方邊欄為使用者所建立的所有標籤，左上方的的是標籤網路由目前標籤所得到的相關標籤。與目前標籤最相關的，也就是說與目前標籤一起同時存在於相同資訊數愈多，就愈相關，會置於相關標籤的愈前面。

# 個人知識重整

按這裡回到首頁

## 標籤網路

 標籤搜尋

目前標籤: [藍天](#)

小提示: 按下目前標籤 做退一步搜尋

相關標籤: [台灣](#) [101](#) [白雲](#) [大樓](#) [戰鬥機](#) [飛機](#) [台北](#)

小提示: 按下相關標籤 做進一步搜尋

3張 頁數

[1 \(37\).jpg](#)

標籤: [飛機](#) [戰鬥機](#) [白雲](#) [藍天](#)

飛機,戰鬥機,白雲,藍天,

確認



相關圖片

[1 \(35\).jpg](#)

標籤: [飛機](#) [戰鬥機](#) [白雲](#) [藍天](#)

飛機,戰鬥機,白雲,藍天,

確認



相關圖片

## 所有標籤

將圖片加上此標籤(關鍵字)

- [可愛](#) [新增](#)
- [鳥](#) [新增](#)
- [棒球](#) [新增](#)
- [花](#) [新增](#)
- [飛機](#) [新增](#)
- [綠色](#) [新增](#)
- [葉子](#) [新增](#)
- [藍天](#) [新增](#)
- [籃球](#) [新增](#)
- [機車](#) [新增](#)

圖 3.3.2 系統介面

## 介面說明

### ● 圖片資料

[1 \(3\).jpg](#)

標籤: [鍵盤](#) [USB](#)

鍵盤,USB,

確認



相關圖片

### ● 從上到下，各行為：

1. 檔案名稱
2. 擁有的標籤
3. 編輯標籤的方塊
4. 確認按鈕

### ● 標籤邊欄

- ◆ [鍵盤](#) [新增](#)
- ◆ [USB](#) [新增](#)

- 新增：將選取的檔案全部加上此標籤

- 邊欄下方

- 全選按鈕:為選取所有圖片。但是當你已有選取時，會變成是反向選擇。

詳細的使用方法於附錄二說明。

### 3.3.2 資料庫設計

本研究使用 MySQL 來做為資料庫，資料庫中主要存放三個表格，分別為標籤與檔案的關係(Tag-File Relation)、檔案資訊(File)及標籤資訊(Tag)，如圖 3.3.3 資料庫架構圖所示。首先，為每一個檔案記錄其編號(File id)和儲放於系統中的路徑(File path)，每一個標籤有編號(Tag id)、標籤名稱(Tag name)和標籤被使用的次數(Used)。當使用者為某一檔案加上一個標籤時，會建立一筆標籤與檔案的關係(Tag-File Relation)，由於不限定一個檔案可以有幾個標籤或一個標籤可以標多少個檔案，所以標籤與檔案的關係對於檔案或標籤都是多對一的關係。標籤與檔案的關係裡存放編號(Relation id)、對應的檔案(File id)、對應的標籤(Tag id)及最後修改這個關連的時間(Access time)。

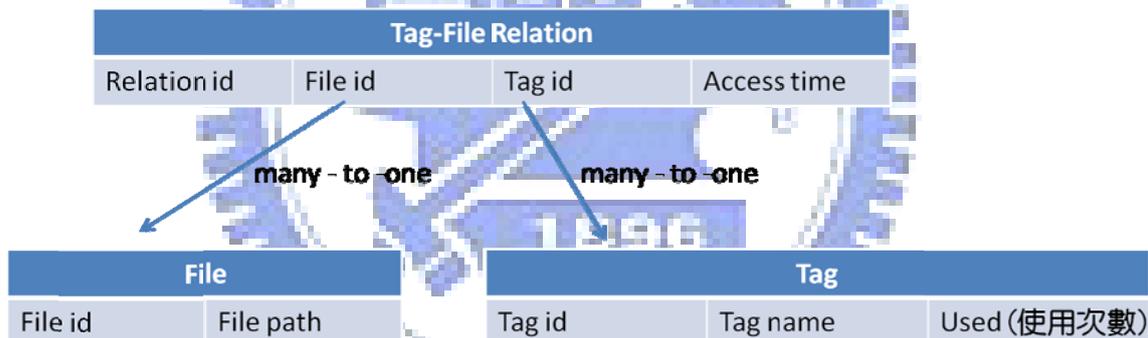


圖 3.3.3 資料庫架構圖

建立起資料庫後，在系統標籤與檔案關連讀/寫的這個單元內，使用 SQL 語法來對表格做合併及查詢的動作，取出目前所需的資料提供給系統使用，以便後續的計算及呈現。

### 3.3.3 資訊內容的選擇

本研究使用圖片來做為實驗時所用的資訊內容，系統本身可支援所有的檔案，例如：影片、音樂、圖片、文件...等的檔案，由於為了簡化實驗的流程，以及降低後續分析的複雜度，我們選擇了圖片當做為後來實驗的內容，原因有下列二點：

1. 選擇圖片當資訊內容不會影響搜尋的方式或結果。因為在系統中要做搜尋的動作，

不會因為圖片有縮圖而與文字檔有差異，原因在於使用者於系統中不論是哪種格式的檔案，皆利用標籤本身及其關連標籤來做搜尋，並非由縮圖來做判斷、搜尋，所以以圖片為內容不會使結果有所偏頗。

2. 圖片本身除了檔名之外，並不包含任何文字，所以在使用現有的搜尋系統，例如：作業系統內建的搜尋、Google Desktop Search...等，使用關鍵字為搜尋基礎的系統難以找到使用者所需要目標，若本研究的系統可以對圖片這類型的檔案做出不錯的成果，相信可以解決目前系統遇到的問題。

系統使用 Python 提供的 API 來讀入圖片，處理作業系統做 I/O 的問題，並秀出於使用者介面中。

### 3.3.4 系統功能計算

此為本研究最主要的部分，實作出之前討論的概念設計，分為三個部分：相關標籤計算、標籤排名、檔案排序，以下以各個小節詳細說明實作部分。

#### 3.3.4.1 相關標籤計算

同一個檔案上，若有標上二個以上的標籤，就認定他們互為相關標籤。二個標籤同時標記一個檔案時，則將此對標籤關係度加一，同時出現於愈多檔案上，則此二標籤關係度愈高。當使用者選定 A 標籤時，系統介面將會列出 A 標籤的所有相關標籤，依照與 A 標籤的關係度愈高，則在相關標籤排愈前面(也就是愈左邊)，如圖 3.3.2 左上方的虛線框內。

#### 3.3.4.2 標籤排名

如同前面 3.3.2 節所說的，我們為每一個標籤記錄使用的次數，每當這個標籤被標到某個檔案上一次，就將它的使用次數加一，使用次數愈多表示使用者愈常用到這個標籤，依常用度來為標籤排名。在系統介面(圖 3.3.2)的右下方，將排序好的標籤放於此處，方便使用者操作。

#### 3.3.4.3 檔案排序

當使用者選定標籤後會在目前標籤欄目上顯示出來，介面下方會秀出所有使用此標籤的檔案，而順序的先後依這個檔案最後標上此標籤的時間來排序，愈晚標上此標籤的檔案表示近期才對其做處理和反芻，對記憶理論來說還位於短期記憶內，所以將之排於愈上方，加速使用者存取。

## 四、 實驗方法與結果分析

### 4.1 研究對象

研究對象為磐石高中三年級的學生，參與本次研究的有二個班級，使用資料夾的學生有 27，使用個人知識重整系統的學生有 37 人。第一階段完成實驗的有效樣本數，資料夾組有 21 人，新系統組有 37 人。第二階段完成實驗及問卷的有效樣本數，資料夾組有 17 人，新系統組有 32 人。

### 4.2 研究工具

在研究工具上我們使用三種工具，最主要的工具是第三章所提到的個人知識重整系統，其他二個工具一個是在使用資料夾做實驗時，記錄使用者動作的小程式(4.2.1 節)，以及最後做實驗者背景調查的問卷。以下將對剩下二種工具進行說明。

#### 4.2.1 記錄結構建立及搜尋的過程

在實驗者使用原作業系統的資料夾來整理檔案時，為避免因電腦效能造成螢幕錄影的失敗，所以採用一個自己寫的小程式，記錄實驗者資料夾的建立、資料夾名稱的變動、檔案的搬移，以及進出資料夾的記錄。

