

國立交通大學

資訊管理研究所

碩士論文

於數位典藏建立社會性標記之研究

Social Tagging Construction on Digital Archives

研究生：鄧有盈

指導教授：柯皓仁 博士

中華民國九十七年七月

於數位典藏建立社會性標記之研究

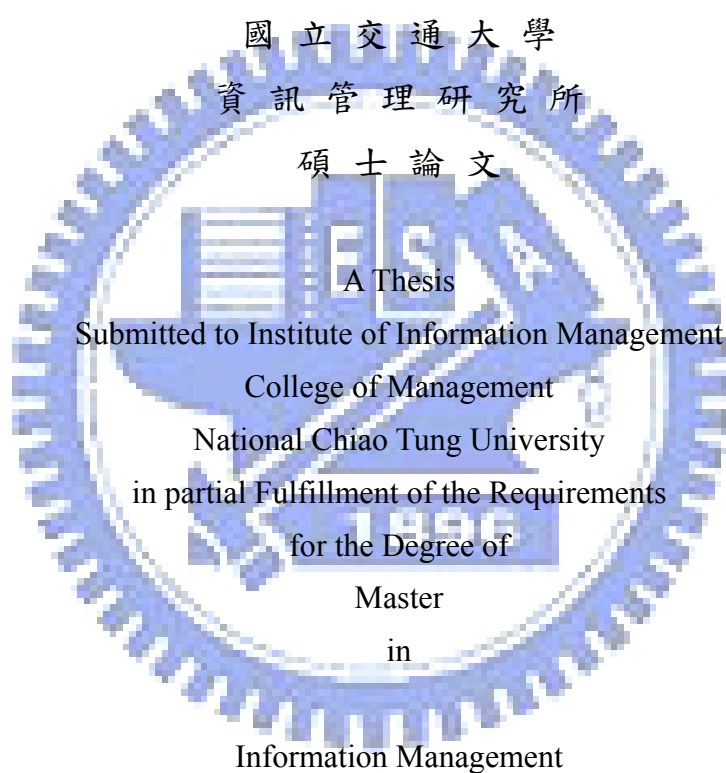
Social Tagging Construction on Digital Archives

研究生：鄧有盈

Student：Yu-Ying Teng

指導教授：柯皓仁

Advisor：Hao-Ren Ke



July 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

於數位典藏建立社會性標記之研究

Social Tagging Construction on Digital Archives

研究生：鄧有盈 指導教授：柯皓仁博士

國立交通大學資訊管理研究所

【摘要】

隨著 Web 2.0 概念及應用的普及，衍生出許多運用 Web 2.0 精神的創作平台，其中社會性標記是近來網路上最熱門的服務之一。社會性標記指的是由使用者針對網站服務內容(如照片、網站連結、藝術品)提出標籤或註解，從而創造出由眾多使用者自行產生的分類，可稱之為大眾分類法。

線上美術館或博物館的優點是讓人們可隨時隨地從網路上觀賞藝術作品，將社會性標記應用於數位典藏網站的目的在於將人們觀賞藝術品的觀感串連起來，希望能透過大眾集思廣益讓每件藝術品的描述能更完整、更多面向，並建立一般人甚至於專家都看不見的巧妙關聯性。

本研究是將 Web 2.0 的精神應用於楊英風數位美術館，透過 CKIP 中文斷詞系統處理每一個藝術作品的內容，然後依據權重萃取其關鍵詞彙，代表此藝術品專業的說明；再透過標記活動，讓一般使用者自由地對楊英風美術作品提供標籤；最後系統會結合作品關鍵詞和使用者加註的標籤進行階層式標籤分群，自動產生每一類作品的標籤地圖，此標籤地圖就是經由大家共同建立的分類，達到大眾分類的目的。

關鍵字：Web 2.0、社會性標記、標籤、大眾分類、分群法、楊英風

Social Tagging Construction on Digital Archives

Student: Yu-Ying Teng Advisor: Dr. Hao-Ren Ke

Institute of Information Management
National Chiao Tung University

【ABSTRACT】

With the widespread of the concept and applications of Web 2.0, a lot of online platforms are created based on the spirit of Web 2.0. One kind of applications is social tagging. Social tagging allows users to provide labels or notes for Web content such as photographs, web links, artwork, and others. Through such behavior, classification of the tagged contents is achieved together by many people. Classification by social tagging is called “folksomony”.

Digital archives let people appreciate artworks from the Internet without the constraints of time and space. Therefore, the purpose of applying social tagging to digital archives is to provide the visitors a tool to express their personal feelings and opinions about the artworks, so that comprehensive comments and descriptions of the artworks will be available to everyone. Hopefully in doing so, latent relationships unnoticed by the experts can be revealed.

This research is to apply the spirit of Web 2.0 to Yuyu Yang Digital Art Museum. First of all, it uses CKIP Chinese Segmentation System to process the metadata of every artwork and extract the keywords according to their weights. The extracted keywords represent professional descriptions about the artwork. Second, public users can tag artworks freely with any vocabulary. Finally, the tagging system will combine the extracted keywords and the tags from the users to form hierarchical tag clusters and generate a tag map for every category of artwork. This tag map reflects the classification by the public users and achieves the goal of folksomony.

Keyword: Web 2.0, Social Tagging, Tags, Folksonomy, Clustering.

誌 謝

在來交大念書之前，我曾經問我自己會不會後悔，捨棄美好的工作，只為了進一步求得知識學位，如今兩年過去了，當初猶豫的決定，在今日卻認為來交大是我這一生中最幸運的事情，尤其完成了學業，並達成了人生的夢想。

這兩年裡，我真的要感謝很多人給予我的幫助，首先感謝恩師柯皓仁老師給予我的指導，他不辭辛勞地修改我的論文寫作及研究過程中觀念架構的釐清，實在是一位難得且認真負責的好老師。其次，要感謝論文口試委員陳光華、謝建成兩位老師的細心的指導，使我的論文更加完整，學生在此對幾位老師獻上由衷感謝。

感謝研究室及所上的同學們，看到你們就像看到過去自己純真洋溢的一面，讓我又再度充滿活力，對人生充滿朝氣，如同剛畢業的莘莘學子對生命滿懷憧憬，感謝學妹在短時間內接下我手中的案子，使我不需分心，得以專注於我的研究論文裡。另外，要感謝母親給我的支持與鼓勵，讓我這兩年完全不用擔心經濟上的壓力，可以毫無後顧之憂繼續完成我的學業，感謝弟弟及眾多親朋好友死黨們，給予我論文無限的支援與協助，讓我順利完成我的論文。

兩年的碩士生涯想不到一下子就過去，在這期間我生活周遭的每一個人都將在我人生佔有一席之地，師長、朋友、同學缺一不可，畢業在即，多少都有些依依不捨，即使將要回到工作崗位上，心裡仍冀望不知未來何時才能再度享受這美麗的校園生活。

有盈 謹誌於交通大學資管所

2008.7

目 錄

【摘要】	I
【ABSTRACT】	II
誌 謝	III
目 錄	IV
表 目 錄	VI
圖 目 錄	VII
壹、緒論	- 1 -
一、研究背景與動機	- 1 -
二、研究方法與目的	- 4 -
三、論文架構	- 4 -
貳、文獻探討	- 5 -
一、WEB 2.0 相關研究	- 5 -
1. Web 2.0 起源及意義	- 5 -
2. Web 2.0 主要原則	- 9 -
3. Web 2.0 的應用類型	- 13 -
二、社會性標記 (SOCIAL TAGGING)	- 16 -
1. 標籤與標記的起源	- 17 -
2. 大眾分類法 (Folksonomy)	- 18 -
3. 大眾分類的優缺點	- 20 -
4. 社會性標記網站的應用	- 22 -
5. 社會性標記與知識管理	- 26 -
6. 標籤之相關研究	- 28 -
三、中文斷詞的相關研究	- 30 -
1. 中文斷詞簡介	- 30 -
2. 中文斷詞的困境及方法	- 31 -
3. 未知詞擷取之研究探討	- 32 -
四、資訊檢索	- 33 -
1. 關鍵字擷取	- 33 -
2. 空間向量模型	- 34 -
五、文件及關鍵字分群	- 35 -
1. 分割法(Partition Method)	- 35 -
2. 密度基礎法(Density-based Method)	- 36 -
3. 格子基礎法(Grid-based Method)	- 36 -

4. 階層分群法(Hierarchical Method).....	- 37 -
參、系統架構與實作	- 38 -
一、核心概念.....	- 38 -
二、系統架構.....	- 40 -
三、資料庫設計.....	- 41 -
四、中文斷詞處理.....	- 43 -
五、線上標籤蒐集.....	- 47 -
六、階層標籤產生器 (TAG HIERARCHY GENERATOR)	- 49 -
七、產出結果.....	- 56 -
八、系統操作畫面.....	- 58 -
肆、問卷評估與結果分析.....	- 62 -
一、問卷設計.....	- 62 -
二、評估項目.....	- 63 -
三、受測者基本資料與網路使用行為分析.....	- 64 -
四、社會性標記應用之需求分析.....	- 67 -
五、關鍵詞、標籤、推薦詞內容分析.....	- 68 -
六、社會性標記功能評估.....	- 70 -
七、社會性標記實作滿意度分析.....	- 72 -
八、綜合討論.....	- 73 -
伍、結論與未來展望	- 76 -
一、結論.....	- 76 -
二、未來發展方向.....	- 77 -
參考文獻	- 79 -
附錄一、中研院平衡語料庫詞類標記集	- 83 -

表 目 錄

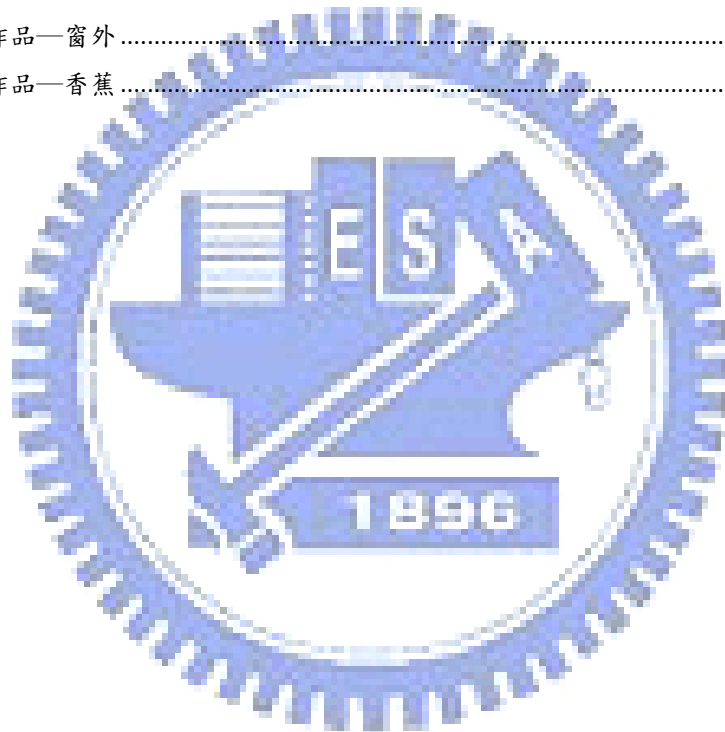
表 1：WEB 2.0 vs WEB 1.0 (整理自TIM O'REILLY "WHAT IS WEB 2.0")[25].....	- 7 -
表 2：楊英風數位美術館開發軟體需求.....	- 41 -
表 3：美術作品擷取資料欄位表 (資料來源： HTTP://YUYUYANG.E-LIB.NCTU.EDU.TW/).....	- 44 -
表 4：CKIP用戶端傳送之XML格式[46].....	- 44 -
表 5：CKIP伺服器傳回XML格式的處理結果[46].....	- 45 -
表 6：CKIP回傳另外三種XML格式的處理結果[46].....	- 45 -
表 7：中文斷詞詞類擷取列表(整理自附錄一).....	- 46 -
表 8：關鍵詞標籤合併範例表.....	- 49 -
表 9：標籤向量表.....	- 50 -
表 10：各標籤被標記的資源數量與總標記次數.....	- 51 -
表 11：計算鋼筆與速寫間的相互關係.....	- 52 -
表 12：計算水墨與速寫間的相互關係.....	- 53 -
表 13：計算鉛筆與速寫間的相互關係.....	- 54 -
表 14：計算水墨與鋼筆間的相互關係.....	- 54 -
表 15：計算鉛筆與鋼筆間的相互關係.....	- 55 -
表 16：計算鉛筆與水墨間的相互關係.....	- 55 -



圖目錄

圖 1：史帝夫標籤計畫 (HTTP://TAGGER.STEVE.MUSEUM).....	- 3 -
圖 2：WEB 2.0 VS WEB 1.0 (資料來源：WEB 2.0: EVOLUTION TOWARDS A READ/WRITE PLATFORM) [7]	- 7 -
圖 3：WEB 2.0 MEME MAP (資料來源: TIM O'REILLY "WHAT IS WEB 2.0")[25].....	- 8 -
圖 4：WEB 2.0 八個核心模式 (WEB 2.0 DEFINITION UPDATED AND ENTERPRISE 2.0 EMERGES)[12]	- 9 -
圖 5：LONG TAIL DIAGRAM [36]	- 13 -
圖 6：URMAP (HTTP://WWW.URMAP.COM/)	- 14 -
圖 7：無名小站 (HTTP://WWW.WRETCH.CC/).....	- 15 -
圖 8：黑米共享書籤 (HTTP://WWW.HEMIDEMI.COM/)	- 16 -
圖 9：社會性標記之概念架構[44]	- 20 -
圖 10：DEL.ICIO.US (HTTP://DEL.ICIO.US)	- 22 -
圖 11：DEL.ICIO.US 架構圖[11]	- 23 -
圖 12：PENNTAGS 讀者分享註解與標籤的範例 (HTTP://TAGS.LIBRARY.UPENN.EDU/).....	- 24 -
圖 13：STEVE.MUSEUM (HTTP://TAGGER.STEVE.MUSEUM)	- 25 -
圖 14：STEVE.MUSEUM 資料模組架構[6]	- 26 -
圖 15：WIIG 知識管理流程[39]	- 26 -
圖 16：本研究之核心概念	- 39 -
圖 17：本研究之系統架構	- 41 -
圖 18：資料庫設計圖	- 43 -
圖 19：中文斷詞處理步驟	- 43 -
圖 20：楊英風網站標記資源範例圖	- 48 -
圖 21：楊英風網站搜尋資源範例圖	- 48 -
圖 22：本研究之關鍵詞擷取流程	- 48 -
圖 23：標籤與數位資源的關係圖	- 50 -
圖 24：第一輪階層分群過程 1	- 53 -
圖 25：第一輪階層分群過程 2	- 53 -
圖 26：第一輪階層式標籤分群的結果	- 54 -
圖 27：第二輪階層式標籤分群的結果	- 55 -
圖 28： Δ 為 0.5 時的階層式標籤分群結構	- 56 -
圖 29： Δ 為 0.8 時的階層式標籤分群結構	- 56 -
圖 30：繪畫類標籤地圖部分擷取畫面	- 57 -
圖 31：楊英風美術作品的推薦詞	- 58 -
圖 32：楊英風美術作品中的關鍵詞、標籤、推薦詞資訊.....	- 59 -
圖 33：楊英風數位美術館標籤搜尋畫面	- 60 -
圖 34：浮雕類標籤地圖部分擷取畫面	- 61 -
圖 35：受測者系所分佈圖	- 65 -
圖 36：受測者年齡分佈圖	- 65 -

圖 37：受測者學歷分佈圖	- 66 -
圖 38：接觸網路的資歷分佈圖	- 66 -
圖 39：每日使用網路的時間分佈圖	- 67 -
圖 40：使用過的社會性標記系統分佈圖	- 67 -
圖 41：受測者對於社會性標記應用之需求分析	- 68 -
圖 42：景觀雕塑類別作品分析	- 69 -
圖 43：繪畫類別作品分析	- 70 -
圖 44：標記功能評估分析	- 71 -
圖 45：標籤地圖功能評估分析	- 72 -
圖 46：受測者評估楊英風網站之標記功能是否具有實用性分佈圖.....	- 73 -
圖 47：受測者評估楊英風網站之標記功能是否具有創意分佈圖.....	- 73 -
圖 48：靜物與香蕉的階層式結構	- 74 -
圖 49：繪畫類作品—窗外	- 75 -
圖 50：繪畫類作品—香蕉	- 75 -



壹、 緒論

一、 研究背景與動機

近年來Web 2.0 逐漸流行，其核心概念強調眾人的集體智慧，使用者可透過網路平台創建各種內容，使用者間也能透過網路彼此互相交流知識。Web 2.0 打破了個人的界線，讓網路上的知識互相連結，形成一個巨大的架構，就好比大腦內的突觸網路型態，藉由密集重複的連結，不斷成長下，讓網路平台變得更緊密更強大[57]。

許多應用Web 2.0 概念的網站，例如：Del.icio.us、Flickr、Youtube等，提供了讓使用者自由地對網站內容定義關鍵字的功能，這些關鍵字通常稱為標籤(Tag)。例如在Flickr上貼標籤的對象是照片，在Del.icio.us上就是一個網路連結(Web Link)[56]。

社會性標記(Social Tagging)是一種運用集體力量收藏和分享標籤的機制，因為收藏的標籤可以讓許多人在網路上分享，因此有人將標籤稱為社會性書籤(Social Bookmark)或網路書籤(Bookmark)。使用者可以利用標籤來收集、分享感興趣的資訊，如新聞、圖片、資料、網站等。同時，也能方便地與其他人分享自己個人的收藏。而社會性標記指的就是由使用者針對服務內容(如照片、網站連結、藝術作品)提供標籤，從而創造出一種極具創意的自行分類，可稱之為大眾分類法(Folksonomy)。這種分類與古典分類法(Taxonomy)形成對比，並非毫無彈性的分類方式，而是讓使用者透過自然的思維來運行。