個人知識重整系統中會在使用者建立標籤、使用標籤，以及使用關連圖片時，自動記錄下這些歷程。

#### 4.2.2 背景調查問卷

主要詢問實驗者的背景，包括使用網路的習慣，常瀏覽的網站，以及是否有使用過網路上標籤的經驗。問卷詳細內容置放於附錄中。

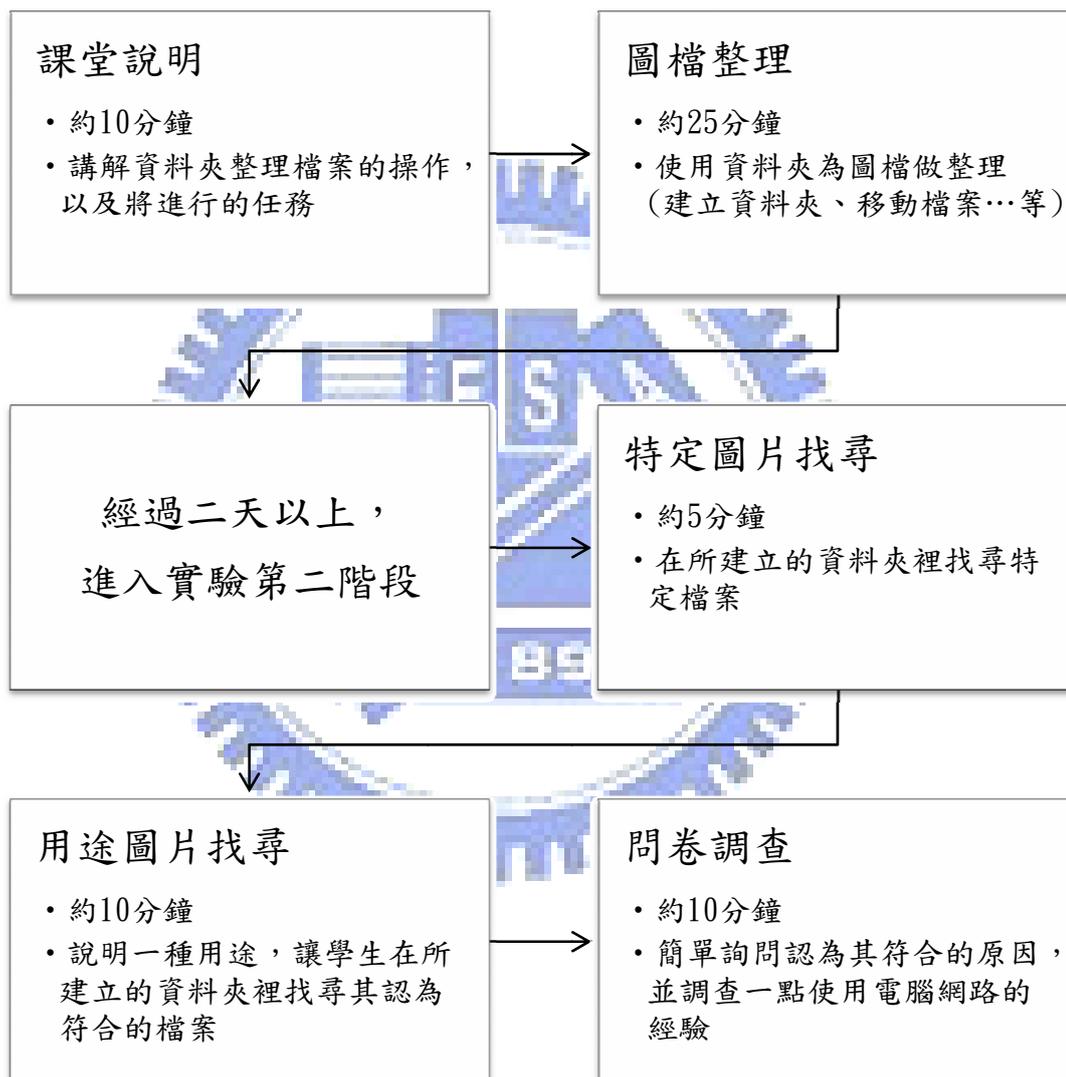
### 4.3 實驗流程

1. 選兩個組成相近的班級，一班使用一種方式來做。
2. 給予實驗者 60 張圖片，圖片內容將平均分佈於各類別。
3. 請實驗者使用作業系統內建的資料夾系統，使用 Windows XP 的縮圖模式，為圖片做分類整理。以程式記錄過程的方式，記下實驗者建立資料夾或搬移檔案的次序。
4. 請實驗者使用個人知識重整系統，為圖片做分類整理。系統內會自動記錄下實驗者

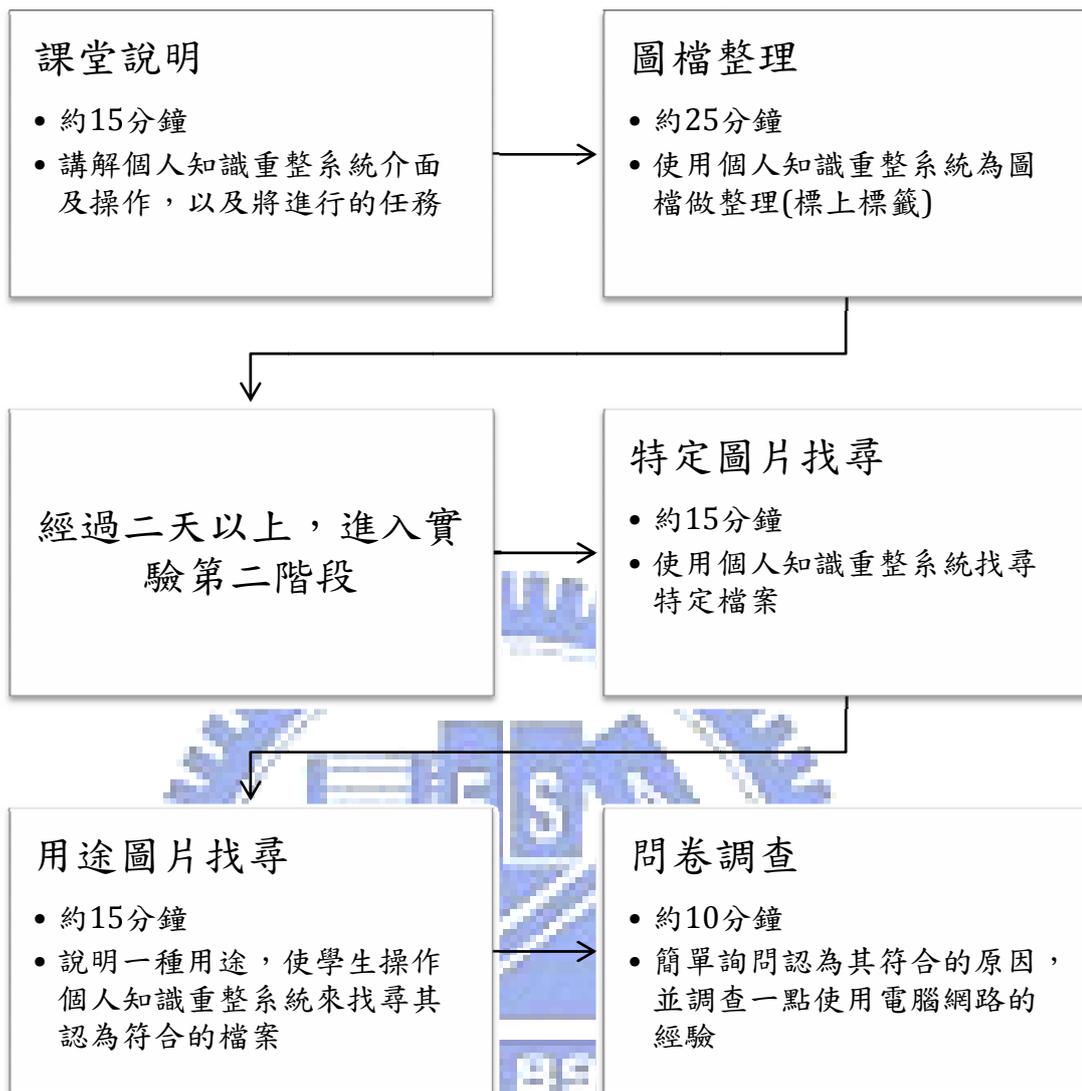
為檔案標上標籤的順序。

5. 請實驗者從建立的資料夾中找出某張特定的圖，不會明確的指出哪張圖，但會給予較模糊的提示。例如：圖的右邊有許多紅色的花。
6. 同上，但是請實驗者用標籤系統來找。
7. 以程式記錄的方式，記下實驗者尋找圖片的過程。

● 資料夾分類組



● 標籤分類組



#### 4.3.1 搜尋任務說明

請使用者從已分類完成的結構當中，尋找指定的圖片。

詳細題目請參考附錄 A。

### 4.4 實驗結果與分析

在這一節中，我們將會分析比較二種方式搜尋到目標檔案的效率、各自結構建立的風格、二種結構的搜尋風格差異，以及個人與整體間是否存在某些認知的共識。一開始將說明系統的使用狀況，做為實驗結果的背景介紹。

#### 4.4.1 系統使用狀況

為了證明系統的各項功能及特性，有被使用者所運用，所以我們將使用下列統計資料來說明。

### 下標籤的情形

由圖 4.4.1 我們可以看到，大部分的使用者對每張圖片會加 3 個左右的標籤，證明有發揮標籤的特性 - 對單一資訊可用多個標籤來描述它。

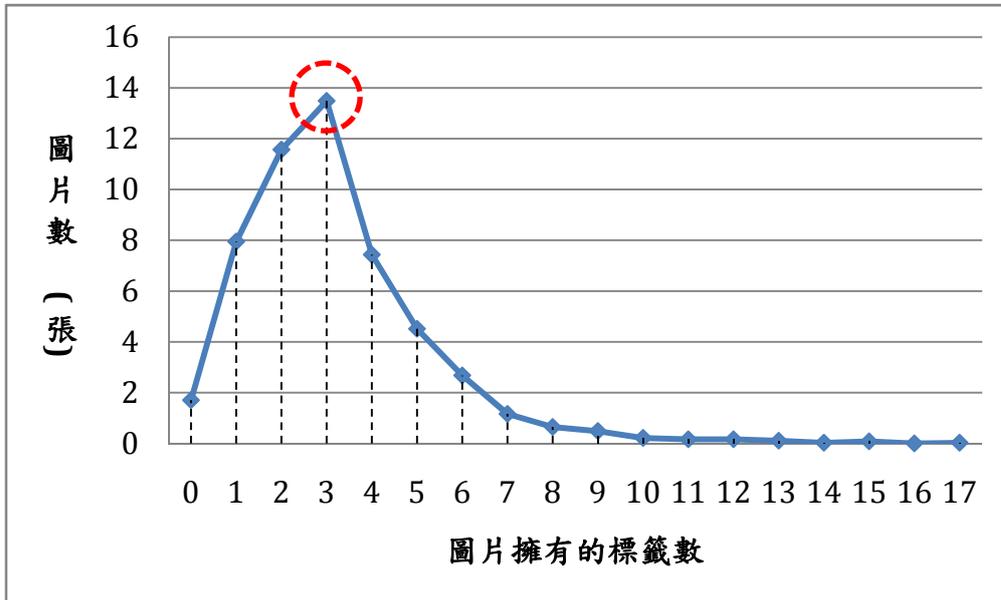


圖 4.4.1 每張圖片平均擁有標籤數的分佈圖，橫軸代表每張圖片擁有的標籤數，縱軸代表擁有特定標籤數的圖片平均有幾張，結果的高峰點顯示一張圖片有 3 個標籤的情形最多