之所以運用社會性標記於線上美術館或博物館的想法原因如下：藝術品或美術作品的創作少則幾十年甚至幾百年的歷史，當走進美術館時，可以很輕易地看到畫作上的名稱及作者，也可以知道是甚麼時代的作品以及畫作的特色風格，但是從線上美術館或博物館尋找藝術或美術作品時，除非知道畫的名稱及作者名字否則根本無法找到想看的畫。

「美術館已經發覺線上收藏品並沒有發揮應有的功能，現在的作法根本是把

內容藏起來，讓專家以外的人無所適從」，多倫多檔案與博物館資訊公司(Archives and Museum Informatics)的合夥人之一珍妮佛川特(Jennifer Trant)這麼說[55]；另外，丹佛美術館新科技部主任布魯斯韋曼(Bruce Wyman)說：「讓大眾下標籤，給了我們一雙從沒有的眼睛」[55]。

社會性標記提供了一種直接且直覺的方式讓大眾透過標籤來了解或評判藝術作品的特質，如梵谷的名畫《星夜》(Starry Night)可能出現的標籤，就會有「星星」、「星球」、「漩渦」或是「精神錯亂」等；佛萊斯(Joachim Friess)的裝飾雕塑作品《黛安娜與雄鹿》(Diana and the Stag)，「鹿角」、「箭術」和「女獵手」等標籤如預期出現，但「珍貴的」、「奢華的」等意想不到的標籤也同樣出現[55]。大都會美術館收藏品資訊規劃總經理鍾蘇珊(Susan Chun)表示：「美術館和大眾之間的確存在著巨大的語義鴻溝」[55]。

史帝夫美術館(Steve.museum)就是由多家美術館聯合進行的標籤計畫[35]，讓一般民眾可以在線上美術館加註標籤，除了縮短學者專家和一般民眾對藝術作品感知的鴻溝，更提供對藝術作品意想不到的視野，同時讓這些專業的說明與標籤並列(圖 1)。除此之外，尚有多個線上美術館進行標籤研究，如：澳洲雪梨動力博物館(Powerhouse Museum in Sydney, Australia)、史密森創新攝影(Smithsonian Photography Initiative)，其目的都是運用Web 2.0的精神，對藝術及美術作品提供新型態的語彙。

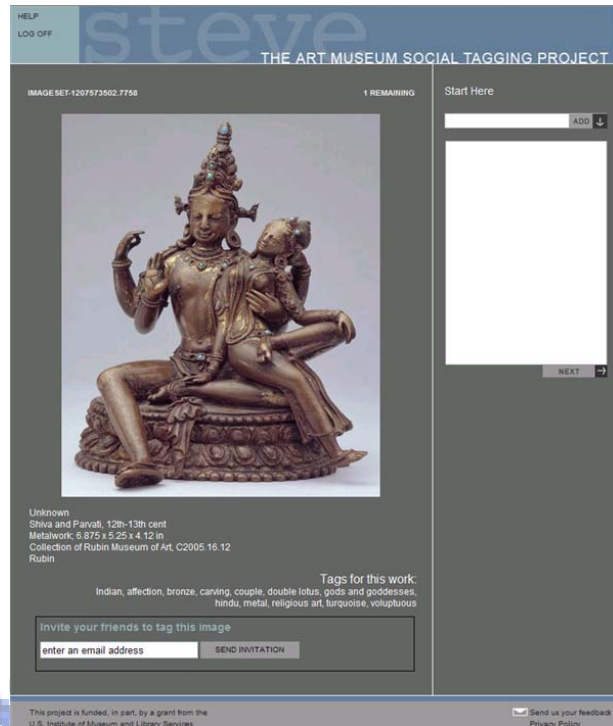


圖 1：史帝夫標籤計畫 (<http://tagger.steve.museum>)

本研究即是將 Web 2.0 的精神應用於楊英風數位美術館 (<http://yuyuyang.e-lib.nctu.edu.tw>)，線上美術館的優點是讓人們可隨時隨地從網路上觀賞藝術作品，將社會性標記應用於線上美術館的目的在於將人們觀賞藝術品的觀感串連起來，希望能透過大眾集思廣益讓每幅畫的描述能更完整、更多面向，並建立一般人甚至於專家都看不見的巧妙關聯性。

本論文系統實作先以中研院中文詞庫小組(Chinese Knowledge and Information Processing, CKIP)所開發的中文斷詞系統依據權重萃取出每一個藝術作品的關鍵詞彙，代表此藝術品專業的說明；再透過標記(Tagging)活動，讓一般使用者自由地對楊英風美術作品提供標籤；之後系統會結合作品關鍵詞和使用者標記的標籤進行階層式標籤分群法(Hierarchical Tag Clustering)，自動產生每一類作品的標籤地圖。此標籤地圖就是結合專業的字彙加上大眾標記的標籤共同建立的大眾分類法，當愈多人對楊英風數位美術館的內容提供愈多標籤時，則標籤地圖便會愈完整，而大眾分類法就會自動創造出更完整的分類。當標籤地圖建立時標籤間的關聯性也建立了，系統會根據此關聯性產生對每一藝術品的推薦詞，其

實也就是符合 Web 2.0 的精神，當愈多人使用得到的回饋也愈多。

二、 研究方法與目的

本研究是以交通大學數位典藏計畫的楊英風數位美術館網站為基礎架構，採用 CKIP 詞庫小組的中文斷詞系統先萃取出網站內藝術品關鍵詞，除了擴充既有作品的詮釋資料(Metadata)外，也可當作使用者建立標籤前參考的來源。接下來會將使用者建立的標籤與系統萃取的詞彙進行合併，自動產生階層式標籤分群，利用此技術可有效彙整資訊來源並利用標籤產生每一類別藝術作品導覽之標籤地圖。本研究之目的如下：

1. 透過社會性標記創造標籤之間的關聯性，並自動建立大眾分類法，破除以往耗時間及耗人力的傳統分類。
2. 利用標籤地圖提供線上美術館完整的索引架構，讓使用者可以快速搜尋並了解站上的資源。
3. 建立標籤地圖後，當使用者瀏覽線上美術館時，可利用標籤之間的關聯性推薦相關聯的詞彙，進一步發掘作品與作品之間的關聯性。

三、 論文架構

本論文第二章為文獻探討，將會依序介紹 Web 2.0 概念、社會性標記、中文斷詞研究、資訊檢索技術、文件及關鍵字分群。第三章說明系統架構及系統實作。第四章以問卷調查分析使用者對社會性標記應用於楊英風數位美術館的需求、標籤內容分析、標記工具評估及使用滿意度。第五章為結論及未來的發展方向。

貳、 文獻探討

本論文將社會性標記應用於數位典藏，結合資訊檢索技術、中文斷詞處理，將標籤與關鍵詞合併後，利用階層式標籤分群法建立標籤地圖，並將結果應用於標籤推薦與資源搜尋上。

第一節將對 Web 2.0 的概念作探討；第二節介紹社會性標記之相關文獻，包含社會性標記與知識管理及現階段標籤之相關研究；第三節介紹中研院詞庫小組中文斷詞的研究；第四節介紹本研究中用到的資訊檢索技術；第五節介紹文件及關鍵字分群。

一、 Web 2.0 相關研究

Web 2.0 並非一種新的技術，而是一種精神，在於讓每個人成為資訊的創作者而非只是使用者，藉由大家使用資料、創作資料，不斷互動與增長，促成一個龐大的知識平台，目前崛起的網站如：維基百科(Wikipedia)、Del.icio.us、Flickr、YouTube 等皆屬 Web 2.0 的應用。

1. Web 2.0 起源及意義

「Web 2.0」的概念是來自於O'Reilly與MediaLive這兩家網路公司於一場國際研討會所提出來的。在 2001 年的秋天，許多美國網路公司股價大跌，大家開始認為網路只是誇大不實的科技而已，認為網路就此泡沫化，相反地，Dale Dougherty這位O'Reilly公司的副總經理認為自 2000 年網路泡沫以來，網路產業正處於一個新的演進時期，新的網路應用、網站經營型態會慢慢誕生，甚至可以稱之為網路的文藝復興時期，因此，他採用軟體版本更新的命名方式，稱呼過去的網路時代為Web 1.0，現在的網路時代為Web 2.0[50]。

事實上雖然網路泡沫化的現象讓網路產業陷入一片低迷，但這只不過是將一些體質不佳的公司淘汰，完全符合適者生存的道理。許多專家發現一些重要、受歡迎的網路新應用順利度過此次網路泡沫化危機，如：網路書店 Amazon、拍賣網站eBay、及一些部落格服務提供商，而這些存活下來的網

路公司，都具有一些共同的特色，即Schauer[30]所提出的Web 2.0 特徵：「使用者貢獻價值」、「長尾定律」、「網路效益」。

O'Reilly公司在研討會中曾經比較Web 2.0 與Web 1.0 差異(如表 1)，說明如下：

- (1) DoubleClick 曾是一家最大廣告網路服務商，在 Web 1.0 時代的網路廣告，多以橫幅廣告的方式呈現，大多置放於網頁最上方，以廣告被刊登的次數向廣告主收取費用，其優點是簡單明瞭，但七彩絢麗的橫幅廣告圖成為網頁瀏覽最大的負擔。於 Web 2.0 時代，Google 推出 AdSense 服務，揚棄華麗的橫幅廣告，改用文字廣告，且以「每次點擊成本」向廣告主收取費用，並且長期收集使用者的習慣來播放廣告。
- (2) Web 1.0 時代 Ofoto 只能以相簿使用者來搜尋相簿，而 Web 2.0 時代 Flickr 則提供使用者標籤功能，讓使用者能依據標籤搜尋相簿。
- (3) Akamai、Mp3.com 提供資料或音樂的網站只能被動讓使用者登錄，挑選檔案來下載，但到了 Web 2.0，BitTorrent 或 Napster 可以讓使用者直接點對點連接分享互相電腦內的資源，不再只是從特定網站上抓取資料。
- (4) 維基百科的誕生就是一個Web 2.0 最重要的特質— 群眾智慧，它摒棄大英百科全書權威智慧的方式，而採用眾人集體編輯，在短短幾年間就已經有一百多種語言的內容[52]，其所呈現的是一種共同創作的模式。
- (5) 個人網頁通常只能瀏覽網站主人所公布的資訊，而部落格不僅擁有個人網頁的各種功能，並且整合了相簿、留言版、日記等，而使用者更可與作者共同參與網站內容的編輯。
- (6) 在 Web 1.0 時代，Evite 是透過網路寄發邀請函，由網站經營者自己對外發佈訊息，而 Web 2.0 社交行事曆的 Upcoming 以及社交軟體(EVDB) 則是將訊息放在網路上由大家共同來編輯及創造。
- (7) 在互動方式上，Web 1.0 時代的資訊通常只由網站提供，使用者只是被動接收，而 Web 2.0 則是由使用者共同參與及貢獻資料。

(8) Web 1.0 的目錄分類俗稱「古典分類」或是「專家分類」(Taxonomy)，是由專家來決定分類的名稱；而 Web 2.0 的標記功能讓使用者自行為喜愛的項目給予標籤，讓大家共同制定的標籤形成大眾分類，使得分類更符合使用者的需求。

表 1：Web 2.0 vs Web 1.0 (整理自 Tim O'Reilly “What is Web 2.0”)[25]

Web 1.0	Web 2.0
DoubleClick	Google AdSense
Ofoto	Flickr
Akamai	BitTorrent
Mp3.com	Napster
大英線上百科全書 Britannica Online	維基百科全書 Wikipedia
個人網站 Personal Websites	部落格 Blogging
Evite	Upcoming.org and EVDB
網域名稱投機化 Domain Name Speculation	搜尋引擎最佳化 Search Engine Optimization
頁面瀏覽次數 Page Views	每次點擊成本 Cost Per Click
螢幕抓取 Screen Scraping	網路服務 Web Service
發佈 Publishing	參與 Participation
內容管理系統 Content Management Systems	Wikis
目錄 Directories (Taxonomy)	標籤 Tag(Folksonomy)
黏性 Stickiness	聚合 Syndication

Jim Cuene[7]提出從使用者觀察Web 1.0 與Web 2.0 的差異(如圖 2)，不難發現Web 2.0 網站的共同點都是以人為基礎，無論是 Wikipedia、Flickr、YouTube、Del.icio.us，這些網站都是由大量使用者各自提供少量的資訊，最後集成龐大的資料庫。



圖 2：Web 2.0 vs Web 1.0 (資料來源：Web 2.0: Evolution towards a read/write platform) [7]

O'Reilly認為Web 2.0 並沒有一個很明確的界線，如同圖 3所示，它呈現的是一種核心精神所形成各種概念及應用的延伸，上方的七個橢圓代表各種網站的應用，如：Flickr、Del.icio.us、Wikipedia、Blog等，下方則是所延伸的概念，如：長尾理論、觀念創新並非技術、參與、豐富使用者體驗等，中間則代表Web 2.0 最核心的精神，如：策略定位、使用者定位、核心競爭力等。

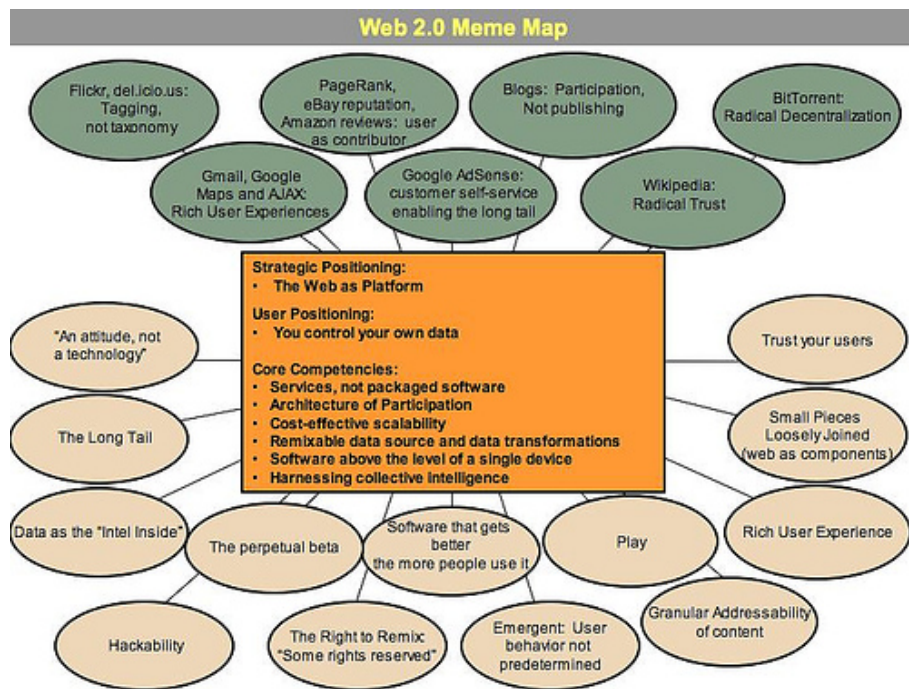


圖 3：Web 2.0 Meme Map (資料來源: Tim O'Reilly "What is Web 2.0")[25]

O'Reilly認為Web 2.0 最重要的觀點就是以使用者為中心，包含下列三項重點[54]：

1. 鼓勵使用者參與內容創作，或是使用者之間的互動，如 YouTube 或部落格都是強調使用者間的互動與參與創作。
2. 開放的重要性，透過開放的討論，互相分享創造，才能豐富使用者彼此的經驗與網站資源。
3. 強調使用者網路的外部延展性，也就是去中心化的型態，如 Skype、BitTorrent。

目前正當 Web 2.0 火紅之際，Web 3.0 這個新名詞也隨之誕生，Web 1.0

時代，主要以單向閱讀為主，Web 2.0 則強調雙向、互動、分享「讀—寫」發展，而現今最新提出的 Web 3.0 概念，一般被闡釋為智慧更高的網路服務，不管是 Web 1.0、2.0 或 3.0，目的地都是希望最終能創造更貼近使用者的網路服務。

2. Web 2.0 主要原則

O'Reilly 在 2005 年發表的一篇文章(What is Web 2.0 : Design Pattern and Business Model for the Next Generation of Software)[25]，提出 Web 2.0 應用的七個原則，以及 Musser 和 O'Reilly 在 2006 年定義的 Web 2.0 八個核心模式(如圖 4)，可說是目前 Web 2.0 的主導思想，將之綜合整理如下[12][57]：

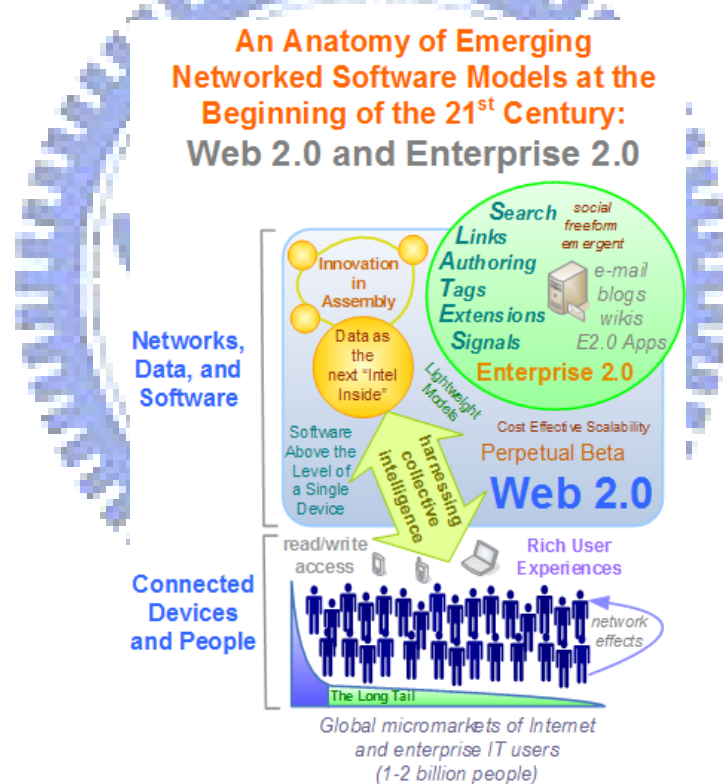


圖 4：Web 2.0 八個核心模式 (Web 2.0 definition updated and Enterprise 2.0 emerges)[12]