### 功能使用率

- 使用相關圖片的人有 21 人，占標籤組有效樣本中的 65.6%
- 使用相關標籤的有 7 人，占標籤組有效樣本中的 21.8%

### 4.4.2 搜尋效率

針對使用二種方式來找尋目標圖片的時間以及步數，我們可以來做一個比較，證明使用哪種方式可以有效率地找到想要的資訊。

### 使用時間

由圖 4.4.2 我們可以看出使用標籤的系統可以比資料夾系統更快速地找到所想找的圖片，證明個人知識重整系統的確較能幫助人們做尋找。證實藉由類似人腦中陳述性網路的節點激發所設計出來的系統，確實可以幫助人們做有效率的搜尋。

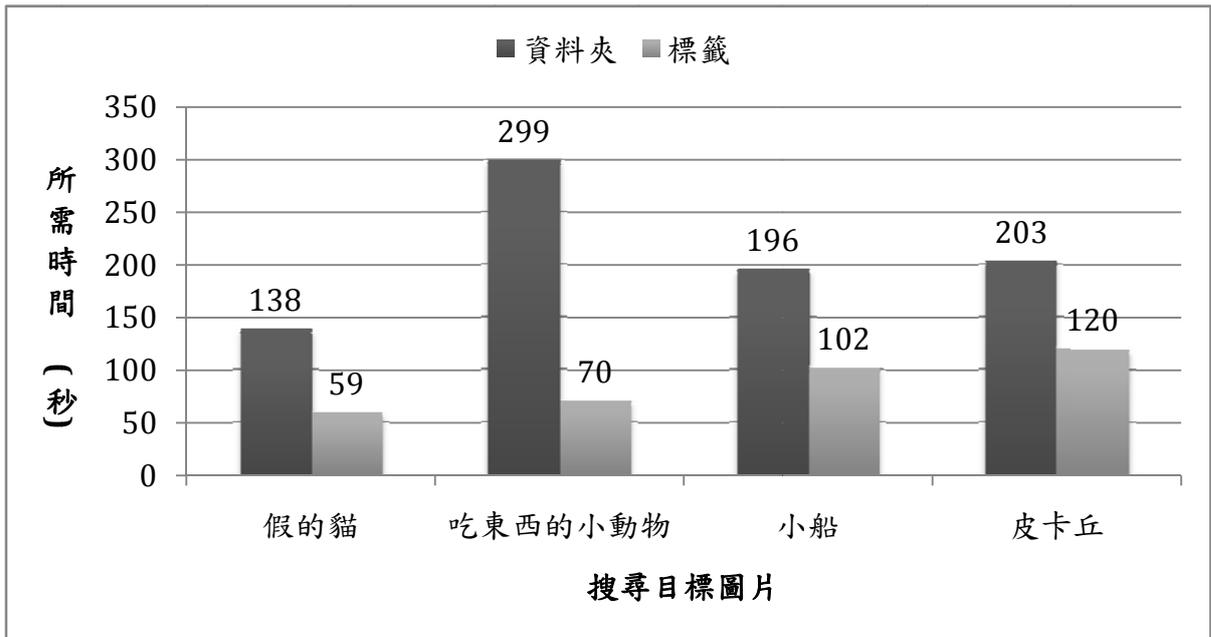


圖 4.4.2 搜尋目標圖片所需的時間圖

將上述圖表轉換成表格，可以看到標籤的所需時間最少可縮減為資料夾的六成或減少更多。

	假的貓	吃東西的小動物	小船	皮卡丘	有效樣本數
資料夾所需時間(s)	138.45	299.00	195.56	203.09	17
標籤所需時間(s)	59.03	70.23	101.73	119.67	32
標籤/資料夾時間比	42.64%	23.49%	52.02%	58.92%	

表 4.4.1 搜尋時間效率比較表

### 使用步數

若將進入一個資料夾算做一步，將使用一個標籤也算做一步，我們可以得到對於每個搜尋任務所需的步數，結果統計於圖 4.4.3。我們可以看到使用標籤系統所需的步數大部分比資料夾來得多，由於標籤可對圖片做較特殊化的描述，所以每個標籤所標記的圖片較少，需使用較多的標籤來找到檔案。但即使步數上較多，在整體的搜尋時間上還是領先資料夾。

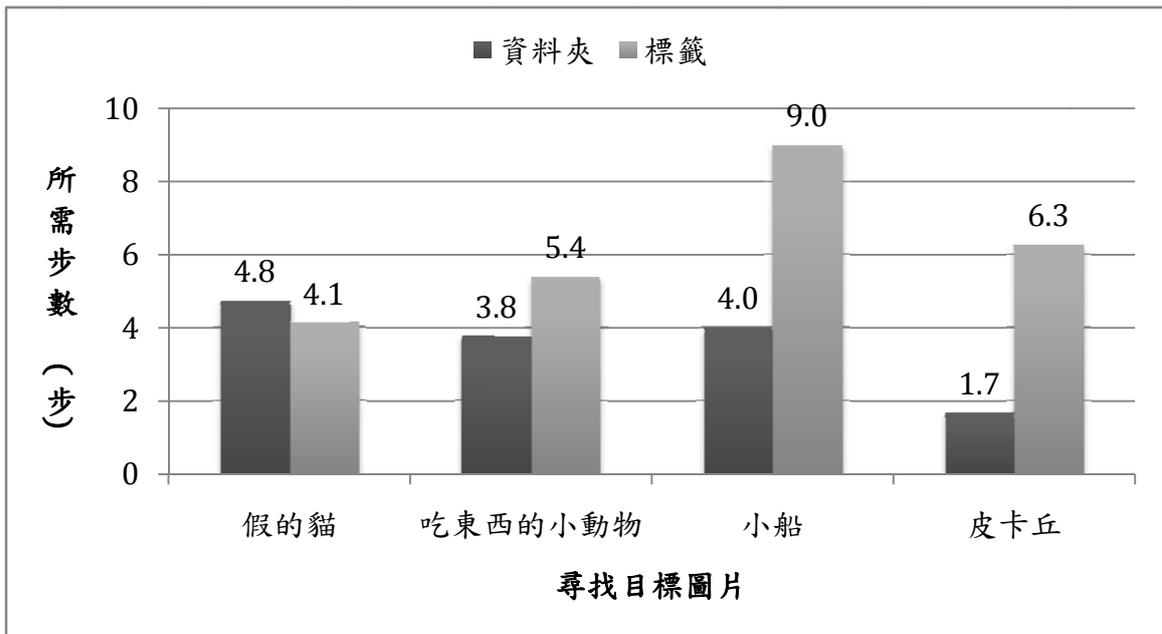


圖 4.4.3 搜尋目標圖片所需的步數圖

將上述圖表轉換成表格，可以看到標籤的所需步數平均而言比資料夾來得多。

	假的貓	吃東西的小動物	小船	皮卡丘	有效樣本數
資料夾所需步數	4.8	3.8	4.0	1.7	17
標籤所需步數	4.1	5.4	9.0	6.3	32
標籤/資料夾步數比	86.93%	143.81%	225.00%	375.56%	

表 4.4.2 搜尋步數比較表

## 4.4.3 結構建立風格

### 4.4.3.1 結構分析

以下就二種方式來觀察其結構：

#### 資料夾結構

大部分的使用者結構，都像下面圖 4.4.4 一樣，深度為 1~2 層的樹狀結構。Test 為實驗時的根目錄，底下為使用者所建立的子目錄。

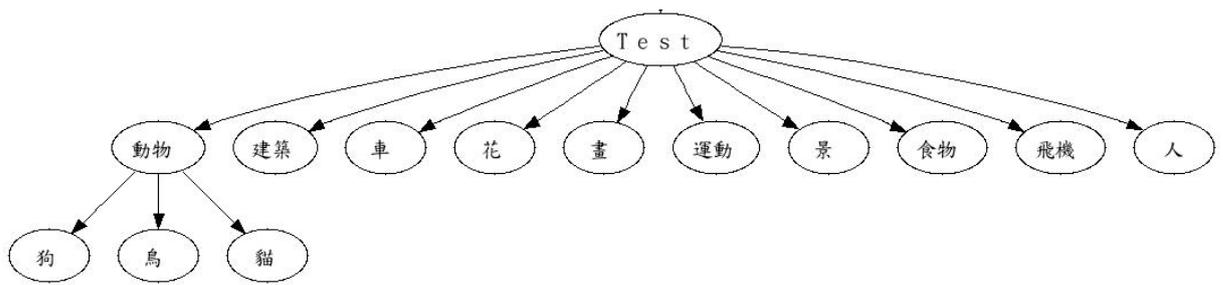
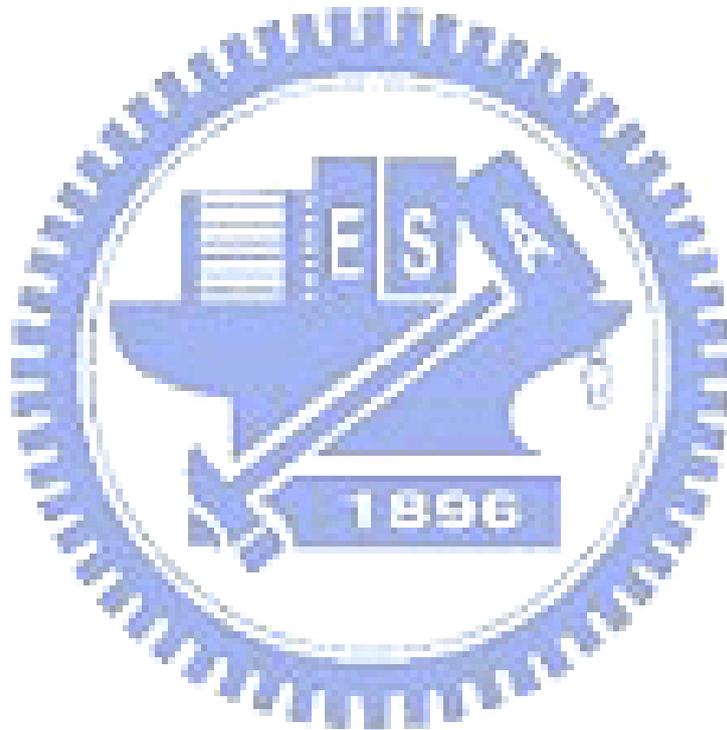


圖 4.4.4 資料夾結構

### 標籤結構

標籤的結構就因個人風格不同，而產生出不小的差異，以下就某幾大類的結構提出說明：



### (1) 連結度高的網路

原本零散的觀念透過某些共通的標籤將各個小網路連起來，從任一節點開始可以到達大部分的節點，呈現大部分相連的狀況，如圖 4.4.5。

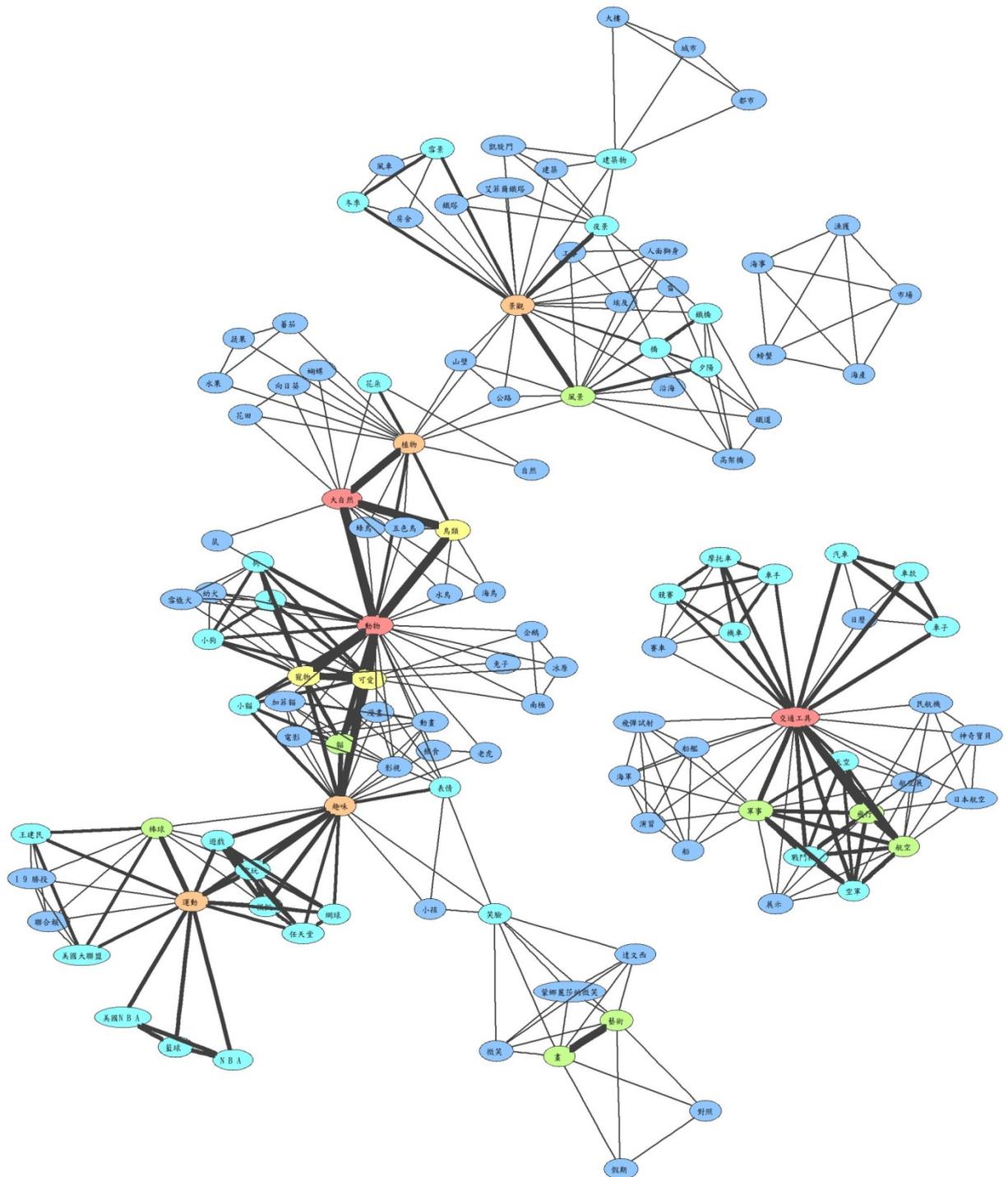


圖 4.4.5 連結度高的標籤網路，此為某一受測者所建立的標籤網路，大部分的節點皆連結在一起，代表對於不同圖片間有相同的標籤，一張圖片含有多個標籤

## (2) 單點散佈的網路

存在著一個一個節點，彼此沒有連結，散佈在網路中。此狀況出現在使用者把標籤運用成類似資料夾的用法時，對每一張圖片只下一個標籤來分類，所以不存在著標籤間的關係，如圖 4.4.6。