(1) 網路應當被視為平台 (The Web as Platform)

Web 2.0 並沒有嚴格的界線範疇，但是卻有一個重要的核心，就是去中心化，它是由「原則」與「實踐」兩個面向所構成的，如同一個擴散式的地圖，環繞於此核心，可向外延伸出許多原則與實踐。

在過去 Web 1.0 時代裡，DoubleClick 及 Akamai 已是網路平台的先驅

者，兩個網站都是利用與其它網站的合作，以獲得營業收入。之後 Web 2.0 的後進者便深入地應用這些特性，如 Google 集合眾多小網站的力量提供廣大的資訊給使用者，這就是一種去中心化，將網路遍及每一處。

(2) 運用集體智慧 (Harnessing Collective Intelligence)

集體智慧是網路平台化後的一個必然發展趨勢。如 eBay、Amazon 皆是不斷地由使用者累積參與網路平台的結果，讓後來的使用者可以獲得更好的服務。另外一種創新的概念是像 Wikipedia 線上百科全書或是 Del.icio.us 線上書籤網站，讓使用者自動自發撰寫資料、維護百科全書的詞條，或是自訂書籤的分類標籤，這無疑是一種使用者彼此之間信任的概念，內容創建更是一種集體智慧的發揮。

O'Reilly 認為在 Web 2.0，使用者願意貢獻他們自有資料與內容，對資料再使用將可獲取更高價值的方式。在 Web 2.0 網路應用的關鍵競爭優勢，就是創造一個參與架構，最重要的核心便是將使用者納入所創造的集體智慧。

(3) 資料是下一個「Intel Inside」(Data is the Next Intel Inside)

「Intel Inside」的含義原本出自於不論何種電腦，只要在外殼上貼上「Intel Inside」的標籤就代表電腦內部的CPU是由Intel所生產。在Web 2.0 裡，資料是下一個「Intel Inside」意味著對於一個成功的網路平台而言，指的是資料而非功能，例如：Google 的搜尋引擎、Amazon 的產品型錄、eBay 的拍賣資料，MapQuest 的地圖資料庫，Napster 的分散式歌曲庫以及 YouTube 的影片資料庫等，許多重要性的網路應用系統背後都有一個專屬的資料庫。所以，資料庫管理將會變成Web 2.0 公司的核心競爭力[57]。

(4) 軟體不斷發行與升級的循環將會終結(「永久的 Beta 版」)(End of the Software Release Cycle)

網路時代最重要的特徵就是以服務為目的，這使得過去軟體公司營運模式產生了徹底的變化。其變化有二：第一是日常營運必須成為核心能力，如

Google必須不斷地搜尋網路，更新其索引目錄，並淘汰無效的連結，還必須持續地回應成千上萬不同步的使用者查詢，並在提供結果的同時顯示與查詢項目有關的廣告，這種日常營運已成為Google主要的核心競爭力。第二，使用者成為共同開發者，即是指開放原始碼與使用者共同開發的過程，現在的軟體開發每一天、每個星期、每個月都推出新的功能，這代表了軟體開發永遠沒有結束的一天，即所謂「永遠的試用版」。如Gmail、Google Maps、Flickr、Del.icio.us，推行好幾年，仍然只是試用版，其中某些使用率太低的功能可能會被取消，受歡迎的功能則會留下來[57]。

(5) 輕量型程序設計模型 (Lightweight Programming Models)

網路之所以獲得如此巨大的成功，其原因之一便是採用一種簡單輕巧的開發模式，如RSS之所以成為應用最廣泛的網路服務，就是因為簡單，Web 2.0是一種創新的網路模式，正是一種鬆散耦合的系統，輕巧的程式撰寫正符合這種模式。

(6) 組合的創新 (Innovation in Assembly)

組合式的創新是透過混合的資料與服務創造更多新的商業機會，如Google近來的Google Maps，就是將簡單AJAX加上新的資料成為新的服務，透過這種狀態，網路服務將出現更多混和運用。

(7) 軟體執行將跨越單一設備 (Software Above the Level of a Single Device)

Web 2.0的另一項特徵便是不再侷限於個人電腦的平台上，軟體與應用服務的設計，需要能夠適用於個人電腦、行動裝置，以便整合更多的資料與服務。iTunes可說是落實這項原則的最佳典範，這項應用從個人手提裝置緊密連結到龐大的網路後台，如iPod/iTunes的結合就著眼於跨越不同裝置[57]。

在Web 2.0的領域，將會有越來越多外在裝置與網路平台相連，當電話、汽車所提供的是輸入資料的話，而這個新興平台將可呈現交通流量、市民新聞等更多潛藏的資訊[57]。

(8) 豐富的使用者體驗 (Rich User Experience)

在歷經網路泡沫化之後，網路曾經被視為誇大不實的應用，一直到了 Gmail 和 Google Maps 的發展，不但具備豐富的使用者介面，也有與個人電腦同等級的互動，才算喚醒主流市場。

Gmail 在電子郵件發展出許多有趣的創新服務，結合網路的優勢，可在任何地點收發以及具有深度的資料庫功能、搜尋能力，利用 AJAX 簡單的技術呈現接近個人電腦方便性的介面，事實上，這種正逐步結合 E-mail、即時通訊軟體(Instant Messenger)、手機，並運用 VoIP 技術讓語音傳輸能力結合至網路應用將會慢慢地展開。

(9) 拉動長尾的能力 (Leverage the Long Tail)

一間商店裡，80%的收入來自 20%的貨品；一所圖書館裡，有 20%的藏書為 80%的讀者借閱；由Chris Anderson在 2004 年發表的「長尾效應」卻顛覆了 80/20 法則(如圖 5)，它是指只要發布渠道夠大，又能夠集合零散的需求，非主流商品的利潤同樣可觀。以亞馬遜(Amazon)網上書店為例：傳統書店裡滯銷的書籍，在亞馬遜卻有不錯的銷量，據統計，熱門的書籍與冷門的書籍同佔銷量的 50%^[36]。

Google 的 AdSense 及 Adwords 廣告策略，就是長尾理論應用的最佳實證，它著重於為數眾多的小企業和個人網站，讓所有小企業都能在全世界任何網站上賣廣告。這種透過網路所賦予低成本且廣泛的接觸性，正是拉動長尾的能力。

Long Tail Diagram

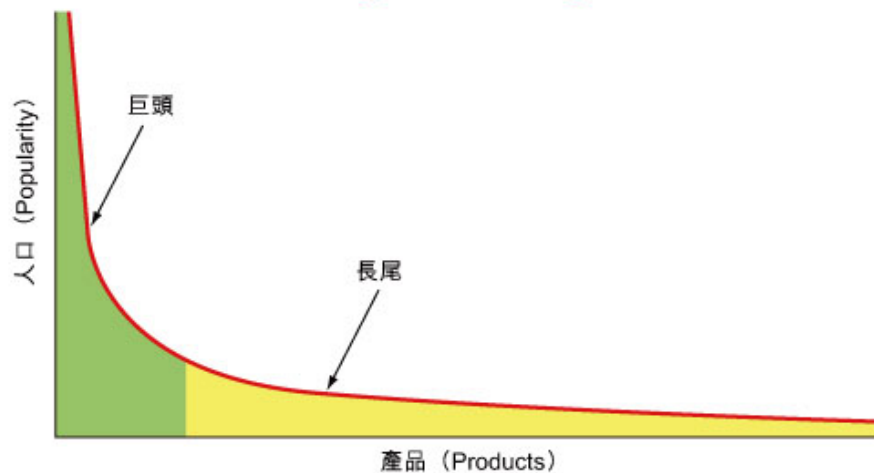


圖 5：Long Tail-Diagram [36]

(10) 輕巧模式且節省成本的架構 (Lightweight Models and Cost-effective Scalability)

透過輕巧的軟體以及經營模式的改變與創新，都是在根本上改變網路軟體的發展，為成本效益的提升，提供強而有力的武器。

3. Web 2.0 的應用類型

Web 2.0 創新應用案例集[59]中將Web 2.0 網站的服務型態，以應用方式區分為混搭服務(Mashup)、網路軟體(Webware)、共享平台(Social Media)、社群網路服務(Social Network Service)以及播客服務(Podcast)等五大類。

(1) 混搭服務 (Mashup)

混搭是將不同的服務、資料搭配在一起，創造出一種新的服務。混搭系統運用巧思與創意提升網站的附加價值，提供使用者多元的服務，主要重點就是網站之內容與功能係整合於其他網站所提供之API。目前如UrMap (如圖 6)、Google Map、HousingMap.com、eBay、Amazon與Yahoo等皆提供API供其他網站整合。

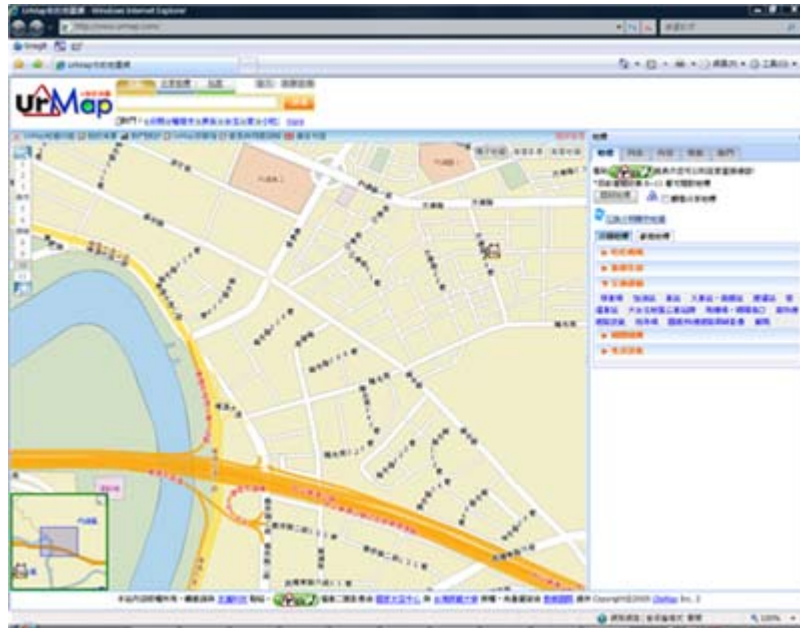


圖 6：UrMap (<http://www.urmap.com/>)

(2) 網路軟體 (Webware)

Web 2.0 概念的崛起使各種不同的網路軟體持續吸引使用者的使用，其功能與操作介面已和傳統桌面電腦的應用軟體相似，可直接在網站上使用，如 Google 買下了 Writely.com 之後提供了 Doc & Spreadsheets 工具，等同於 Microsoft 的 Word 跟 Excel 的功能，網路軟體將成未來的趨勢，亦即只要上網，需要甚麼軟體就可以用到甚麼軟體。

(3) 共享平台 (Social Media)

共享平台是由服務業者在網站上提供一個儲存與展示的空間，而使用者一般需註冊為會員，將數位內容上傳到網站上，由眾多使用者上傳的內容形成廣大的資源，基本上以文字、圖片、影片、音樂等數位內容檔案，讓使用者之間互相分享，如無名小站(圖 7)。



圖 7：無名小站 (<http://www.wretch.cc/>)

(4) 社群網路服務 (Social Network Service, SNS)

社群網路服務是讓使用者根據不同的理念、主題、興趣在社群網站中形成線上社群，服務業者在網站上提供各式工具與技術，包括共享書籤、共享照片，讓使用者藉此進行聯絡、溝通與分享，因而創造出非主流的新聞、音樂、照片、影片等媒體內容。目前社群網路服務所提供的型態越來越多樣化，這也是經營者之間差異化的基礎，如黑糯米共享書籤(圖 8)透過網友收集不同網路書籤，進而產生各式各類的社群，讓使用者在其上分享彼此的興趣與喜好。



圖 8：黑米共享書籤 (<http://www.hemidemi.com/>)

(5) 播客服務 (Podcast)

Podcast 是由「iPod」與「Broadcast」兩字組合而成，意即 MP3 Player 與廣播功能的結合，是數位廣播技術的一種，它改變了過去人們被動收聽廣播的方式，使聽眾成為主動參與者。而 Podcast 與傳統廣播最大的不同在於能夠透過 RSS 訂閱的方式，讓聽眾即時下載電台或網站上最新的 MP3 檔案，透過隨身裝置即可進行收聽，如 iPod。Podcast 可以自由選擇收聽的內容、收聽的時間以及以何種方式與其他人共享，讓創作者能夠擁有自己的宣傳管道，找到喜愛自己創作的聽眾，顛覆過去傳統媒體的播放權，人人都可能成為網站的主播。

二、 社會性標記 (Social Tagging)

社會性標記是一種 Web 2.0 的概念，透過共同建立標籤的方式，讓使用者在同一平台上共同參與及分享資源，一些代表性的網站如：Del.icio.us、Flickr 受到大家的歡迎，而且也引起廣泛討論；WWW 2006 研討會中亦舉辦 Collaborative Web Tagging Workshop 探討如何從標籤中擷取資訊。

在社會性標記中，使用者從以往單純的資料接收者轉換成資料提供者。一個

標記行為包含三個層面：使用者、資源、標籤，並衍生一個重要的概念，就是大眾分類，詳細內容會在本節所述。

1. 標籤與標記的起源

標籤起源於社會性書籤，即個人儲存與分享網站資源的服務，因為收藏的超連結可以被許多人在網路上分享，因此也有人稱之為網路書籤或是線上書籤[40]。它就像內建於瀏覽器的「我的最愛」功能。社會性書籤的優點是第一：任何電腦都可以連結到收錄在社會性書籤中的網站或網頁，並不侷限於單機電腦中；第二：社會性書籤網站可以將喜歡的網頁大量地收錄起來，並且給予標籤，即可達到分類、管理、收錄龐大的網頁資料等目的，形成一種個人的資料管理中心。

最早提供書籤服務的是 Del.icio.us(<http://del.icio.us>)這家公司，其網址的三個部份組合起來是英文字 delicious (美味、好吃)，所以又稱「美味書籤」。在臺灣至少有兩家公司提供了社會性書籤的服務：一家是智邦的 MyShare 服務；另一家就是前一節所提到的黑糯米共享書籤 HEMiDEMi。

使用者為網站內容給予標籤的行為稱之為標記，是 Web 2.0 時代出現的典型應用，所收藏的內容不再侷限於網址，開始擴展包括音樂、影片、相簿、書籍等任何事物，社會性書籤變成眾多資源類型中的一種應用，舉例來說，某一篇文章講述的內容是社會性書籤，如果使用者想把它記下來，日後保留或跟別人分享，就可以把它收錄進書籤裡，並貼上標籤如：「網路書籤」、「社會性書籤」、「共享書籤」、「網路服務」、「Web 2.0」、「服務介紹」或是其他想的到的詞做為標籤。這些標籤可以作為搜尋資源的依據，不僅能拓展個人閱讀的範圍，更可找到與自己興趣相同的人，了解該領域其他人所搜集、閱讀的訊息。

社會性標記已儼然成為 Web 2.0 的典型應用，在 2005 年 O'Reilly 掀起了一股 Web 2.0 相關議題與應用的熱潮後，同年 3 月 Yahoo! 正式收購影像分享

網站Flickr，並於同年12月，再度收購網路書籤分享網站Del.icio.us。身為大型主要入口網站的Yahoo! 收購這兩家提供社會性標記服務的網站之主要原因來自於：「透過龐大的志工，依照詮釋資料來作搜尋，並且共同替資訊作分類」[26]。龐大的使用者自由地進行標記，累積出的獨特分類目錄，更能貼近使用者體認，以有別於傳統以知識為中心的分類方式，如圖書館所習用的分類系統[19]。根據Mathes[38]的研究從給予的標籤可以發現總共有60,000名到300,000名使用者(累計從各地的Del.icio.us和Flickr使用者)，並創造出多達四百萬個文件。相信隨著日後使用者人數的逐日攀升，其影響力將會擴及整個網路的生態。這種由共同創作所形成自由標記的分類系統，已經衍生一種新的分類方式名詞，一般稱為大眾分類法或通俗分類，將在下一小節介紹。

2. 大眾分類法 (Folksonomy)

「Folksonomy」這個新的詞彙是由Tomas Vander Wal在討論Flickr和Del.icio.us的資訊架構時，將「Folks」和「Taxonomy」組合，創造出新的詞彙。他定義Folksonomy是：“A folksonomy is the result of personal free tagging of information and objects (anything with a URL) on the internet for one’s own retrieval. The tagging is performed in a social environment (shared and open to others). The tagging action is done by the person consuming the information.”[37]

「Folksonomy」是由Taxonomy(古典分類、學科分類或專家分類)所發展而來，並以「Folks」顯示其自由分類的特性。Folk表示一群人，一伙人的意思，-sonomy是由Taxonomy一字演變而來，表示有系統、專門的學科知識分類法；合而為一的意思為大眾所產生的一種分類知識，也可稱為一種「由下而上的社會分類法」(Bottom-up Social Classification)[37]，或是Shirky[32]所形容的「社會性創造的平面化命名空間」(Socially Created, Typically Flat Name-spaces)，強調其社會性分類(Social Classification)的概念，而較不去區

分專家與大眾之別。在翻譯上則有「分眾分類法」，「通俗分類法」，「大眾分類法」，「民眾分類法」，「民俗分類法」等。

Folksonomy因其涵義甚廣，還可發現存在多種說法，包括如協力式分類(Collaborative Classification)、協力式標記(Collaborative Tagging)、自由標記(Free Tagging)、標籤分類學(Tagsonomy)、基層分類(Grassroots Classification)[20]、民眾分類(Ethnoscience)[20]、社會性分類(Social Classification)[9]等。相關概念還包括強調社會性或互動取向的社會性軟體(Social Software)、社會網路(Social Networks)；或強調資訊組織取向的命名(Labeling或Naming)、關鍵詞索引(Keyword Indexing)、分類(Categorization)、分類(Classification)、分面分類法(Faceted Classification)、分面索引法(Faceted Indexing)、知識本體(Ontology)、語意網(Semantic Web)等概念，相當多元與龐雜[44]。

在此可以發現Folksonomy存在許多譯名，同一個名詞卻代表許多種意念，實不易單用一個名詞來定義，每個名詞所關注焦點不同，也反映出不同領域對同一概念的各種觀點。卜小蝶[44]指出Folksonomy可由三種概念思考，依序為使用者、資源、及分類(如圖 9)，說明如下：

- (1) 使用者角度：產生如 Social、Ethno 的概念。
- (2) 資源角度：有各類物件資源如 Bookmark、Photo 的概念產生。
- (3) 分類角度：Tagging、Labeling、Classification、Categorization、Indexing 等則是常見的概念。

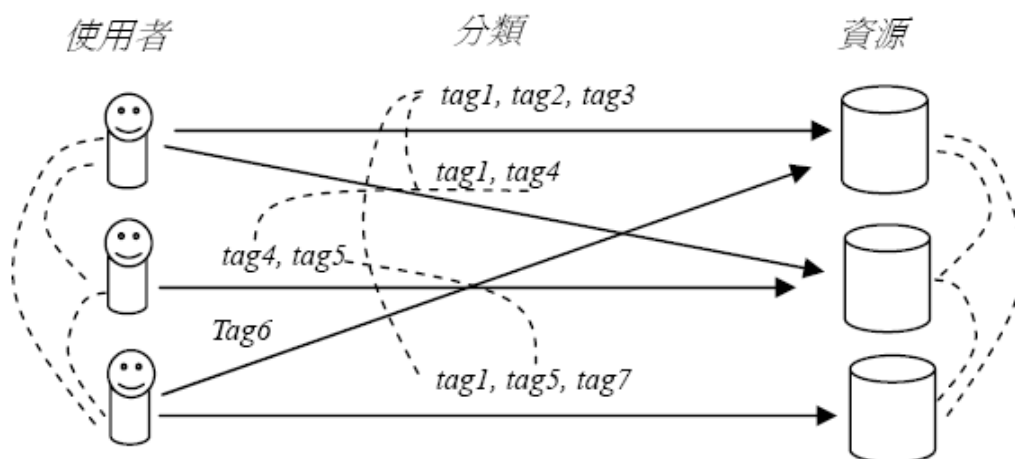


圖 9：社會性標記之概念架構[44]

故可知 Social Tagging 強調的是 Social 與 Tagging 兩種構面，Social 關心的是使用者，而 Tagging 則較關心資源及分類的內容與結構，而 Social Tagging 的結果，最後形成 Folksonomy。

總體來說，大眾分類是一種分散式分類系統的類型，根據 Wikipedia 歸納具有三項特點[41]：

1. 大眾分類是由個人自發性標記與定義而來的。
2. 標籤是公開分享，可以被所有使用者看到。
3. 大眾分類是由使用者群體標記詞彙的頻率來決定分類法。

由使用者、分類、資源(網站內容)組成的大眾分類與專家分類進行比較，雖然大眾分類的架構並不嚴謹，也未事先定義，且內容也是因個人所好、自由發揮，而顯得品質參差不齊。反觀專家分類其架構多事先製定也較為嚴謹，但大眾分類擺脫了固化的現象，跟使用者的認知程度密切的結合，而且方便、靈活，不受條件限制。所以這種以自訂標籤形式的大眾分類在目前流行的社會性網路服務中得到了廣泛的應用。

3. 大眾分類的優缺點

從前一節最後可知大眾分類法雖然有許多項優點，但也有許多缺點需要

解決，列舉優點如下[43]：

1. 定義標籤者即為內容使用者，使用者認同感較高。
2. 具回饋性，可幫助社群創造溝通與分享空間。
3. 具集體智慧，如透過字彙與概念的變化，呈現流行主題。
4. 直接反應使用者需求，如資源排序是以使用者點擊數為主，較其它以連結數為計算基礎的排序方式，更符合使用者需求。
5. 具語言及文化的豐富性，例如同一網站，各國人士所訂定關鍵字，可以反映不同文化觀點。
6. 可以包含少數人興趣，較無偏差，網路應用兼容並蓄，可以大者恆大，也容許少數人的興趣存在。
7. 成本低廉。
8. 開放性、機動性強。
9. 具啟發性，可協助使用者發現、探究，以尋得原先未知的資源。

大眾分類法也有不少缺點，舉例如下[43]：

1. 單詞索引(Single Word Indexing)，關鍵字只能以一個單詞表示，造成的不一致現象。如 Google Maps 就有 googlemaps、google_maps、google-maps 等不同形式，造成檢索時相當困擾。此外，單詞也不易表達複雜概念。
2. 缺乏控制字彙，降低資料的有用性與接受度。許多標籤語意模糊、不精確、無區分性，又缺乏同義詞、同形異義詞等控制，標籤的重複比例相當高。
3. 錯別字太多。
4. 無分類架構，較無脈絡可循。
5. 不易瀏覽，查詢功能過於簡略。
6. 無標記指導原則，標籤格式及給定原則缺乏標準，不易達成一致性。
7. 有心人士做為商業廣告管道，影響其有用性。

4. 社會性標記網站的應用

由於社會性標記網站眾多，本節將介紹三個應用於不同領域的標記網站，第一個是一般社會性書籤網站 Del.icio.us，第二個是將標籤運用於大學圖書館的 PennTags，第三個是與本研究應用相關，將標記的概念應用於線上博物館網站 Steve.museum。

(1) Del.icio.us

Del.icio.us (圖 8)是早也是最著名的社會性書籤網站，是由 Joshua Schachter 於 2003 年 9 月成立的，它的概念很簡單，源自於個人瀏覽器的「我的最愛」，讓大家將自己的「我的最愛」書籤儲存於網路上，主要目的就是透過每個人儲存的書籤達到分享群眾知識的目的。Del.icio.us 是一個網路化的書籤管理工具，除了可以保存網站連結之外，還能與同好分享書籤，也可以透過標記自行為網站分類。

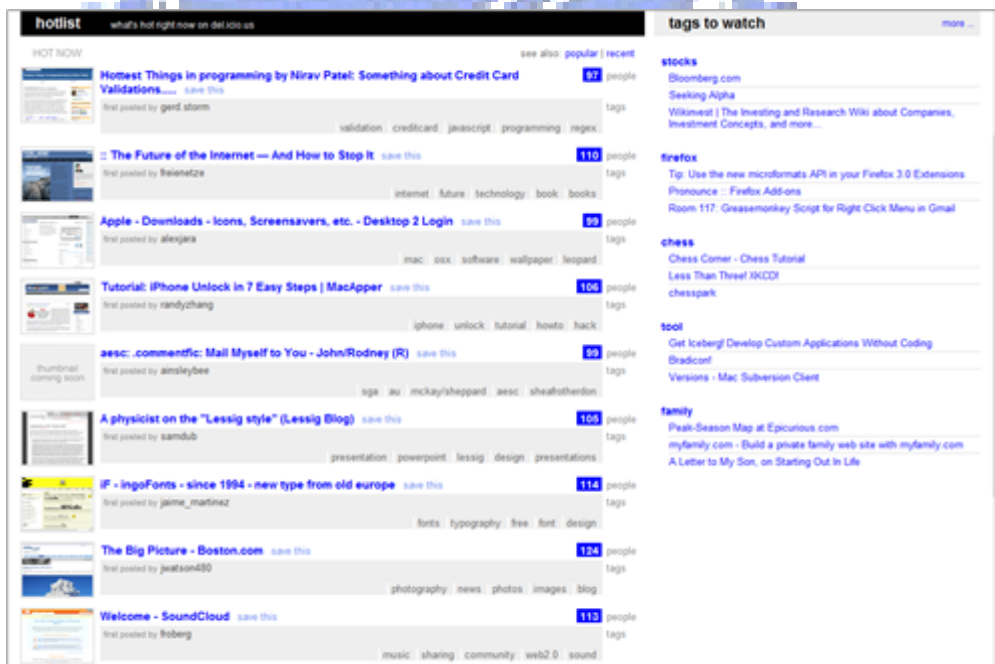


圖 10：Del.icio.us (<http://del.icio.us>)

Hend and Hugh[11]描繪出一個Del.icio.us的架構圖(如圖 11)，第一部分 Fetch a resource是將標籤、資源標題、URL等資訊截取出來，作為標籤的修正與擷取，為下一階段的前置作業。第二部分Tags則是透過使用者運用標籤

進而對知識本體予以加值。第三部分Resource with added value則產生資訊語義化的詮釋資料。

Del.icio.us 的標籤可讓使用者靈活運用，並可自由的新增或刪減，以符合自己的需求，隨著使用時間的增長，則會發現有愈來愈多相同興趣的使用者與資源，進而促進網站使用率提升。

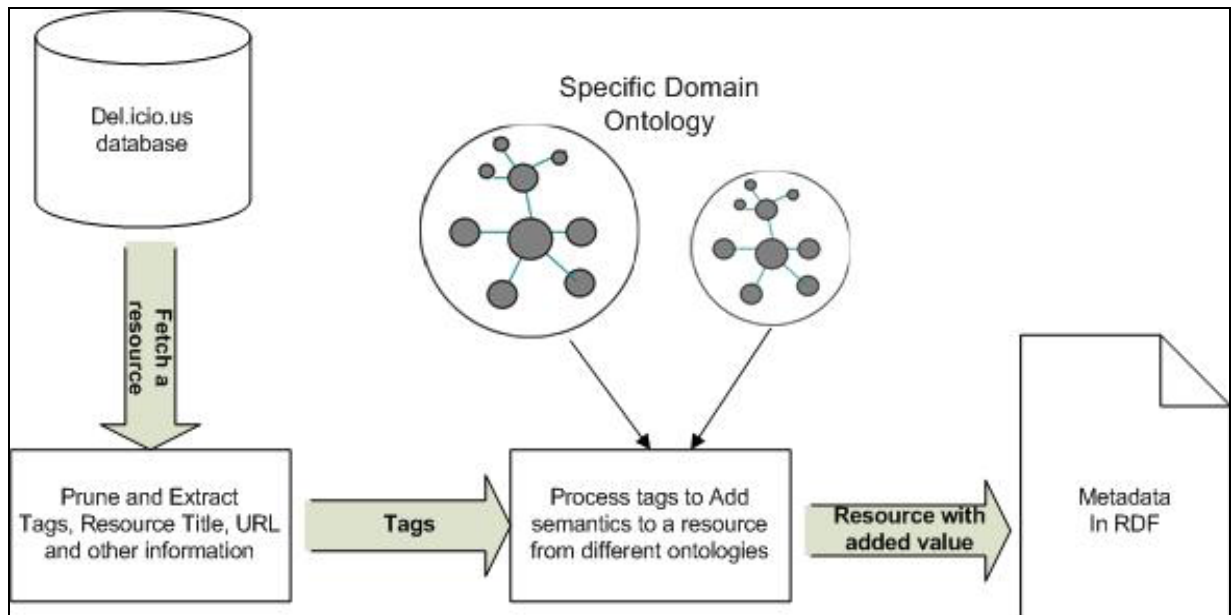


圖 11：Del.icio.us架構圖[11]

(2) PennTags

美國賓夕法尼亞州大學將社會性標記應用於圖書館而創造了 PennTags，它可以用Bookmarklet工具，只要在瀏覽器上安裝Bookmarklet之後，在書目記錄頁面時，點選一下瀏覽器上的「Add to PennTags Link」後便可以建立標籤。舉例來說：找到某本書名叫「Broadcasting it」時，會顯示出有哪些人為它加入了哪些標籤，當點擊到「television」這個標籤時，則可以連到television這個標籤類別，查看這個類別裡有哪些人加入了哪些書(如圖 12)。

在 PennTags 網站上，有五項功能，介紹如下：

1. 可以讓使用者訂閱某一類別的 RSS feeds，例如：訂閱 television 的 RSS。

2. 可以用分類標籤來搜尋。
3. 顯示出相關的標籤，例如：顯示出和「television」有關的標籤。
4. 顯示出某位使用者分類過哪些標籤。
5. 顯示出所有的計畫(Projects)、標籤擁有者(Owners)、和標籤(Tags)。

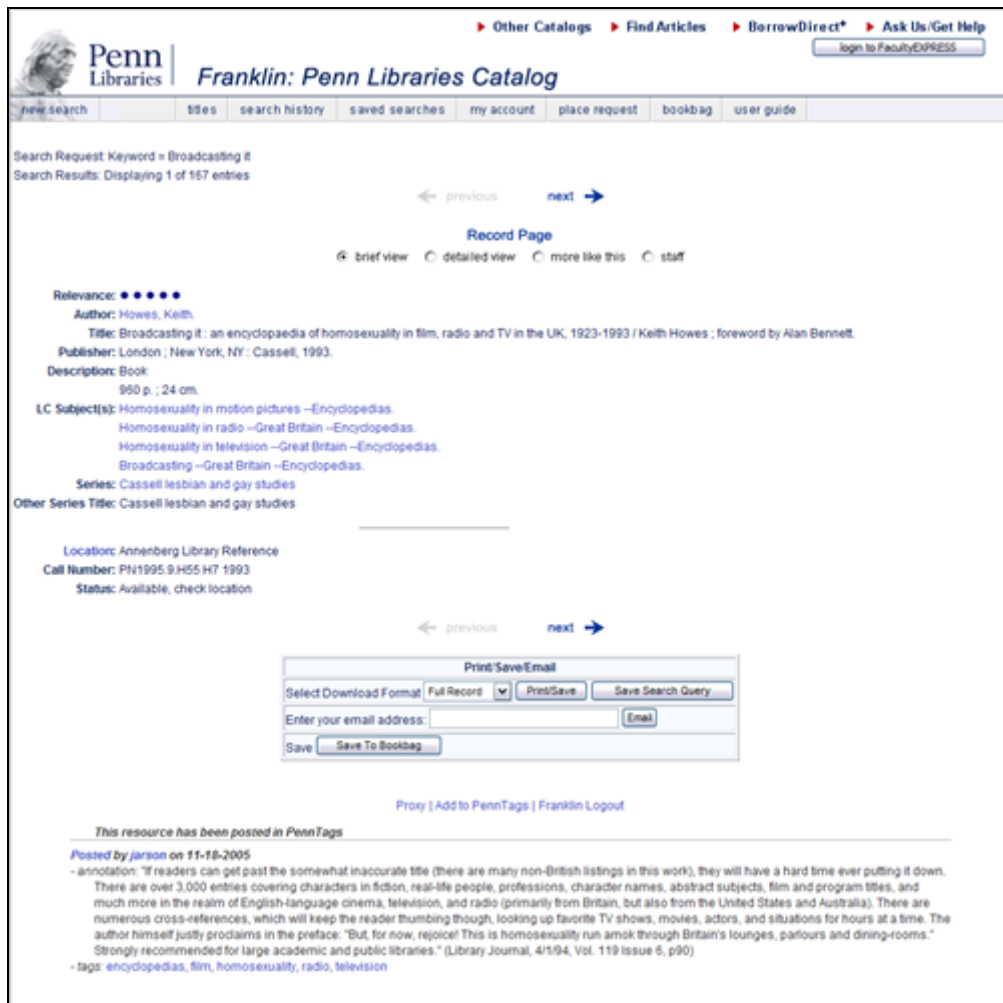


圖 12：PennTags 讀者分享註解與標籤的範例 (<http://tags.library.upenn.edu/>)

(3) Steve.museum

Steve.museum (圖 13)在本論文研究動機與背景曾經提到，它是一個應用社會性標記於線上博物館的系統，在 2005 年 4 月由 Cataloguing by Crowd 組織將社會性標記概念帶入博物館館藏內容裡[5]。主要任務是探索使用者對藝術作品給予的標籤，透過分析來改善館藏的擷取。過往博物館的館藏文件皆由專家進行詮釋[34]，但卻讓內容不容易為大眾所了解，並且失去以使用者

的興趣與觀點來描述文件[35]，該計畫採用社會性標記於數位博物館的目的有三種[6]：

1. 改善博物館線上館藏的使用。
2. 透過非正式的人員給予標籤，去驗證社會性標記可融入專家分類中。
3. 透過標籤讓線上博物館也享有社群的參與力量。

Steve.museum 的研究主要對使用者來源及詞彙內容作分析，以了解標籤的錯誤來源，進而改善系統對詞彙控制的依據，目前分析皆是為了用社會性標記及大眾分類來搜尋藝術作品效能所作的前置研究。現在已有多個博物館受到 Steve.museum 的影響，如：丹佛美術館(Denver Art Museum)、古根漢博物館(Guggenheim Museum)、克里夫蘭美術館(The Cleveland Museum of Art)、印地安納波里美術館(Indianapolis Museum of Art)等，逐漸想將社會性標記應用於數位美術館，以貼近使用者的方式增進資料檢索的效能。