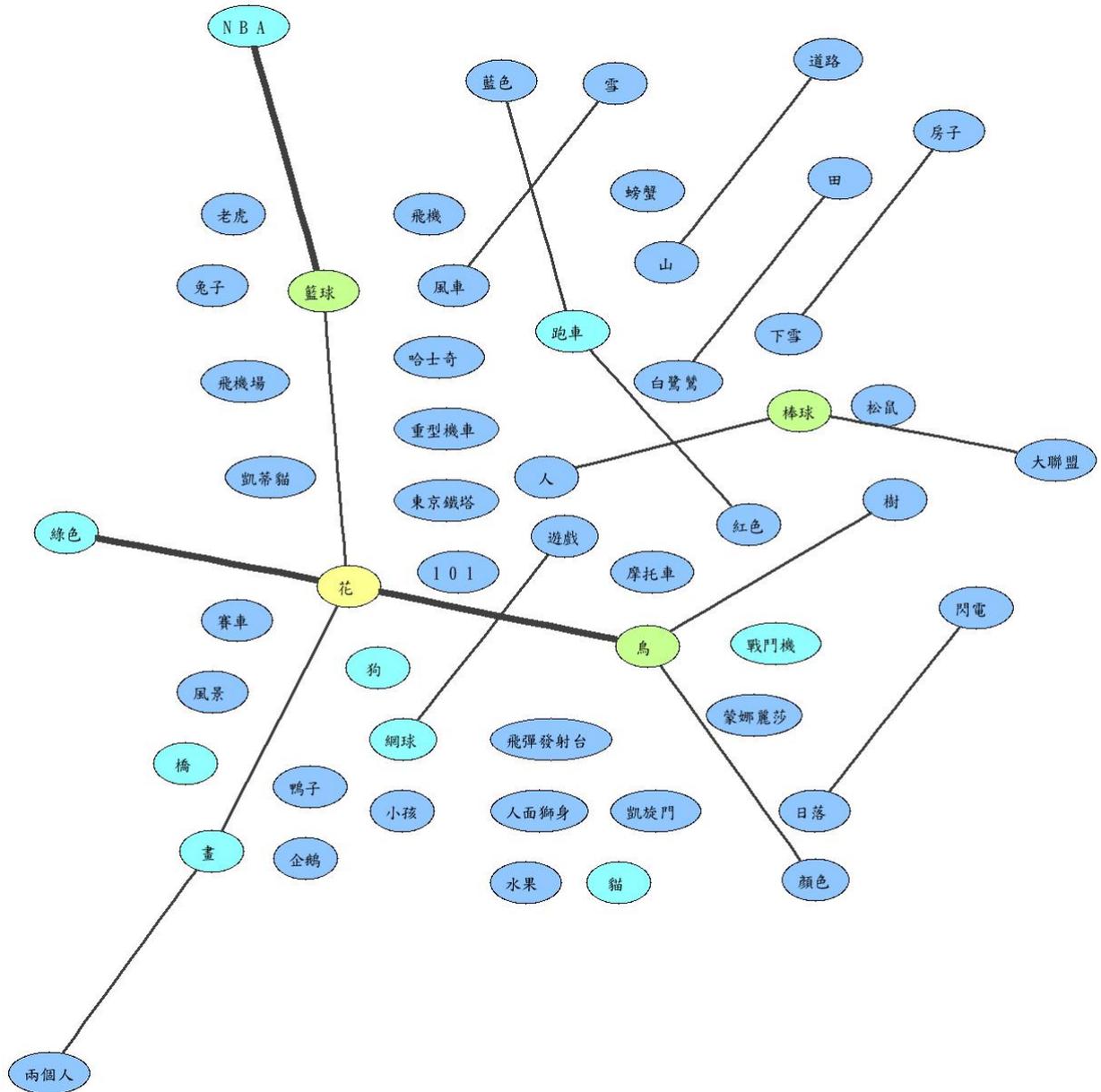


圖 4.4.6 單點散佈的標籤網路，此為某一受測者所建立的標籤網路，大部分的節點之間無相連，代表此受測者大部分圖片只下一個標籤

### (3) 區塊分佈的網路

當使用者對於單張圖片下多個標籤，但是不同圖片間不含有相同的標籤時，就會出現區塊分隔的情況，如圖 4.4.7。

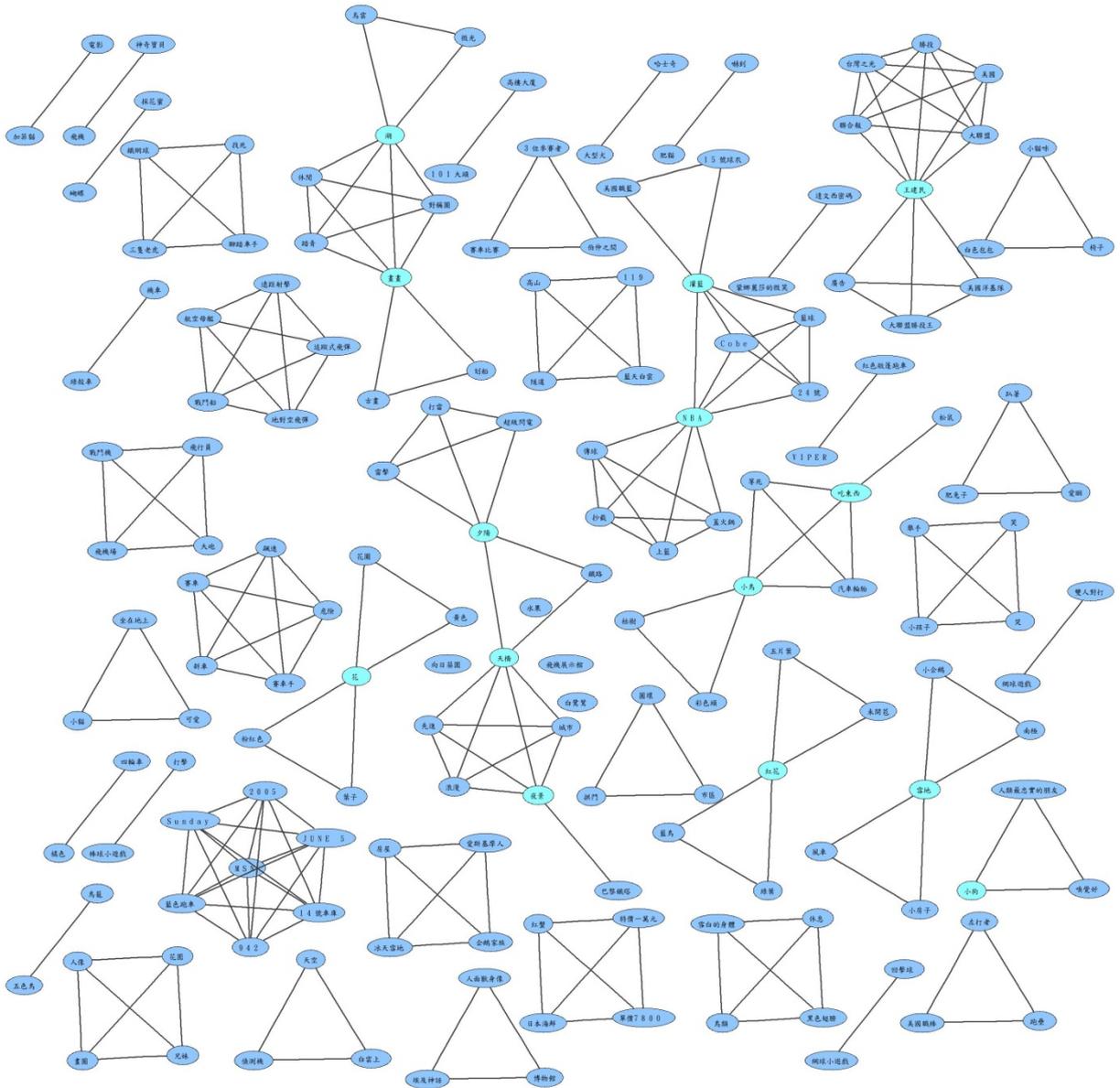


圖 4.4.7 區塊分隔的標籤網路，此為某一受測者所建立的標籤網路，節點相連成一個個的區塊，區塊與區塊間無相連，代表此受測者不同圖片間很少下相同的標籤

#### 4.4.3.2 結構統計資料

以下對於使用者建立結構的統計資料做說明：

#### 資料夾和標籤的使用比率和使用範圍

首先說明圖 4.4.8 的二個座標軸：

- 使用比率—指的是多少百分比的使用者有使用此標籤或資料夾
- 使用範圍—平均使用多少個標籤在檔案(圖片)上面，或是平均一個資料夾包含幾個檔案。

我們可以發現不論資料夾或標籤都存在有高使用比率的例子，像是圖右上角的「動物」資料夾有 88.24% 的使用者有建立，和圖左上角的標籤「風車」有 94.59% 的使用者有使用此標籤，代表二種方法都存在認知的共識或偏好的分類。二種方式的分佈從高到低的使用比率都有。但是很明顯地，幾乎每個標籤平均只會加到 4 個以下的圖片上面，說明標籤描述較精確的部分。大部分每個資料夾都包含 3 張以上的圖片，因為使用者傾向用較一般化的概念去建立資料夾，好包含多項待整理的圖片。

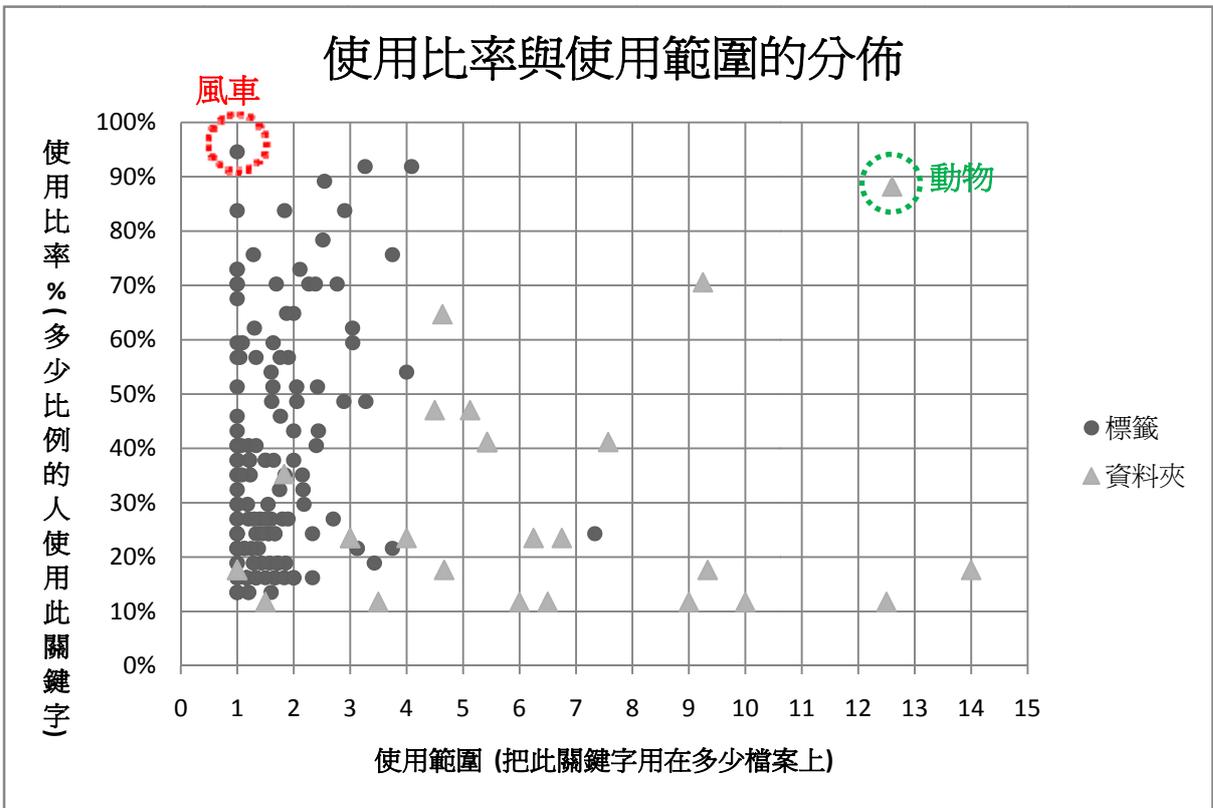


圖 4.4.8 使用比率與使用範圍的分佈圖

### 特定圖片在二種方法上的差異

例如說下圖 4.4.9，我們將分析其資料夾與標籤所使用詞彙的差別。



圖 4.4.9 特定圖片

在資料夾使用的名稱上面，由於一張圖片只能存在於一個資料夾中，所以統計其所存在的資料夾，我們可以畫出圖 4.4.10。大部分的人傾向使用較概觀的詞彙—動物來分類這張圖，有少部分人會直接使用此圖片的主題「企鵝」來建立一個新的資料夾，另外還有少數使用鳥和大自然來分類的。