圖 13：Steve.museum (<http://tagger.steve.museum>)

如圖 14 顯示 Steve.museum 一個簡單的資料模組架構，表示使用者

(User)、藝術品(Image)、標籤(Tag)之間的關係。物件(Object)以藝術品影像(Image)的方式呈現，使用者透過博物館(Institution)與物件(Object)產生聯繫，然後跟資訊環境(Environment)互動提供標籤(Tag)給予藝術品(Image)。所以在 Steve.museum 中，使用者、標籤、藝術品是主要核心，這與圖 9 的社會性標記架構不謀而合。

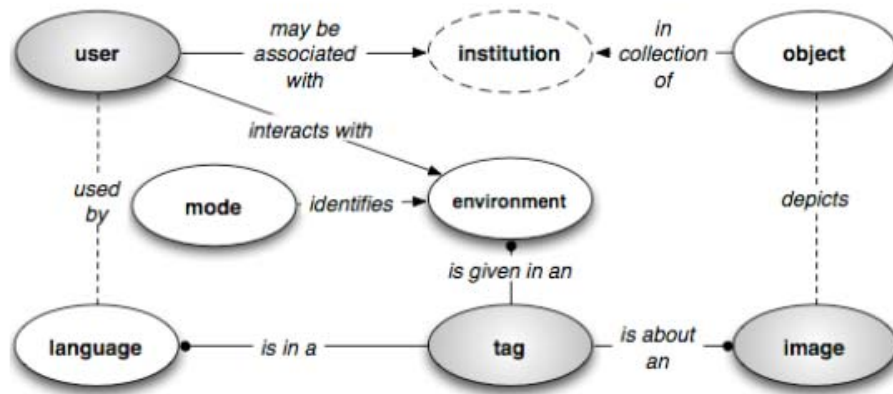


圖 14：Steve.museum 資料模組架構[6]

5. 社會性標記與知識管理

標籤與知識管理都是資訊與知識探勘的重要工具。於前幾節所述，標記活動是經由電腦網路的媒介，透過人們的溝通、互動與合作，為數位資源加註標籤，是一種由下而上的資訊分類方法，由使用者儲存與組織標籤的行為，逐漸延伸為一種知識管理的工具。

學者對知識管理的定義，多是由上而下的靜態設計，Wiig[39]認為包括四個流程：

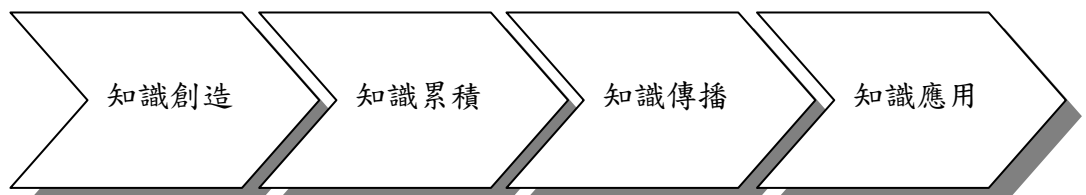


圖 15：Wiig知識管理流程[39]

- (1) 知識創造：知識創造不一定是新發明的知識，而是將既存的資訊賦予新的價值，知識不斷地創新，才能建立永續經營的活動。

- (2) 知識累積：藉由對資訊的有效編輯、分類、整理，讓使用者得以有系統的快速搜尋知識，並透過完整的資料庫進行知識累積的工作。
- (3) 知識傳播：將知識彼此分享，達到交流的目的，並進一步產生創新，才是真正知識傳播。
- (4) 知識應用：將知識達到真正重複的利用。

另外，O'Dell[24]也提出知識管理模型包含七個步驟：確認、收集、採用、組織、應用、分享、創造；Despres and Chauvel[8]也定義知識管理流程為描繪/掃描、獲取/捕捉、彙整/儲存、分享/移轉、再利用。由上可知，學者們所定義的知識管理流程雖然在名詞或順序上略有不同，但是其內涵是大同小異的，對應於社會性標記則演變為一種參與式的知識管理，以下根據Despres and Chauvel所定義的知識管理流程，將社會性標記知識管理活動闡述如下[47]：

- (1) 知識描繪/掃描：透過瀏覽及建立標籤的方式，對知識重新描繪與定義，並建立高效率的參考資源。
- (2) 知識獲取/捕捉：透過標籤分享將知識重新組織與連結。
- (3) 知識彙整/儲存：對於社會性標記而言，知識彙整/儲存常與知識獲取/捕捉相互連結，因其使用上相當直觀，故程序上常是一起進行，資訊是立即被發現，然後便彙整/儲存起來。
- (4) 知識分享/移轉：使用者運用標籤產生知識的內容，達到分享，並透過標籤協助搜尋。
- (5) 知識再利用：經由標籤的重複使用，並衍生出新的知識。

社會性標記是利用使用者在網路上的互動，彼此互相銜接，形成一種社會網路，將議題延伸到真實生活互動中，促成彼此相互的理解，並透過一些管理工具，達到動態及參與式的知識管理，它的優點在於能夠適應使用者，而非強迫使用者改變。傳統的知識管理系統大多包含數種資訊技術，如：資料倉儲、資料採擷、知識庫、決策支援軟體、文件管理資料庫、群組軟體、

電子郵件等，必須結合數種科技，才能建構完整知識管理系統，不像社會性標記建構方便且使用上非常直覺，其與傳統的知識管理的主要差異如下：

- (1) 在知識的搜尋上，經由人們加註標籤後，搭配搜尋檢索功能，就能產生檢索效果；而知識管理則需要由系統建立資料倉儲或關聯資料庫，搭配網站後端管理才能啟動。
- (2) 知識管理是屬於規劃的性質，需要開發、建置、知識儲存、導入等階段，比較有知識歸納的特質，用在可族群化資訊與社群的分類；標籤是經由使用者瀏覽與經驗累積，採用共通的語句，來找出資訊的交集與共同點。
- (3) 在知識產生的過程與程序差異上，知識管理系統知識的產生與定義是由組織者與資訊發行者決定；標籤則是由使用者自行對數位資源加註標籤，透過知識描繪、獲取、儲存、分享而達到知識的連結。

6. 標籤之相關研究

標籤的研究是近年來一個新興的主題，包括：社會性標記、大眾分類、社會性書籤網站等，相關文獻大多著重於描述性的介紹。Seldow[31]將標籤的研究分為兩種類型：

- (1) 標籤趣味面：內容多為介紹社會性網路，傾向發揚社會性網路、頌揚社會性標記的優點，鼓勵使用者體驗及應用標籤。
- (2) 標籤實作面：內容以標籤資料庫的技術性實作、標籤群聚的演算規則、或探討標籤、使用者與資源的關係等。

本節將著重於標籤的實作面，並介紹部分與本研究相關的分群演算。

Millen等人[21]將社會性書籤網站Dogear的使用者設定為研究對象，以網站資源、標籤與使用者來觀察社會性書籤的特性，研究顯示使用者經常運用相似詞彙來標記資源，於2007年，Millen, Yang, Whittaker and Feinberg又繼續以Dogear為實驗對象，觀察三種搜尋行為：社群式瀏覽、個人書籤的搜尋、

關鍵字搜尋，觀察結果顯示社群式瀏覽使用頻率最為頻繁，證實社會性標記網站的的確有其應用價值[22]。

從社會性標記建立標籤間的關連最常用的方式就是分群法，Begelman 等人[1]是透過標籤與文件，將標籤間的關聯建立一標籤關係圖，然後利用遞迴執行分割演算法(Partition Algorithm)建立標籤分群；而Yu and Chien [42]則是採用Web-based clustering approach與Graphical approach的方法建立標籤分群。研究結果發現三點：(1)自動分群與手動分群的差距不大；(2)不同使用者分群的結果大不相同；(3)分群後的標籤能幫助使用者更快熟悉其搜尋概念。

Hsieh 等人[13]提出的是一個概念階層法(Concept Hierarchy)，是利用 Concept Space Generator(CSG)分析一群標籤(關鍵字)集合(Concept Space)，利用標籤與文件之間的關係，建立標籤階層的架構，將標籤的階層關係建立起來後，則標籤(關鍵字)之間的關係也建立起來，在本研究中，將運用這個觀念來建立標籤間的關係，以下是CSG's Algorithm的步驟：

- (1) 使用 tag vector 描述每一個 tag：對每一個 tag t_j ，用 vector T_j 代表 tag 與所有 resource 的關係。 $T(i)$ 代表一個 tag 標記了 resource i 多少次，例如：tag “programming” 被描述在三個 resource 上分別為 8、3、7 次，則 tag vector 為 $T_j = (8, 3, 7)$ 。
- (2) 將所有的 tag vectors 依照下列順序排序：
 1. 一個 tag 被描述在多少個 resource 的數量(resource count)。
 2. 在條件 1 中相同時，則依照 tag 被描述在所有 resource 的總使用次數(total use count)排序。
 3. 當條件 1 與條件 2 都相同時，則按照 tag 最早使用的時間排序。
- (3) 使用在步驟二排序好的結果建立 Concept Hierarchy：
 1. 已知一個 tag A，判斷是否任何一 tag B 所描述的 resource set(R_B) 滿足公式 2-1

$$\frac{|R_A \cap R_B|}{|R_A|} > \lambda \quad \text{公式 2-1}$$

2. λ 是一個臨界值，代表 tag A 與 tag B 之概念距離(Concept Distance)的下界，其介於 0 到 1 之間， λ 愈接近 1 時，則兩者關係愈強烈，愈靠近 0 時，兩者愈無關係。當 tag A 與 tag B 滿足公式 2-1 時，則 tag B 為 tag A 的父節點，概念距離則是利用 Cosine Similarity 計算其相似度。

三、 中文斷詞的相關研究

中研院的中文斷詞系統(CKIP)已廣泛運用於各個領域中[58]，只要研究涉及中文字的處理多會利用此工具或是與其效能做比較。本研究藉由中文斷詞的可靠性運用於數位典藏網站中。以下將介紹中文斷詞、目前的困境及方法與未知詞的處理。

1. 中文斷詞簡介

詞是最小有意義且可以自由使用的語言單位。任何語言處理的系統都必須先能分辨文本中的詞才能進行進一步的處理，例如機器翻譯、語言分析、語言了解、資訊抽取。因此中文自動斷詞的工作成了語言處理不可或缺的技術。基本上自動斷詞多利用詞典中收錄的詞和文本做比對，找出可能包含的詞，由於在中文裡，當中斷的地方不同時，就會產生不同意義的字詞，稱為「斷詞歧義性」[51]，因此多數的中文斷詞程式多討論如何解決斷詞歧義的問題，而較少討論如何處理詞典中未收錄的詞出現的問題(新詞如何辨認)。

中文詞集是一個開放集合，任何一個詞典或方法都無法盡列所有的中文詞。當處理不同領域的文件時，領域相關的特殊詞彙或專有名詞，常常造成參考詞彙的不足而產生錯誤的斷詞。為了解決這個問題，最有效的方法是補充領域詞典並加強詞彙的搜集。因此新的詞彙或關鍵詞的抽取成為斷詞的前置作業。領域相關的詞彙多出現在該領域的文件中而少出現在其它領域，因此抽取關鍵詞時可多利用此特性[45]。

2. 中文斷詞的困境及方法

中文與英文的不同在於其基本單位並非為每一個「字」，而是「詞」，英文在擷取「詞」時，可以利用空白將之拆解為最小單位，但做中文字的語句分析時每一個「字」並無法代表最小的基本單位。因為當「字」的位置不同時，字義也就不同，並沒有明顯的區隔。

Chen和Ma曾提出的中文詞擷取的困難度在於[3]：

1. 中文字詞沒有空格可以當作詞的邊界。
2. 大部份的中文字與詞都已是語言中最小的單位。
3. 字詞的語法與字義是多元意義且複雜不清。
4. 相同語法型態的字，可能會有不同的語義種類。
5. 沒有一個簡單的規則可以涵蓋各種型態的未知詞。
6. 很難直接從線上來辨識短文的未知詞，因為出現的未知詞頻率太低以致於無法用統計方法來達成。

而中文斷詞對各種的應用都有很深的影響，例如：文件檢索、中文輸入、光學字體辨識、語音辨識、機器翻譯等，都必須經過中文斷詞的處理，其準確度的優劣也影響著後續的效能，使得中文斷詞處理變成一種很重要的工作。

目前針對中文斷詞處理，已有許多研究，基本上可分為詞庫斷詞法[4]、法則式斷詞法[49]、統計式斷詞法[33]、混合式斷詞法[23]，及遺傳演算法式斷詞法[51]、N-Gram選詞法[58]等方法，以下僅作簡要介紹。

詞庫斷詞法是一種常用的斷詞方法，它是以詞庫內收集的語詞與文件進行比對，找到斷出的詞句，此法雖直覺易懂，但其準確度與收集詞彙資料的大小有關。

法則式斷詞法是將詞的構造原則運用到詞庫或辭典一起運作，再根據規則排除掉不可能的詞語組合，以達到較好的斷詞效果，代表性的法則式斷詞

法為「長詞優先法」(Maximum Matching Method or Longest Matching Method)[16][17]。

統計式斷詞法是利用語料庫來歸納字詞的方法，主要是依據機率分布來決定斷詞的位置，先計算計算字詞出現的頻率，再利用機率模型反覆求得各詞的機率分布，以找到最佳的字詞組合[33]。統計式斷詞法的優點是執行效率高，但只能處理兩字詞或單字詞，當詞的長度大於二時，效率會大幅降低，且斷詞正確率不高[23]。

N-Gram 選詞法則是將文件以二字詞(2-Gram, Bi-Gram)、三字詞(3-Gram, Tri-Gram)逐字進行分割，若斷出字詞之詞頻高於門檻值，則視為一個獨立字詞。

因為以上各種斷詞法皆有其優劣，因此有學者提出綜合的斷詞法，結合兩種以上的段詞法、例如先利用詞庫斷詞，再透過 N-Gram 找出剩下的新詞。

3. 未知詞擷取之研究探討

傳統中文斷詞系統碰到最大的瓶頸就是未知詞(辭典中沒有的詞)的擷取，統計一篇文章當中約有 3%~5%的詞彙是未知詞，新聞類的文章更是遠高於此。因此未知詞的擷取問題一直是中文語言處理上一個重要且困難的研究課題。

前人的研究主要是針對特定類型的詞彙(如人名、地名、組織名)以詞構分析或是語言模型(Language Model)擷取，可獲致不錯的成果。不過仍有許多未知詞的類型更加複雜且多變，例如像「泛藍」、「泛綠」、「拉法葉」、「矜島」、「非典(SARS)」等等。這些詞彙往往扮演了文章中的關鍵角色必須加以擷取辨認。

傳統的作法是利用語料庫的統計方法(如 co-occurrence、mutual information、dice、t-score)。其基本假設是認為那些統計值較高的樣式(pattern)，因為具有較強的統計特性，所以較有可能是一個未知詞。

本研究所引用的未知詞斷詞技術，為中央研究院詞庫小組所研究，已開放線上操作系統，目前的作法依序可分為以下幾個步驟[48]：

1. 初步斷詞：採取長詞優先演算法[4]，絕大部分的未知詞會被斷成較小的單位，此即未知詞的詞素。
2. 未知詞偵測：將斷出來的詞素重新組合成未知詞，透過未知詞偵測的步驟去判定哪些是詞素，哪些是獨用詞彙[2]，之後再判斷是否能和其相鄰的詞素合併成未知詞。
3. 中國人名、歐美譯名、複合詞擷取：針對一些特定的未知詞，如中國人名、歐美譯名、複合詞作詞構分析及簡單的語言模型將之擷取[3]。
4. 由上而下合併演算法(Bottom-up Merging Algorithm)：剩餘的未知詞交由此演算法做最後的擷取[18]，擷取出來的樣式稱為「未知詞候選者」。
5. 重新斷詞：將「未知詞候選者」搭配原始詞庫再做一次斷詞，即可得到最後結果。

四、 資訊檢索

資訊檢索的目的主要是從大量文件中自動抽取符合使用者需求的特定資訊，所以檢索技術就顯得非常重要，目前已有許多數學模型被應用在處理非結構化自然語言的文字上，如布林模型(Boolean Model)、機率模型(Probabilistic Model)、向量空間模型(Vector Space Model)等，而向量空間模型也成為資訊檢索系統中標準的工具，在面對在大量的文件集時，以關鍵字為擷取的方式仍然是最可行的方法[15]。

1. 關鍵字擷取

每一份文件皆是由詞所組成，文件中的每個詞在該文件中的重要性也會有不同，利用每個詞在該文件的出現頻率可以用來計算該詞在文件中的重要

性。

一個詞在一文件中出現的次數稱為詞頻(TF, Term Frequency)，文件頻率(DF, Document Frequency)是指一詞在多少篇文件中出現過，其倒數稱之為反轉文件頻率(IDF, Inverted Document Frequency)。IDF愈高代表該詞集中出現在少數文件裡，表示具有較高的獨特性，且對於這些文件也較具代表性[53]。

字詞權重計算最著名的方法是TFIDF(Term Frequency · Inverse Document Frequency，詞頻 · 反轉文件頻率)[29]，也就是用詞頻及反轉文件頻率來決定權重，就是將一個文件中的詞算出TF及IDF然後相乘，即可得到每個詞相對於該文件的權值，其原理是考量字詞對一篇文件的重要性多與其在該文件中出現之次數成正比，但與其出現在所有文件之次數成反比，公式如下：

$$w_{ij} = tf_{ij} \times \log\left(\frac{N}{df_i}\right) \quad \text{公式 2-2}$$

其中， w_{ij} ：表示字詞*i*在文件*j*的權重

tf_{ij} ：字詞*i*在文件*j*中的詞頻

df_i ：表示字詞*i*在整個文集中出現的文件頻率

N ：為整個文集的文件量。

如果將 tf_{ij} 作正規化，則公式如下：

$$tf_{ij} = \frac{freq_{ij}}{\max freq_{lj}} \quad \text{公式 2-3}$$

其中， $freq_{ij}$ ：字詞*i*在文件*j*中的詞類

$\max freq_{lj}$ ：字詞*l*是在文件*j*中出現頻率最多次的字詞

2. 空間向量模型

Salton於1975年提出向量空間模型(Vector Space Model)，在傳統資訊檢索領域中是一種最簡單、最具生產力的模型[29]。此模型主要以文件、段落或句子為單位，因此任何查詢語句與文件皆可以向量的形式來表示[27][28]。如公式 2-4、公式 2-5 所示：

$$D_i = (w_{i1}, w_{i2}, w_{i3}, \dots, w_{it}) \quad \text{公式 2-4}$$

$$Q = (w_{q1}, w_{q2}, w_{q3}, \dots, w_{qt}) \quad \text{公式 2-5}$$

係數 w_{it} 與 w_{qt} 分別表示在文件及查詢語句中詞彙的權重，當此詞彙不存在時，權重則設為零，表示詞彙權重的方式有許多種，最常見的是透過 TFIDF 字詞權重計算公式，則可得出 w_{it} 、 w_{qt} 的值。一旦得知每個詞彙在文件與查詢語句上的權重之後，便可以計算它們間的相似度，最常使用計算相似度技術為 Cosine Coefficient。其計算相似度如公式 2-6：

$$\text{sim}(Q, D_i) = \frac{\sum_{j=1}^t w_{qj} \times w_{ij}}{\sqrt{\sum_{j=1}^t w_{qj}^2 \times \sum_{j=1}^t w_{ij}^2}} \quad \text{公式 2-6}$$

五、 文件及關鍵字分群

資料分群(Data Clustering)的技術已經廣泛應用於各領域中，例如決策支援(Decision Making)、機械學習(Machine Learning)、資料探勘(Data Mining)、圖像辨識(Pattern Recognition)、文件檢索等[14]。

分群方法可分為幾種如：分割法(Partitioning Method)、階層法(Hierarchical Method)、密度基礎法(Density-based Method)、格子基礎(Grid-based Method)等[10]。以下對現有的分群法依序簡介：

1. 分割法(Partition Method)

分割法是一種對比資料，以找出相近資料的技術，最常用的為 K 平均法(K-mean)。假設有 n 個資料物件，欲分割成 k 群且 $k \leq n$ (k 代表目標分群個數，必須事先定義)，每一群都必須滿足以下二點：(1)每一群至少須包含一個物件；(2)每個物件只能屬於其中一群。

分群的原則是依各物件特徵劃分到相似的群集中，再藉由不斷修正各群集的特徵屬性，來調整群集的範圍。在分群的過程中，須確保兩個條件：(1)位於相同群內的物件，彼此間相似度高，此處的相似度定義為物件與群中心

點的歐基里德距離；(2) 位於不同群內的物件，彼此間相似度低，亦即屬於不同群的物件其歐基里德距離愈大愈好。詳細步驟如下：

1. 先從資料物件中任意選取 k 個物件作為群心，並以其為群中心點。
2. 重複以下步驟，直到群的分佈不再改變：
 - (1) 計算所有物件與 K 群中心點的距離，再將其物件分配給最近的群中心點。
 - (2) 重新計算該群的群中心點。

而另一種類似的分群法 k -medoids 是 k -means 的變形，它不使用群心，而是用所謂的代表物件為中心點(Representative Object，即 Medoid)。每個分群會有一個代表物件，它是該群中最接近中心而真實存在的一個物件。除此之外，分群方法的過程則和 k -means 類似。

2. 密度基礎法(Density-based Method)

密度基礎法是利用相鄰區域的觀念來發現群聚之方法，當相鄰區域內資料點的密度到達預先設定的門檻值時，便自動形成一個群聚。

密度基礎法較分割法具有去除或控制偏差值的能力。目前較常見的密度基礎法有 DBSCAN，OPTICS，DENCLUE。

3. 格子基礎法(Grid-based Method)

格子基礎法的概念是將包含資料點的資料空間，量化為許許多多格子(Grid Cells)的概念，藉由計算格子內的統計資訊，進行群聚的方法。其每一個格子內皆隱藏許多資訊，因而大量的減少群聚的時間。

每一個格子內皆包含其資料的統計資訊(如平均值、標準差、最大值與最小值)，因為在演算的過程無需面對所有資料點，所以能夠大量減少演算的時間，提高此類型群聚演算法效能，尤其擅長大量且高維度資料庫。

此分群法的優點就是處理速度很快，通常與資料點的個數無關的，只與有多少個資料空間有關。目前常見的格子基礎法有 STING、WaveCluster、

CLIQUE 演算法。

4. 階層分群法(Hierarchical Method)

階層式分群方法的目的是在為資料物件以階層分解的方法建立樹狀結構的分群。在過程中藉由訂定終止條件，當滿足終止條件時即停止程序。判斷停止程序的終止條件通常為達到目標分群數，或是經由合併或分化後，群與群之間的相似度達到門檻值等。階層式分群方法主要可分成凝聚式

(Agglomerative)與分裂式(Divisive)兩種方法：

1. 凝聚式(Agglomerative)：又可稱為由下而上(Bottom-up)的方法，先將資料中的每一個物件自成一分群，計算兩兩群的相似度，當相似度大於既定的臨界值時，則合併兩群為一個較大的群，直到所有的物件都屬於同一個群，或是符合終止條件才停止。
2. 分裂式(Divisive)：一種由上而下(Top-Down)的分群方式，先將所有項目視為同一群，再依照物件間的相似度分割為較小子群，直到每個子群都只有一個物件或是符合終止條件才停止。



參、 系統架構與實作

在本章中將介紹楊英風數位美術館建置社會性標記的技術細節，包含如何利用資訊檢索技術萃取關鍵詞，蒐集使用者於網站上標記的標籤，並將兩者結合形成大眾分類，以達成 Web 2.0 應用所能期待的效果。

第一節說明本研究的核心概念；第二節介紹系統架構及技術開發所需的環境及應用軟體；第三節說明系統的資料庫設計圖；第四節到第六節詳細說明本研究開發的核心技術，依序為中文斷詞處理、線上標籤搜集、階層標籤產生器(Tag Hierarchy Generator)；第七節將解釋產出結果的應用，包含標籤地圖、標籤推薦、搜尋標籤；第八節介紹系統操作畫面。

一、 核心概念

圖 16 為本研究之核心概念，主要有三個步驟，第一：透過一個線上標籤維護的介面，蒐集使用者提供的標籤；第二：除了使用者提供的標籤之外，系統會運用資訊檢索技術從線上美術館萃取出每一幅藝術作品的關鍵詞；第三：運用標籤階層的概念建立標籤地圖。

根據上一章所介紹的社會性標記，標記行為即為主要功能之一，而標籤成為使用者提供的資料索引，在本研究中所萃取的系統關鍵詞擴充了網站資源的詮釋資料，結合兩者以期能達到知識整合的效果。圖 16 為本研究的主要核心概念，由以下各點分別說明：

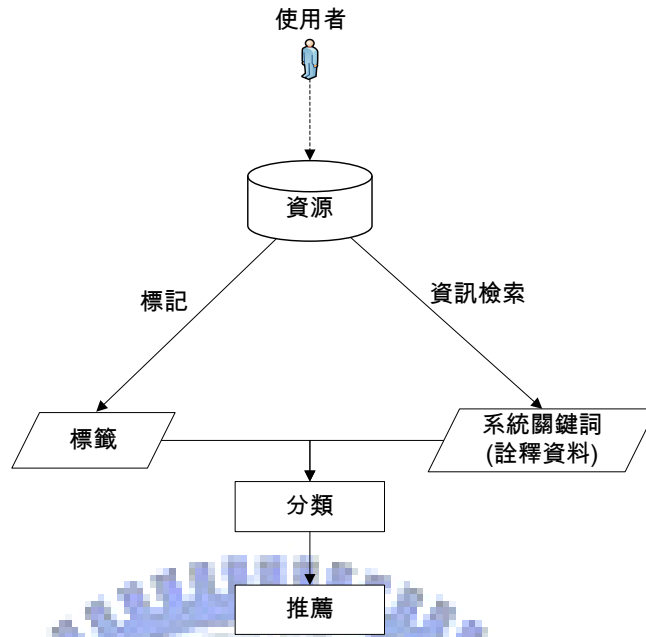


圖 16：本研究之核心概念

1. 標籤：代表使用者的想法，而標記是使用者加註標籤的行為，所以標籤實際上是一種意念的表徵，就意義上來說，是將外界的輸入轉化為資訊的結果。
2. 系統關鍵詞：由本系統擷取每個藝術作品的作品名稱、副標題、體裁、材質、技法、摘要等欄位，利用資訊檢索技術計算每個詞彙的權重，再根據萃取權重較大者做為系統關鍵詞。其目的是當作推薦資訊的基礎。當使用者看到系統關鍵詞，可能會聯想到其他相似意涵的標籤，所以它扮演的是激發大眾靈感的輔助角色。在本研究中，使用者除了可提供標籤外，也會透過兩種方式影響系統關鍵詞的權重，第一是資料搜尋，第二是標記行為，詳細的作法將會在後續介紹。
3. 分類：由標記資源所形成的分類稱為大眾分類，相對於傳統專家分類則是一種自由的分類方法。在本系統中，經由使用者不斷地標記及調整系統關鍵詞權重，分類的呈現會愈來愈完整，且不同於傳統的專家分類。
4. 使用者：在本研究是最核心的角色，所有的概念皆由使用者角度出發，標籤必須經由標記行為產生，系統關鍵詞也會透過使用者影響其權重。

所謂 Web 2.0 就是希望讓使用者與使用者、及使用者與系統間互動，讓網路資源更豐富地呈現。

5. 推薦：使用標籤地圖提供推薦詞。標籤在地圖內為節點，節點與節點中間的連線，即為兩者同時擁有的資訊內容。

本研究是改良Hsieh等人[13]的研究，應用於楊英風數位美術館上，與前研究的差異在於本研究將數位內容用中文斷詞擷取出關鍵詞，而非只是單純的讓使用者提供標籤；再者，本研究結合關鍵詞與標籤，共同建立標籤階層結構，並改良了Concept Space Generator Algorithm(CSG's Algorithm)，強調標籤與標籤之間的共用性，而不只是用相似度來判斷兩者之間的關係。

二、系統架構

本研究之系統架構如圖 17 所示。使用者進入楊英風數位美術館即可對站內美術作品進行標記。

系統模型於第一階段會將站內資料進行前置作業，針對每個藝術作品先利用 CKIP 中文斷詞系統將文字加以處理，再經過權重計算擷取其關鍵詞；第二階段則蒐集使用者於網站上所提供的標籤及搜尋資料時使用的關鍵字，並調整系統關鍵詞的權重；第三階段為系統關鍵詞與標籤的合併，然後運用階層式標籤分群法將標籤及關鍵詞建構為階層式標籤；第四階段將產生的階層式標籤分群以標籤地圖呈現，它建立標籤/關鍵詞之間的連結，透過其間的連結，對每個藝術作品提供相關的標籤推薦，稱為推薦詞，經由此推薦詞可搜尋到其他相關的藝術作品。在後續小節將針對系統的每一個模組詳細說明。

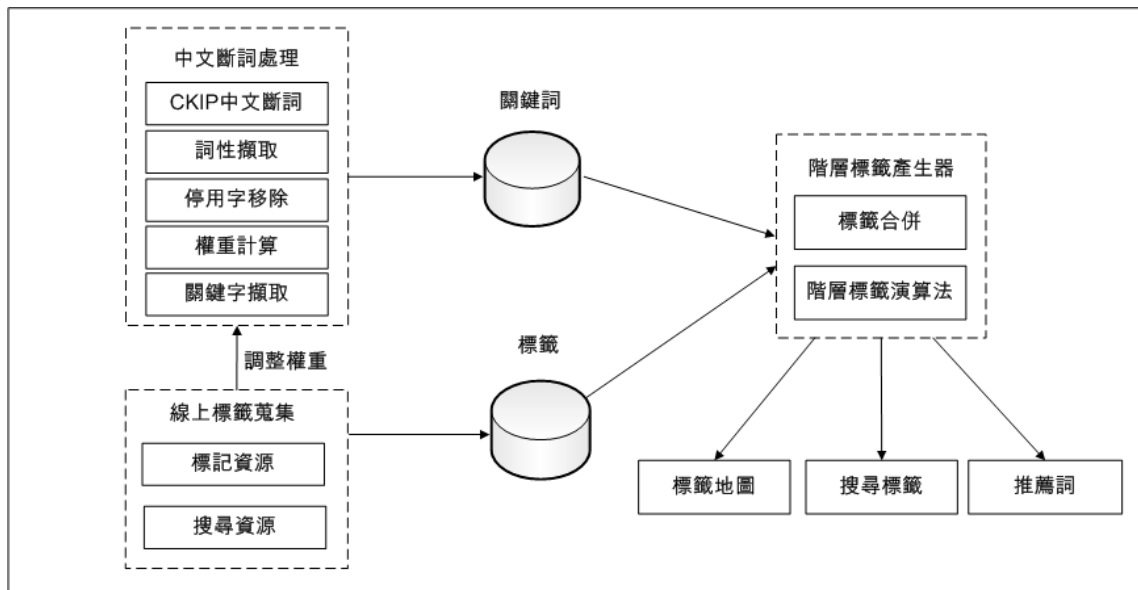


圖 17：本研究之系統架構

楊英風數位美術館之標記功能開發環境如表 2。它是建構在Microsoft IIS 6 之上，原始網站程式語言採用ASP，而本研究後端標籤資料彙整是由Java語言撰寫，包含中文斷詞、詞性擷取、權重計算、關鍵字擷取、標籤合併及階層式標籤分群法，採用JDK 1.5.0_11-b03 版本開發，階層分群演算因含大量資料，並於每日執行資料更新，故為批次處理，現階段於每天晚上 11:00 執行。

表 2：楊英風數位美術館開發軟體需求

功能	軟體名稱
Web Server	Microsoft IIS 6 Microsoft Windows 2003 Standard SP2
Development Tools	ASP Java SE Development Kit(JDK) 1.5.0_11-b03 Borland JBuilder 2006 Enterprise UltraEdit
Database Server	Microsoft SQL Server 2005 Enterprise SP1
ODBC Driver	Microsoft ODBC

三、 資料庫設計

本研究以Microsoft SQL Server 2005 Enterprise作為資料庫伺服器，在資料庫中主要設計四個表格作為完整標記功能資料的存放，分別為系統關鍵詞 (CKIPTAG Table)、標籤資訊(TAG Table)、系統關鍵詞與標籤合併表(TAGMERGE Table)及標籤階層表(TAGHIERARCHY Table)。此外，artWork Table則是存放所

有藝術作品的原始資料，也是本研究的資料來源，它與CKIPTAG Table和Tag Table呈現一對多的關係，資料模型如圖 18所示。

CKIPTAG Table 儲存資料前置處理後所產生的資訊，經 CKIP 中文斷詞系統作初步斷詞後，篩選所需要的詞性存放於該 Table 內，再由公式計算每個字詞的權重，最後會根據權重取出適當的關鍵詞，呈現於藝術作品的頁面上。每筆資料皆有記錄編號(ckiptag_id)、藝術作品編號(no)、所屬類別(workstype)、關鍵詞(keyword)、詞頻(tf)、反轉文件頻率(idf)、權重(weight)、建立日期(dates)。

TAG Table 是使用者為某一藝術作品加入標籤時，就會建立一筆標籤與作品的關聯，每一個藝術作品皆可給予多個標籤，因此每筆資料皆有記錄編號(tag_id)、藝術作品編號(no)、所屬類別(workstype)、標籤名稱(tag)、使用次數(count)、建立日期(dates)。

TAGMERGE Table 是由關鍵詞與標籤合併之後所存放的 Table，以利後續進行階層式標籤分群法，目前標籤/關鍵詞的合併是以藝術作品的類別來合併，也就是說在同一分類下，不會有重複的標籤/關鍵詞，合併之後每個標籤/關鍵詞會有記錄編號(tagmerge_id)、所屬類別(workstype)、標籤名稱(tag)、資源使用次數(resourcecnt)、總使用次數(totalcnt)、最早建立日期(dates)。

TAGHIERARCHY Table 是將關鍵詞/標籤合併之後，由階層式標籤分群法計算的結果，每次計算時皆會刪除 Table 內所有的資料並重新建立標籤/關鍵詞之間的關聯，每個標籤/關鍵詞皆會有記錄編號(taghierarchy_id)、所屬類別(workstype)、標籤名稱(tag)、標籤編號(id)、群組編號(groupid)、父節點編號(parentid)、與父節點相似度(similarity)，階層式標籤分群法的計算是依據各個類別的藝術作品分別建立其關聯。