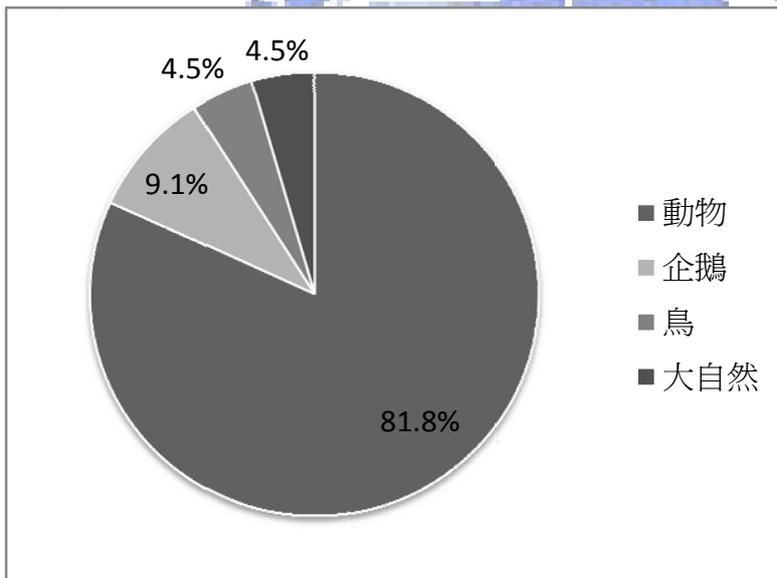


圖 4.4.10 特定圖片的資料夾名稱使用統計

在標籤的名稱上，因為標籤可以對一張圖片下多個標籤，所以我們以長條圖來說明，見圖 4.4.11。我們可以看到標籤「企鵝」遙遙領先其他標籤，使用者傾向使用直接描述圖片主題的方法來為圖片下標籤，反觀剛才在資料夾名稱使用最多的「動物」，反而只有 10.8%使用者會用它，因為它是個比較一般化概括的詞彙，不符合使用者在挑選標籤描述時的直覺描述。

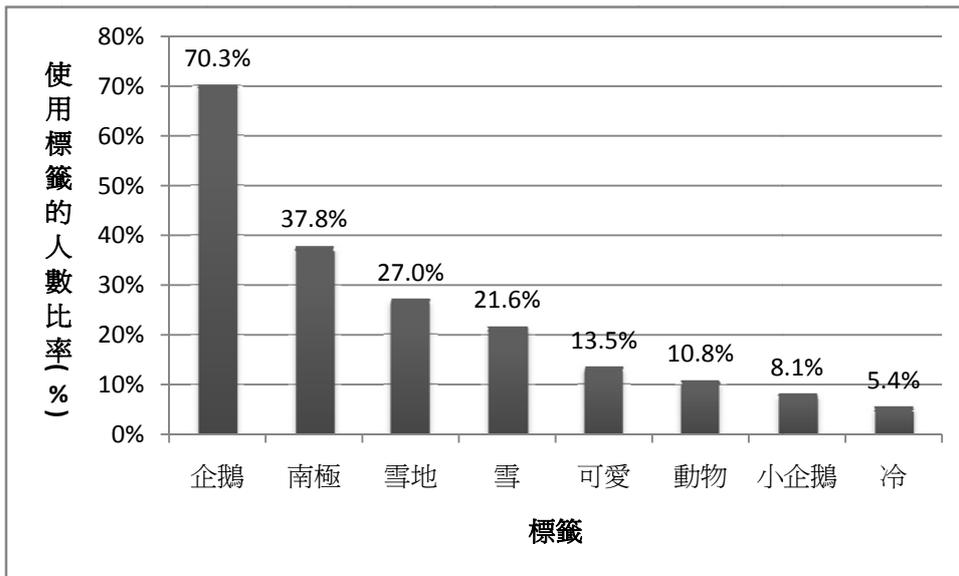


圖 4.4.11 特定圖片的標籤名稱使用統計

#### 4.4.4 搜尋風格

##### 資料夾搜尋

我們舉一個例子來說明，像下圖 4.4.12 是使用者所建立出來的資料夾結構，他在尋找「吃東西的小動物」時，先找了「人」、「車」、「花」、「食物」、「飛機」的資料夾，找不到後又反覆回去找「食物」、「飛機」還是找不到，接著進入「動物」底下的「狗」和「鳥」都找不到後，回到「動物」的資料夾才找到目標的圖片。大部分其他的使用也大致上是類似的過程。

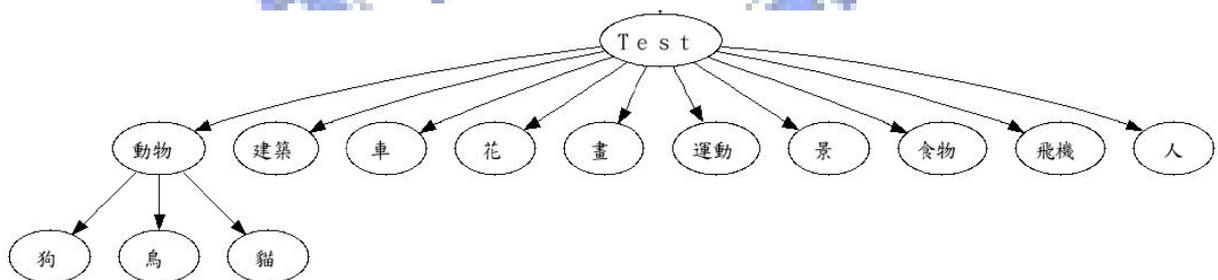


圖 4.4.12 資料夾搜尋範例

##### 標籤搜尋

同樣舉一個例子來說明，尋找的圖片為「吃東西的小動物」，使用者先從「狗」這個標籤來搜尋，在圖 4.4.13 上可以看到此標籤，尋找連結狗這個節點旁的連接線段，也就是含有「狗」這個標籤的圖片，結果找不到要的圖，點選一、二張標籤「狗」的圖片的相關圖片時，同時激發該圖片其他的標籤，進行內容找內容的搜尋，還是找不到。

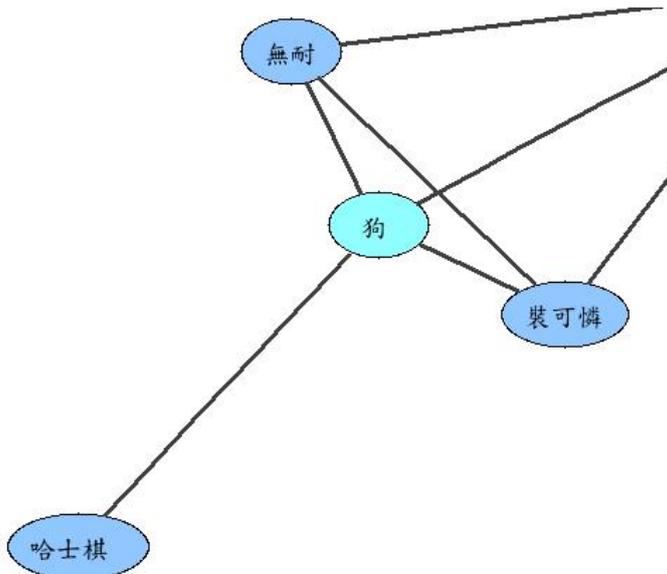


圖 4.4.13 搜尋標籤範例－步驟 1

接著就跳至「吃」的節點，出現的是超級市場中販賣的螃蟹與花叢中的鳥兩張圖。與目標的吃東西小動物都有差距。使用者加入標籤時可能不完整標示出所有可能的屬性，而目標的吃東西小動物圖並沒有標上這個標籤因此不在此搜尋結果之內。使用者接下來點選螃蟹的圖再次進行相關圖片搜尋(具有螃蟹、餐廳、吃與肚子餓四個標籤) 參考圖 4.4.14。此時由於吃東西的小動物與螃蟹具有肚子餓的共通標籤，使用者得以完成從吃擴散到肚子餓的標籤，順利找到目標圖片。

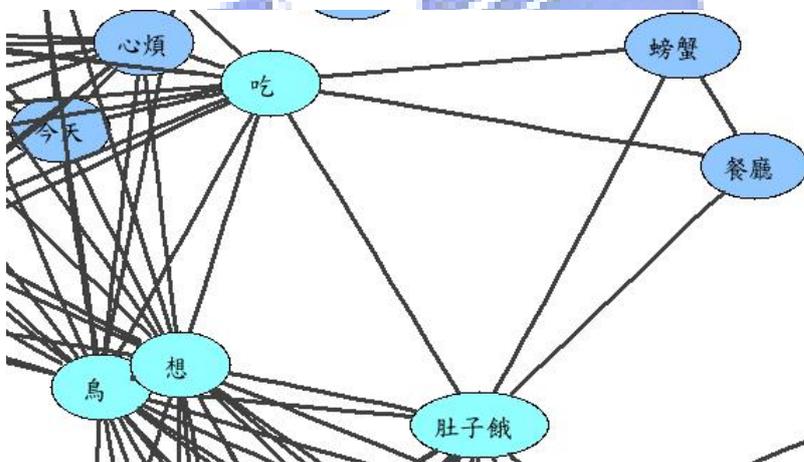


圖 4.4.14 搜尋標籤範例－步驟 2

#### 4.4.5 個人與群體差異

##### 群體共識和個人認知

使用者使用標籤系統時，在沒有參考其他人的答案情況下，對於很具代表性的圖片(如企鵝)在進行加入標籤的動作時所下的標籤也會產生一致的情形。這種情況反

應出對於一張圖片的普遍觀感。我們計算標籤系統受測者的結果，將具有普遍性的標籤關係整理如圖 4.4.15，由圖中可以發現形成一個較小但具有一定結構，並且具有共通性的標籤關連性子集合網路，此種放諸四海皆準的標籤間關係代表了不同使用者間對圖片所共有的常見描述。

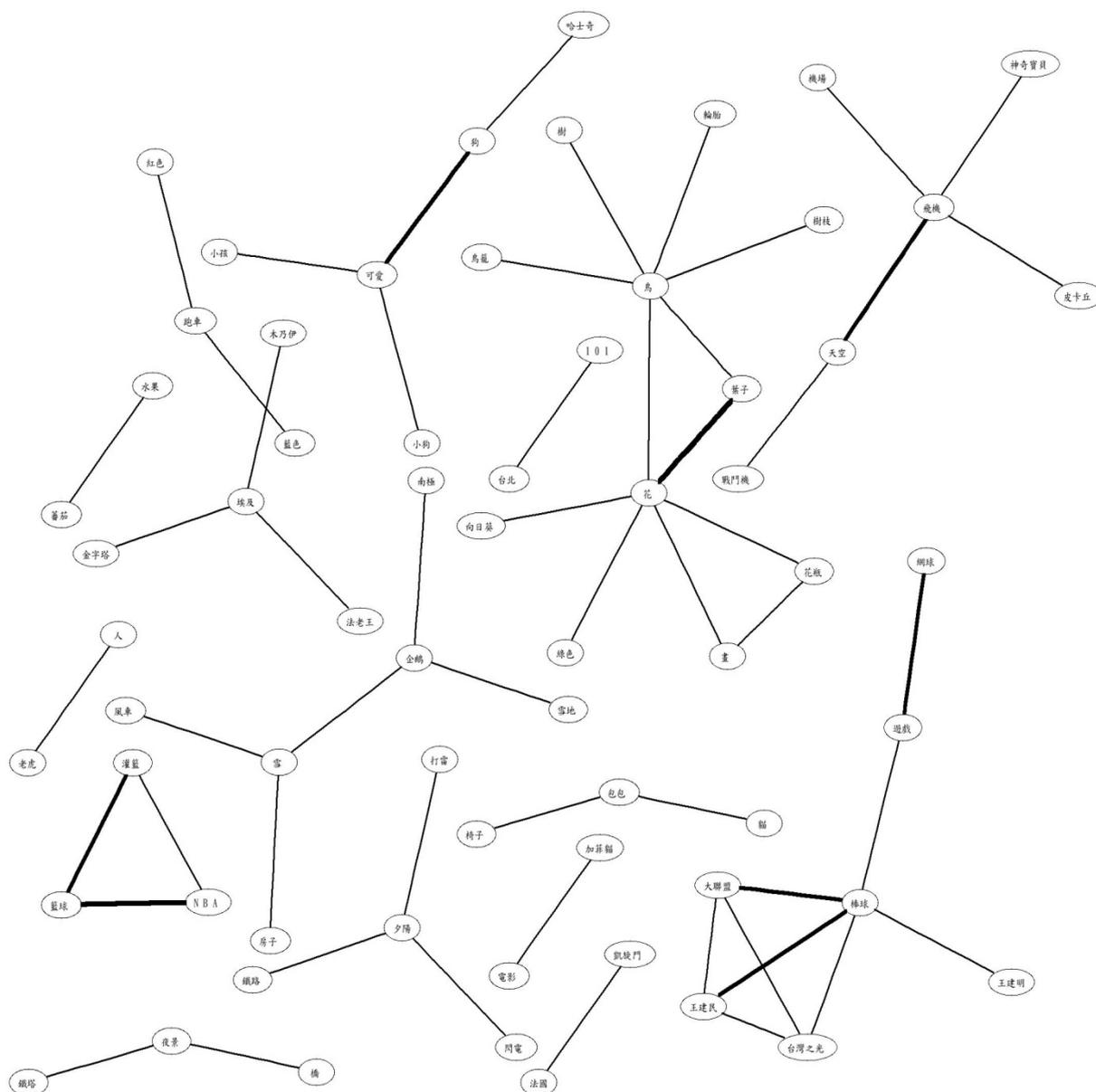


圖 4.4.15 群體共識標籤網路

在形成這樣的共有標籤關係子集合後，我們將個別使用者的個人標籤網路中屬於共通擁有的部分進行標示，將共通節點的邊框與共通邊加深。(如圖 4.4.16 和圖 4.4.17) 可以觀察到在不同類型的網路圖上，都具有相似的共通標籤結構。



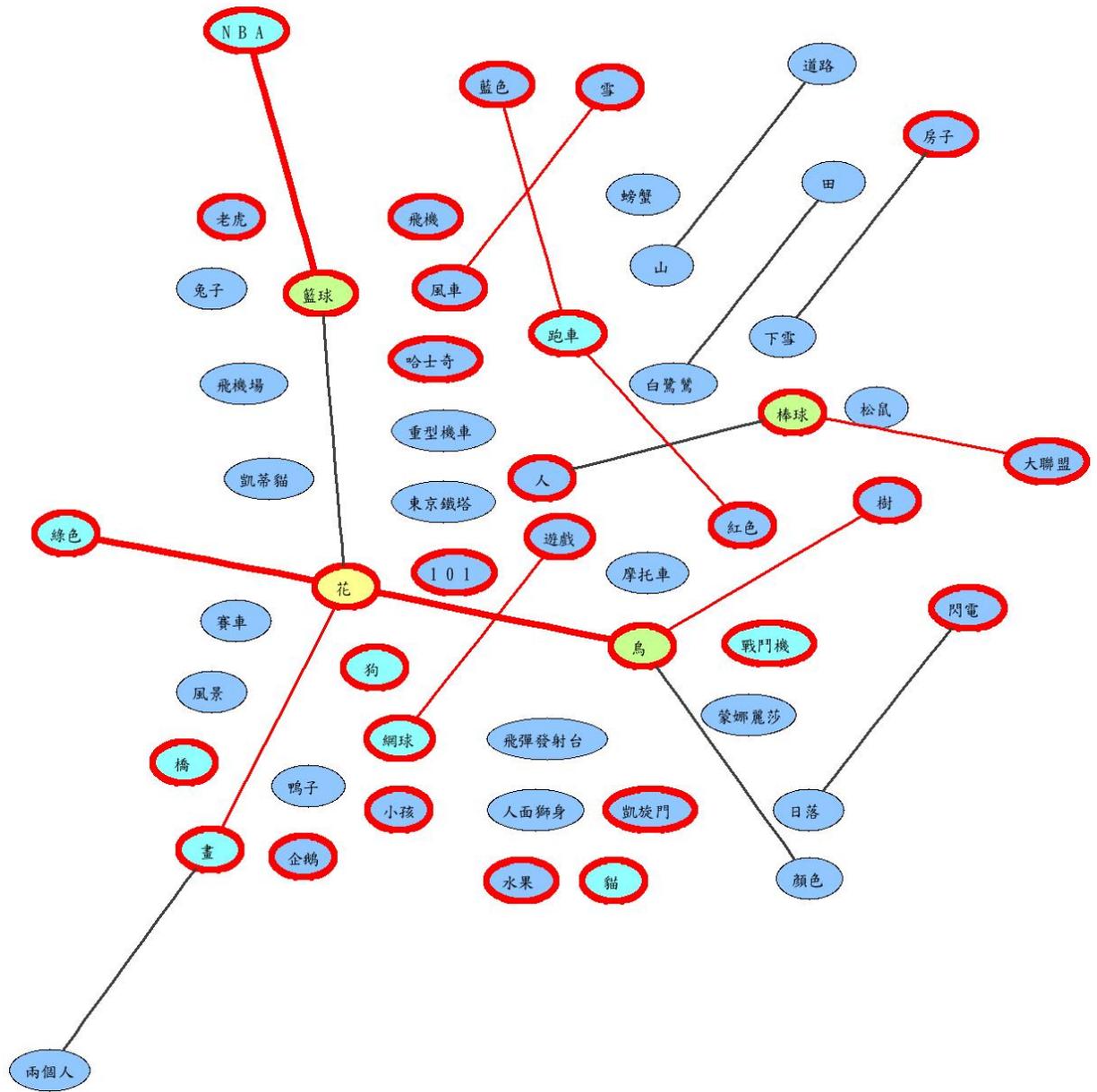


圖 4.4.17 共通標籤在個人標籤網路上的位置(範例 2)

## 五、 結論

在本研究中提供一種理論框架，將認知心理學的記憶相關理論拿來解釋和對應標籤的運作，把二種領域的觀念做一個整合。以記憶相關理論為主要的觀念，實際設計出一個應用的系統，將網際網路上成功的標籤系統搬到個人電腦上來，作為個人知識整理的工具。以教學實驗來確定系統的可使用性，以及比傳統資料夾在搜尋、回憶知識上效能的提升，並討論出標籤網路和資料夾階層的性質與比較。