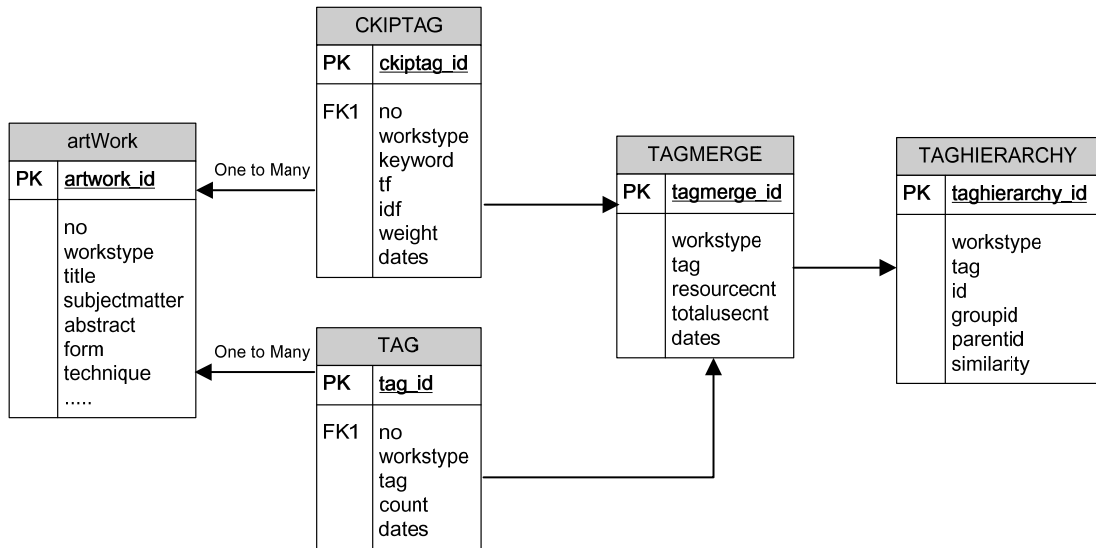


圖 18：資料庫設計圖

四、中文斷詞處理

本系統在此階段會將資料作前置處理，也就是當沒有任何人提供標籤時，系統自動內建一些關鍵詞，供使用者加註標籤時參考。其目的有二：第一是當後續合併彙整之後可以產生更多標籤/關鍵詞的推薦，以便發掘出更多潛藏於系統的資源；第二是將內建關鍵詞放在每個藝術作品的網頁上，可以快速告訴使用者這個作品的重點為何及原始作者的想法，激發使用者聯想更多其他的標籤。

本研究採取的是中研院詞庫小組的中文斷詞系統，圖 19 是中文斷詞的處理步驟，依序說明如下：

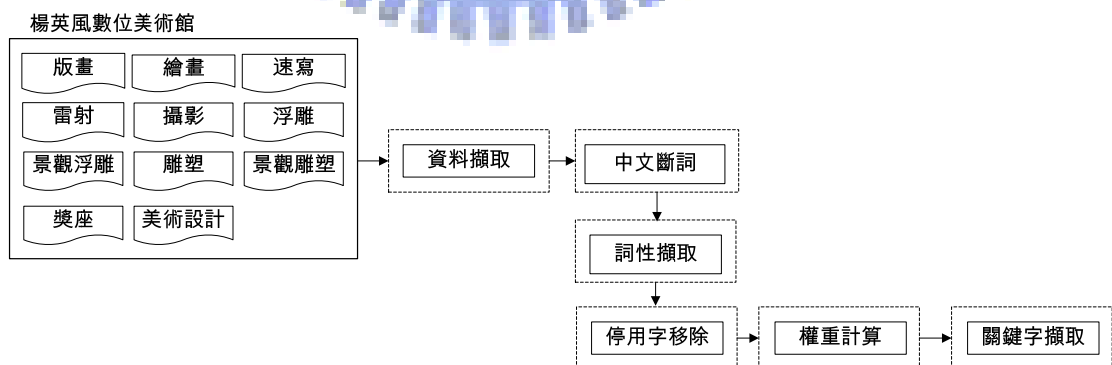


圖 19：中文斷詞處理步驟

1. 資料擷取

楊英風美術作品總共有 11 個類別，如圖 19，在做資料採擷時是將每一

類別分別來作，因在不同類別的美術作品，其所呈現的體裁、技法、描述、內容都不太一樣，所擷取出來的關鍵詞也不盡相同。表 3 為每一個藝術作品在網頁上所呈現出來的資料列表，也是本研究需要斷詞的資料，分別為作品名稱、副標題、體裁、材質、技法、摘要，選擇這些欄位的資料斷詞是因為其內容能確切描述此美術作品，也代表其特徵。

表 3：美術作品擷取資料欄位表 (資料來源：<http://yuyuyang.e-lib.nctu.edu.tw/>)

欄位	範例
作品名稱	茁生
副標題	抽象不鏽鋼雕塑
體裁	景觀雕塑
材質	不鏽鋼
技法	
摘要	〔茁生〕顧名思義乃為了表現萬物順應自然成長、茁壯的生息過程，是揚英風尊重自然及強調宇宙和諧平衡美學觀的再一次具現，作品由各組件組成，最大者至最小者的整體排列、組合方式，乃象意了宇宙自然的循環不息與無盡生機。.....

2. 中文斷詞

本研究是透過CKIP中文斷詞系統作文字的斷詞，中研院詞庫小組開放一個API供用戶端呼叫，資料的交換方式採用 XML，用戶端自行撰寫程式經由TCP Socket 連線傳送驗證資訊及文本至伺服器端，伺服器處理後經由原連線傳回結果。用戶端傳送資料之 XML 格式如表 4：

表 4：CKIP用戶端傳送之XML格式[46]

```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<option showcategory="1" />
<authentication username="XYZ" password="ABC" />
<text>台新金控 12 月 3 日將召開股東臨時會進行董監改選。</text>
```

username 及 password 為用戶端所申請之帳號及密碼，<text> 區段內包含文本，文本不應包含會使 XML 格式錯誤之字元如 "<" 及 ">"，此等字元需加以適當處理(例如轉換為全形符號)，以免造成處理錯誤的結果。文本不需預先進行斷句，伺服器會自行判斷句子的邊界。

CKIP的處理結果亦由伺服器以XML格式回傳，如表 5：

表 5：CKIP伺服器傳回XML格式的處理結果[46]

```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<processstatus code="0">Success</processstatus>
<result>
  <sentence> 台新(N) 金控(N) 1 2月(N) 3日(N) 將(ADV) 召開
(Vt) 股東(N) 臨時會(N) 進行(Vt) 董監(N) 改選(Vt) ，
(COMMACATEGORY) </sentence>
</result>
</wordsegmentation>
```

其中 <processstatus> 區段為處理的結果，除了 code=0 表示成功外，另外有三種可能的處理結果如表 6：

表 6：CKIP回傳另外三種XML格式的處理結果[46]

```
<processstatus code="1"> Service internal error</processstatus>
<processstatus code="2">XML format error</processstatus>
<processstatus code="3">Authentication failed</processstatus>
```

code=1 表示伺服器內部發生錯誤，可能是由不預期的字元或是過於複雜的句子結構所造成；code=2 表示接收到的XML格式有錯誤；code=3 表示帳號或密碼錯誤。斷詞過的所有句子包含在 <result> 區段內，每一句儲存於 <sentence> 區段中，每個詞以全形空白隔開[46]。

3. 詞性擷取

經過斷詞後，並非所有斷出來的詞都需要擷取，本研究觀察楊英風網站的美術作品，其中能表示美術作品特徵者大多為名詞及某些動詞。表 7整理出需要截取的詞性。如果斷出來字詞的詞性屬於精簡詞類裡的N及Vi，則不管在哪個欄位一律全部擷取，如果詞性屬於精簡詞類的Vt時，則只擷取作品名稱欄位，因Vt包含多種不同的動詞，若是全部欄位都擷取，則會選到太多不必要的字詞，故只有在作品名稱有包含Vt詞性的才會擷取出來。

表 7：中文斷詞詞類擷取列表(整理自附錄一)

化詞類	精簡詞類	意義	擷取欄位
Na	N	普通名詞	全部
Nb	N	專有名詞	全部
Nc	N	地方詞	全部
Ncd	N	位置詞	全部
Nd	N	時間詞	全部
VA	Vi	動作不及物動詞	全部
VB	Vi	動作類及物動詞	全部
VH	Vi	狀態不及物動詞	全部
VI	Vi	狀態類及物動詞	全部
SHI	Vt	是	作品名稱
VAC	Vt	動作使動動詞	作品名稱
VC	Vt	動作及物動詞	作品名稱
VCL	Vt	動作接地方賓語動詞	作品名稱
VD	Vt	雙賓動詞	作品名稱
VE	Vt	動作句賓動詞	作品名稱
VF	Vt	動作謂賓動詞	作品名稱
VG	Vt	分類動詞	作品名稱
VHC	Vt	狀態使動動詞	作品名稱
VJ	Vt	狀態及物動詞	作品名稱
VK	Vt	狀態句賓動詞	作品名稱
VL	Vt	狀態謂賓動詞	作品名稱
V_2	Vt	有	作品名稱

4. 停用字移除

停用字指的是在一群文集當中，出現太多頻率的字詞，因而無法具有識別性，在英文中如 the、of、and，移除停用字可以提升檢索效率及降低擷取不必要的關鍵詞。停用字的作法通常會採取停用字清單(Stoplist)來過濾不要的字詞。在本研究中是採用中文停用字清單來移除不要的中文字，如：你們、但是、且、或等一些連接詞、語助詞及代名詞。

5. 權重計算

本研究的字詞權重計算是沿用 TFIDF 的計算公式，將公式內容作修改以符合於本研究之應用，如公式 4-1，說明如下：

$$tf_{ijk} = \frac{freq_{ijk}}{\max freq_{lk}} \quad \text{公式4-1}$$

其中， $freq_{ijk}$ ：在所屬類別 k 中，字詞 i 在美術作品 j 中的詞頻

$\max freq_{lk}$ ：字詞 l 是在所屬類別 k 中，出現頻率最多次的字詞

$k = 1, 2, \dots, 11$ ，表示美術作品類別版畫、繪畫、雕塑...等等

$$w_{ijk} = tf_{ijk} \times \log\left(\frac{N_k}{df_{ik}}\right) \quad \text{公式 4-2}$$

其中， w_{ijk} ：在所屬類別 k 中，字詞 i 在美術作品 j 的權重

tf_{ijk} ：在所屬類別 k 中，字詞 i 在美術作品 j 經過正規化的詞頻

df_{ik} ：表示字詞 i 在類別 k 中，所出現的美術作品頻率

N_k ：在類別 k 中之美術作品數量

6. 關鍵詞擷取

關鍵詞是根據不同類別來擷取，在不同類別裡可能會有相同的關鍵詞，但所計算的權重卻會不一樣，最後產生的標籤地圖其關聯的標籤也會有所不同。關鍵詞的擷取如以下三個步驟：

- (1) 根據每一類別，將所有字詞的權重由大排到小。
- (2) 找出每一類別字詞權重的中位數。
- (3) 取出每一類別字詞權重大於中位數以上的值作為該類別之關鍵詞。

經由擷取權重大於中位數的字詞，在網頁上不同類別的每一個美術作品中，就會呈現屬於自己的關鍵詞。

五、 線上標籤蒐集

楊英風數位美術館是一個開放的網站，將楊英風藝術作品放在網路上讓民眾觀賞，本系統提供使用者在瀏覽藝術作品的同時可於線上加註標籤，同時系統開始啟動標籤蒐集的功能。線上標籤蒐集主要有二個來源：第一部分為標記資源，如圖 20所示，即是讓使用者自行對站內藝術作品給予適當的標籤；第二部分為搜尋資源，如圖 21所示，其概念是當使用者利用站內檢索功能搜尋資料時，經由輸入關鍵字所搜尋出來的文件，即表達了資料與關鍵字的關係，在此將這種利用輸入的關鍵字搭配搜尋結果視為另一種標籤輸入的方式。

並非所有用來搜尋的關鍵字都可以成為標籤，所以必須經過搜尋結果的檢

驗，以防止產生沒有意義的標籤。如圖 21，當輸入關鍵字所搜尋出來的藝術作品符合某些規則，則會將此關鍵字加入到搜尋結果中的每個藝術作品。於本研究定義搜尋結果產生標籤的規則如下：

1. 當搜尋的藝術作品數量等於零時，代表搜尋關鍵字沒有跟任何作品有關，不會加入任一作品成為標籤。
2. 當搜尋的數量大於所有藝術作品的二分之一，則代表此關鍵字過於普遍，不具有獨特性，也不會加入到所搜尋的藝術作品中。



圖 20：楊英風網站標記資源範例圖



圖 21：楊英風網站搜尋資源範例圖

另外，標籤蒐集的另一個功能是會隱性調整系統斷詞詞庫的權重，系統斷詞詞庫為針對所有藝術作品的作品名稱等欄位(詳表 3)加以斷詞後所產生的詞彙，當使用者透過標記資源與搜尋資源兩種提供標籤的方式時，如果所給予的標籤與系統斷詞詞庫相同時，則會主動調整權重，目前針對使用者給予的權重是將原來權值加 0.5，當重新擷取關鍵詞時，會因為使用者調整了某些字詞的權重，而擷取到新的關鍵詞，如圖 22顯示本研究關鍵詞的擷取流程。

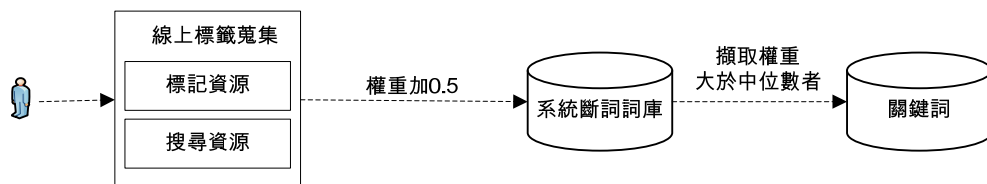


圖 22：本研究之關鍵詞擷取流程

六、 階層標籤產生器 (Tag Hierarchy Generator)

如圖 17 所示，楊英風數位美術館提供了一個標記系統的架構，當使用者瀏覽站內資源並給予喜好的標籤後，系統會將標籤與關鍵詞合併，然後利用階層式標籤分群法建立標籤的階層結構。階層式標籤分群法是利用標籤與數位資源間的共用性來產生標籤間的從屬關係，並建構出概念分群的效果。產生了階層式標籤分群之後，即可利用它做為後續的應用。在本節中根據每一個作品類別，分別產生階層式標籤分群，故不同的類別就有不同的分群結果，並將其結果產生標籤地圖及推薦詞。

(1) 標籤關鍵詞合併

在同一個美術作品類別裡，會有兩種資料，一個是關鍵詞資料，一個是標籤資料，關鍵詞是擷取大於字詞權重中位數者，標籤則是使用者自訂的，合併的方式是在同一類美術作品中，當標籤與關鍵詞的字詞一樣時，則將關鍵詞所出現在整個類別的詞頻與標籤的使用次數相加，儲存於標籤關鍵詞合併資料表(TAGMERGE)，若是兩者不同時，則可直接合併，表 8 是一個簡略的合併範例。

表 8：關鍵詞標籤合併範例表

CKIPTAG	
Tag	TF
海鷗	2
海浪	1
飛翔	3

TAG	
Tag	count
海鷗	1
凌越	2

+

↓

TAGMERGE	
Tag	count
海鷗	3
海浪	1
飛翔	3
凌越	2

(2) 階層式標籤分群法 (Hierarchical Tag Clustering)

合併關鍵詞與標籤的下一步驟為執行階層式標籤分群法，詳細步驟如下所述：

圖 23 為數位資源加註標籤的範例圖示，以方便後續說明。文字**速寫**、**鋼筆**、**鉛筆**、**水墨**代表儲存於資料庫的標籤(或關鍵詞)，而標籤指向某數位資源(Resource 1 – Resource 3)則代表使用者以此標籤為該數位資源進行標記，旁邊的數字則是以該標籤標記數位資源的次數。

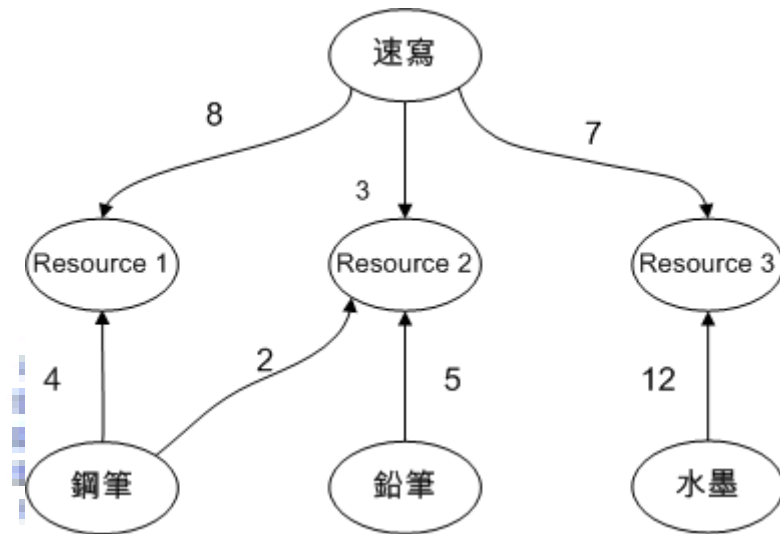


圖 23：標籤與數位資源的關係圖

1. 將儲存於標籤資料庫的標籤用向量來表示，標籤**速寫**的向量是(8,3,7)，代表對數位資源 1-3 分別標記的次數為 8 次、3 次、7 次，結果如表 9 所示。

表 9：標籤向量表

標籤	向量
速寫	(8,3,7)
鋼筆	(4,2,0)
水墨	(0,0,12)
鉛筆	(0,5,0)

2. 計算每一個標籤被標記的資源數量(Resource count)以及總標記次數 (Total use count)。

表 10：各標籤被標記的資源數量與總標記次數

排序	Tags	向量	被標記資源數量	總標記次數
1	速寫	(8,3,7)	3	18
2	鋼筆	(4,2,0)	2	6
3	水墨	(0,0,12)	1	12
4	鉛筆	(0,5,0)	1	5

3. 將標籤及標籤向量根據以下順位排序，結果如表 10所示：
 - (1) 將標籤及標籤向量依照被標記資源數量的值由高到低排列。
 - (2) 若兩標籤被標記的資源數量相同，則依總標記次數由高到低排列。
 - (3) 若前二者皆相同，則依照標籤最早使用的時間來排序。

4. 利用表 10排序好的標籤向量建立階層式標籤分群：

- (1) 若標籤資料庫中有 N 個標籤，則下述步驟共需執行 N-1 輪。
- (2) 在第 i 輪中，以表 10排序第 i 位的標籤 T_i 為主，分別計算 $R(T_i | T_j)$ 及 $R(T_j | T_i)$ ，以釐清 T_i 與排序其後之其他標籤 $T_j (j > i)$ 間的關係。
- (3) 假設任兩標籤 T_A 、 T_B ：

$$P_1 = (T_B | T_A) = \frac{R(T_A \cap T_B)}{R(T_A)} \quad \text{公式 4-3}$$

$$P_2 = (T_A | T_B) = \frac{R(T_A \cap T_B)}{R(T_B)} \quad \text{公式 4-4}$$

P_1 的計算方式如公式 4-3 所示，其中 $R(T_A)$ 表示標籤 T_A 被標記的資源數量， $R(T_A \cap T_B)$ 表示標籤 T_A 和 T_B 所共同標記的資源數量，依公式 4-3 計算標籤 T_A 被標註時標籤 T_B 亦被標註的機率 (P_1 值介於 0 到 1 之間)，如果 P_1 值愈接近 1，對 T_A 來說，代表 T_A 與 T_B 的共用程度愈強；若是愈接近 0，則對 T_A 來說，代表 T_A 與 T_B 的共用程度愈弱。同理， P_2 的計算方式如公式 4-4 所示，如果 P_2 值愈接近 1，對 T_B 來說，代表 T_A 與 T_B 的共用程度愈強；若是愈接近 0，則對 T_B 來說，代表 T_A 與 T_B 的共用程度愈弱。

- (4) 若是 $P_2 \geq P_1$ ，表示 T_A 標註的資源數量比 T_B 多或相同，也就是

$R(T_A) \geq R(T_B)$ ，故 $P_2 \geq P_1$ 。若 $P_2 \geq P_1$ 且 P_2 則滿足公式 4-5，則將 T_A 設為 T_B 之父節點。

$$P_2 = R(T_A | T_B) = \frac{R(T_A \cap T_B)}{R(T_B)} > \lambda \quad \text{公式 4-5}$$

(5) 若 P_2 值為 1 時，則代表以 T_A 所標註的資源包含所有 T_B 標註的資源；反之，如果該值為 0 時，則代表兩個標籤之間沒有關聯。

(6) 利用 Consine Similarity 計算兩標籤 T_i 和 T_j 的相似度，如公式 4-6、4-7 所述。

$$\begin{aligned} T_i &= (w_{i1}, w_{i2}, \dots, w_{it}) \\ T_j &= (w_{j1}, w_{j2}, \dots, w_{jt}) \end{aligned} \quad \text{公式 4-6}$$

$$\text{sim}(T_i, T_j) = \frac{\sum_{q=1}^t w_{iq} \times w_{jq}}{\sqrt{\sum_{q=1}^t w_{iq}^2 \times \sum_{q=1}^t w_{jq}^2}} \quad \text{公式 4-7}$$

5. 將表 10 根據前述方法進行階層式標籤分群的過程，並假設臨界值 λ 為 0.5：

第一輪運算：依以下步驟運算後，可得出如圖 26 的結果。

(1) 計算鋼筆與速寫的關係：依表 11 計算鋼筆與速寫間的關係，在此步驟可決定速寫為鋼筆之父節點，二者的相似度為 0.78，可得出如圖 24。

表 11：計算鋼筆與速寫間的相互關係

$$P_1 = R(T_{\text{鋼筆}} | T_{\text{速寫}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{鋼筆}})}{R(T_{\text{速寫}})} = \frac{2}{3} = 0.67$$

$$P_2 = R(T_{\text{速寫}} | T_{\text{鋼筆}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{鋼筆}})}{R(T_{\text{鋼筆}})} = \frac{2}{2} = 1$$

$$\therefore P_2 \geq P_1, \text{ 且 } P_2 > 0.5$$

\therefore 速寫為鋼筆之父節點

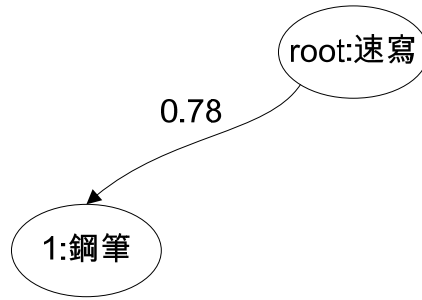


圖 24：第一輪階層分群過程 1

(2) 計算水墨與速寫的關係：依表 12 計算水墨與速寫間的關係，在此步驟可決定速寫為水墨的父節點，二者的相似度為 0.63，可得出如圖 25。

表 12：計算水墨與速寫間的相互關係

$$P_1 = R(T_{\text{水墨}} | T_{\text{速寫}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{水墨}})}{R(T_{\text{速寫}})} = \frac{1}{3} = 0.33$$

$$P_2 = R(T_{\text{速寫}} | T_{\text{水墨}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{水墨}})}{R(T_{\text{水墨}})} = \frac{1}{1} = 1$$

$\therefore P_2 \geq P_1$ ，且 $P_2 > 0.5$

\therefore 速寫為水墨的父節點

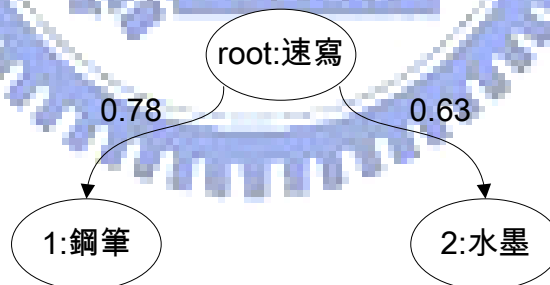


圖 25：第一輪階層分群過程 2

(3) 計算鉛筆與速寫的關係：依表 13 計算鉛筆與速寫間的關係，在此步驟可決定速寫為鉛筆的父節點，二者的相似度為 0.27，可得出如圖 26。

表 13：計算鉛筆與速寫間的相互關係

$$P_1 = R(T_{\text{鉛筆}} | T_{\text{速寫}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{鉛筆}})}{R(T_{\text{速寫}})} = \frac{1}{3} = 0.33$$

$$P_2 = R(T_{\text{速寫}} | T_{\text{鉛筆}}) = \frac{R(T_{\text{速寫}} \cap T_{\text{鉛筆}})}{R(T_{\text{鉛筆}})} = \frac{1}{1} = 1$$

$\therefore P_2 \geq P_1$ ，且 $P_2 > 0.5$

\therefore 速寫為鉛筆的父節點

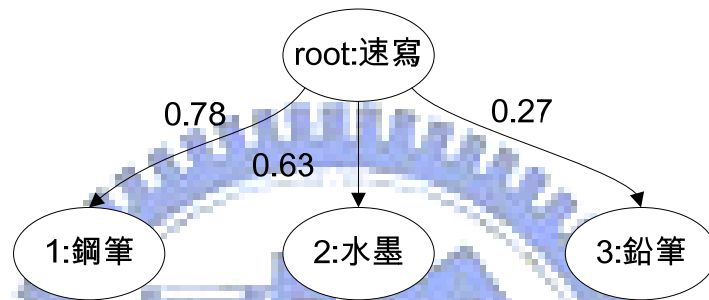


圖 26：第一輪階層式標籤分群的結果

第二輪運算：依以下步驟運算後，可得出如圖 27 的結果。

- (1) 計算水墨與鋼筆的關係：依表 14 計算水墨與鋼筆間的關係，在此步驟可決定二者無關聯。

表 14：計算水墨與鋼筆間的相互關係

$$P_1 = R(T_{\text{水墨}} | T_{\text{鋼筆}}) = \frac{R(T_{\text{鋼筆}} \cap T_{\text{水墨}})}{R(T_{\text{鋼筆}})} = \frac{0}{2} = 0$$

$$P_2 = R(T_{\text{鋼筆}} | T_{\text{水墨}}) = \frac{R(T_{\text{鋼筆}} \cap T_{\text{水墨}})}{R(T_{\text{水墨}})} = \frac{0}{1} = 0$$

- (2) 計算鉛筆與鋼筆的關係：依表 15 計算鉛筆與速寫間的關係，在此步驟可決定鋼筆為鉛筆的父節點，二者的相似度為 0.45。

表 15：計算鉛筆與鋼筆間的相互關係

$$P_1 = R(T_{鉛筆} | T_{鋼筆}) = \frac{R(T_{鋼筆} \cap T_{鉛筆})}{R(T_{鋼筆})} = \frac{1}{2} = 0.5$$

$$P_2 = R(T_{鋼筆} | T_{鉛筆}) = \frac{R(T_{鋼筆} \cap T_{鉛筆})}{R(T_{鉛筆})} = \frac{1}{1} = 1$$

$\therefore P_2 \geq P_1$ ，且 $P_2 > 0.5$ ，又相似度為 0.45 大於鉛筆與速寫的相似度 0.27

\therefore 鉛筆的父節點速寫被鋼筆取代，故鋼筆為鉛筆新的父節點

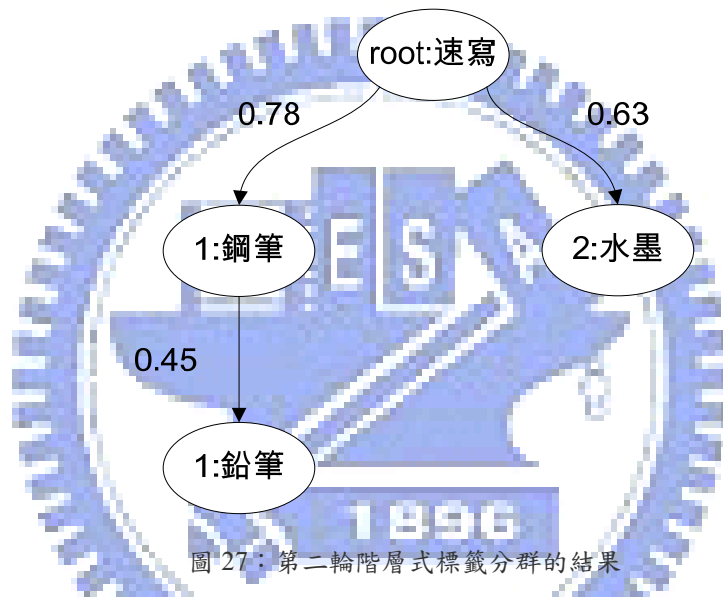


圖 27：第二輪階層式標籤分群的結果

第三輪運算：依以下步驟運算後，可發現圖 27 的分群結果不產生變動。

- (1) 計算水墨與鉛筆的關係：依表 16 計算水墨與鉛筆間的關係，在此步驟可決定二者無關聯。

表 16：計算鉛筆與水墨間的相互關係

$$P = R(T_{鉛筆} | T_{水墨}) = \frac{R(T_{水墨} \cap T_{鉛筆})}{R(T_{水墨})} = \frac{0}{1} = 0$$

$$P = R(T_{水墨} | T_{鉛筆}) = \frac{R(T_{水墨} \cap T_{鉛筆})}{R(T_{鉛筆})} = \frac{0}{1} = 0$$

臨界值 λ 的設定將影響階層式分群的結果，亦即當 λ 值愈高，表示標籤之間的共用性須愈高才能群聚在一起，此時所產生出來的群組會愈多，而每

一群的標籤數量會較少；反之，當 λ 值愈低，表示標籤之間不用太多的共用性就可以群聚在一起，故所分出的群組就愈少，而每一群內的標籤數量則較多。

以本研究階層式標籤分群的實際的例子來看，圖及圖 29 分別顯示 λ 為 0.5 及 0.8 時的階層式標籤分群，可以看出當 λ 設定為 0.5 時，在廟會這個標籤分群下，明顯比 λ 為 0.8 時的分群結構還要龐大。另外，相同的標籤數量作階層式分群時，當每一群組內的標籤數量愈多，即代表全部產生的群組數會較少；反之亦然。



圖 28： λ 為 0.5 時的階層式標籤分群結構

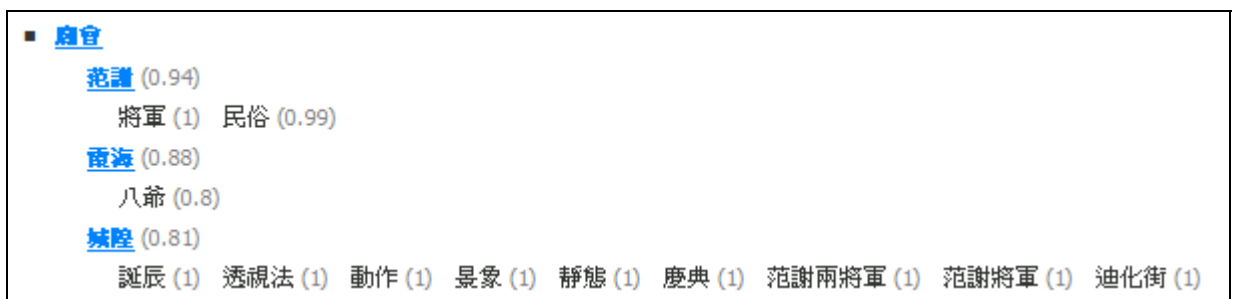


圖 29： λ 為 0.8 時的階層式標籤分群結構

七、 產出結果

1. 標籤地圖 (Tags Map)

在經過階層式標籤分群之後，將會產生各個美術作品類別的標籤地圖，如圖 30 所示。它是結合關鍵詞與標籤，創造出大眾分類的方式。標籤地圖也可以當作是一種導覽的模式，幫助使用者快速瀏覽及發掘整個網站的資源。經由眾多人標記之後，其標籤地圖便會愈趨完整，用標籤展示方式，可以幫助使用者找尋其有興趣的資源，以達到快速搜尋的目的。



繪畫 Tags:425

標籤地圖 Tags Map (Tag-Groups: 25) updated: 2008/6/12 下午 03:57:16 [new](#)

- **水彩**
 - [生活](#) (0.69)
 - [遠望](#) (0.79)
 - [悠閒](#) (0.76)
 - [窗外](#) (0.95)
 - 街景 (1) 茶 (1) 窗景 (1) 畫家 (1) 桌椅 (1) 飲茶 (1) 憶 (1) 溫暖 (1) 樓房 (1) 屋內屋外 (1) 城市 (1) 景物畫 (1) 異文化 (1) 悠賞 (1) 日常生活 (1) 水壺 (1) 千變萬畫 (1) 2 (1) 午後 (1)
 - [書](#) (0.85)
 - [靜物](#) (0.6)
 - [香蕉](#) (0.78)
 - 畫面 (1) 水果 (1) 蘋果 (1) 中央 (1) 臺師大 (1) 流洩 (1) 猴子 (1) 水彩素描 (1)


- **寫意**
 - [風景](#) (0.52)
 - [遠景](#) (0.56)
 - 松樹 (0.79)
 - [熱人](#) (0.73)
 - [郊外](#) (0.89)
 - 羅馬郊外 (1) 中景 (1) 近景 (1) 廣闊 (1)
 - [生動](#) (0.63)

- **美術**
 - [台北](#) (0.79)
 - [城隍廟](#) (0.82)
 - 寫生稿 (0.93)
 - [造型](#) (0.73)
 - [差差](#) (0.53)
 - [雕塑](#) (0.83)
 - 毀滅 (1)
 - [背景](#) (0.59)

圖 30：繪畫類標籤地圖部分擷取畫面

2. 推薦詞及搜尋標籤

階層式標籤分群的另一個用途是產生每一個數位資源相關聯的推薦詞。推薦詞是從每一個藝術作品既有的標籤及關鍵詞所衍生出來的，以標籤及關鍵詞當作節點，再透過標籤地圖建立的關聯性去尋找其上層的父節點及相似度大於 0.5 的子節點，如圖 31 所示。利用推薦詞可以搜尋到其他的相關的數位資源。



標記:

(連續輸入請用空白鍵隔開)

標籤使用說明(一定要看)

關鍵詞: 藝術系.學院.
瀟灑.聚落.塵沙.黃昏.晚霞.老厝.房子.傍晚.台北.古厝.建築.沙塵暴.清真寺.老台北.屋頂.故鄉.新公園.

標籤: 夕照.台北街廓.浪漫.餘暉.祥龍獻瑞.美術.都市.好.悠然遙想.

推薦詞: 城隍廟.燕尾.樹.鄉村.水彩小品.林家.

標籤地圖 <標籤地圖資料多,點選後請稍待>

圖 31：楊英風美術作品的推薦詞

八、系統操作畫面

當使用者進入交大楊英風數位美術館網站，可點選呦呦作品集的美術作品，裡面收錄了楊英風所有藝術作品，其分類依據楊英風全集，包括：版畫、繪畫、速寫、雷射、攝影、浮雕、景觀浮雕、獎作、美術設計、漫畫、插