對於一般電腦使用者而言，一個有效率整理及再度提取檔案的方法，是他們所渴望的，個人知識重整系統可以滿足多面向、個人化的需求，有效地提供檔案線索和描述細節，廣泛地應用到個人知識重整上。另外，對於認知心理學家而言，在進行卡片分類法(card sorting)這類心理學實驗時的平台選擇，以及從人們標標籤的行為來研究了解人類如何對文字圖片進行認知辨識。

本研究所討論的個人標籤網路同時具有應用面與學術面的豐富價值。例如在學術面，透過此系統使用者所建立的對檔案標籤描述，分析其中的標籤所具有的描述對像、主觀客觀差異。可以轉化成類似概念圖(Concept Map)的方式，幫助學者進行具有自動化、背景式的系統化來對使用者行為進行分析。而在應用面，除了能夠透過個人的標籤網路分析個人喜好與特性，進行針對性的網路服務如社群的興趣交流與合作式篩選垃圾郵件等服務外。未來透過網路進行使用者個人標籤網路匯集後，透過廣範圍共識網路的建立，一方面減少使用者在處理共同資訊時的整理常見重覆動作，在另一方面，將這些大量合作產生的常識性(Common Sense)的知識做為提供未來語意網路與 Web3.0 技術發展所需。

## 參考文獻

- [1] F. Wolf, T. Poggio, and P. Sinha, "Human Document Classification Using Bags of Words," 2006.
- [2] H. L. Roediger 3rd, "Memory metaphors in cognitive psychology," *Mem Cognit*, vol. 8, pp. 231-246, 1980.
- [3] R. C. Atkinson and R. M. Shiffrin, "Human memory: A proposed system and its control processes (Vol. 2)," New York: Academic Press, 1968.
- [4] R. C. Atkinson and R. M. Shiffrin, "The control of short-term memory," *Sci Am*, vol. 225, pp. 82-90, 1971.
- [5] F. I. M. Craik and R. S. Lockhart, "Levels of processing: A framework for memory research," *Journal of Verbal Learning and Verbal Behavior*, vol. 11, pp. 671-684, 1972.
- [6] S. C. Brown and F. I. M. Craik, "Encoding and retrieval of information," *The Oxford handbook of memory*, pp. 93-108, 2000.
- [7] F. I. M. Craik and E. Tulving, "Depth of processing and the retention of words in episodic memory," *Journal of Experimental Psychology: General*, vol. 104, pp. 268-294, 1975.
- [8] A. Baddeley, "Working memory.(In) *The Cognitive Neurosciences*," Cambridge, MA: MIT Press, 1995.
- [9] J. Cantor and R. W. Engle, "Working-memory capacity as long-term memory activation: an individual-differences approach," *J Exp Psychol Learn Mem Cogn*, vol. 19, pp. 1101-14, 1993.
- [10] M. Daneman and P. A. Carpenter, "Individual differences in working memory and reading," *Journal of Verbal Learning and Verbal Behavior*, vol. 19, pp. 450-466, 1980.
- [11] R. W. Engle, J. Cantor, and J. J. Carullo, "Individual differences in working memory and comprehension: a test of four hypotheses," *J Exp Psychol Learn Mem Cogn*, vol. 18, pp. 972-92, 1992.
- [12] J. R. Anderson, "The architecture of cognition," Cambridge, MA: Harvard University Press, 1983.

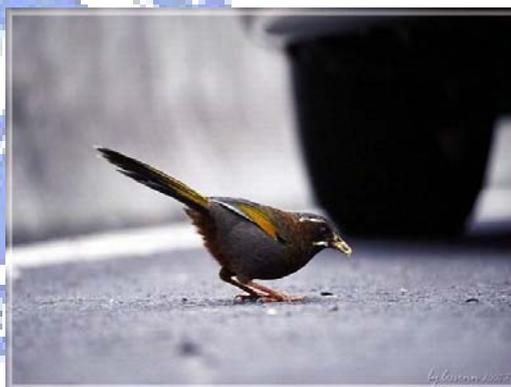
- [13] A. Baddeley, "Human memory: Theory and practice (rev. ed.) Boston: Allyn & Bacon," 1998.
- [14] A. D. Baddeley, "Short-term and working memory," *The Oxford handbook of memory*, pp. 77-92, 2000.
- [15] J. D. Bransford and M. K. Johnson, "Contextual prerequisites for understanding: Some investigations of comprehension and recall," *Journal of Verbal Learning and Verbal Behavior*, vol. 11, pp. 717-726, 1972.
- [16] L. Grossman and M. Eagle, "Synonymity, antonymity, and association in false recognition responses," *Journal of Experimental Psychology*, vol. 83, pp. 244-248, 1970.
- [17] W. A. Bousfield, "The occurrence of clustering in the recall of randomly arranged associates," *Journal of General Psychology*, vol. 49, pp. 229-240, 1953.
- [18] E. Tulving and Z. Pearlstone, "Availability versus accessibility of information in memory for words," *Journal of Verbal Learning and Verbal Behavior*, vol. 5, pp. 381-391, 1966.
- [19] G. H. Bower, M. C. Clark, A. M. Lesgold, and D. Winzenz, "Hierarchical retrieval schemes in recall of categorized word lists," *Journal of Verbal Learning and Verbal Behavior*, vol. 8, pp. 323-343, 1969.
- [20] J. Nielsen, "Ten Usability Heuristics." vol. 1, 2001, p. 2001.
- [21] E. Tulving and D. M. Thomson, "Encoding specificity and retrieval processes in episodic memory," *Psychological Review*, vol. 80, pp. 352-373, 1973.
- [22] M. J. Watkins and E. Tulving, "Episodic memory: When recognition fails," *Journal of Experimental Psychology: General*, vol. 104, pp. 5-29, 1975.
- [23] A. G. Greenwald and M. R. Banaji, "The self as a memory system: Powerful, but ordinary," *Journal of Personality and Social Psychology*, vol. 57, pp. 41-54, 1989.
- [24] T. Mantyla, "Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 12, pp. 66-71, 1986.

## 附錄 A. 實驗中搜尋任務的詳細說明

1. 請在你整理完成的圖片中，找到一張「假的貓」的圖。



2. 請在你整理完成的圖片中，找到一張「吃東西的小動物」的圖。



或

3. 請在你整理完成的圖片中，找到一張有「小船」的圖。



或

4. 請在你整理完成的圖片中，找到一張有「皮卡丘」的圖。



## 附錄 B. 系統介面的操作方法

### 建立標籤的方法

#### A. 對單一檔案編輯標籤

1. 假設想為下圖加上標籤：鍵盤, USB

[1 \(3\).jpg](#)

標籤:

確認



相關圖片

2. 輸入標籤：鍵盤, USB，每個標籤中間以逗號「,」做為分隔

[1 \(3\).jpg](#)

標籤:

鍵盤,USB

確認



相關圖片

3. 按下「確認」後即可，下面為標籤編輯完成的樣子

[1 \(3\).jpg](#)

標籤: [鍵盤](#) [USB](#)

鍵盤,USB,

確認



相關圖片

#### B. 對多個檔案加上標籤

4. 對想加上標籤的檔案，用滑鼠單點，檔案的背景會變為綠色。重覆此動作，直到點完所要的檔案為止。

1 (1).jpg  
標籤: [光碟機](#)

光碟機,

 相關圖片

1 (2).jpg  
標籤: [光碟機](#)

光碟機,

 相關圖片

1 (3).jpg  
標籤: [USB](#)

USB,

 相關圖片

1.jpg  
標籤: [主機](#) [鍵盤](#)



將圖片加上此標籤(關鍵字)

- ◆ [光碟機](#) [新增](#)
- ◆ [主機](#) [新增](#)
- ◆ [鍵盤](#) [新增](#)
- ◆ [USB](#) [新增](#)
- ◆ [全選](#)

5. 在右方邊欄中的標籤群裡，找到所想加的標籤之後，點選右方的新增連結。

- ◆ [鍵盤](#) [新增](#)
- ◆ [USB](#) [新增](#)

6. 標籤就會加到所選取檔案上。

1 (1).jpg  
標籤: [鍵盤](#) [光碟機](#)

鍵盤,光碟機,

 相關圖片

1 (2).jpg  
標籤: [光碟機](#)

光碟機,

 相關圖片

1 (3).jpg  
標籤: [鍵盤](#) [USB](#)

鍵盤,USB,

 相關圖片

1.jpg  
標籤: [主機](#) [鍵盤](#)



將圖片加上此標籤(關鍵字)

- ◆ [鍵盤](#) [新增](#)
- ◆ [光碟機](#) [新增](#)
- ◆ [主機](#) [新增](#)
- ◆ [USB](#) [新增](#)
- ◆ [全選](#)

## 搜尋標籤

1. 在頁面上方找到搜尋的方塊

2. 輸入想搜尋的標籤名，中間以空白做分隔。

3. 按下「標籤搜尋」後，頁面出現所有標籤交集的結果。

目前標籤: [鍵盤](#) [光碟機](#)

小提示: 按下目前標籤 做退一步搜尋

相關標籤:

小提示: 按下相關標籤 做進一步搜尋

1張 頁數

[1 \(1\).jpg](#)

標籤: [鍵盤](#) [光碟機](#)



[相關圖片](#)

鍵盤,光碟機,

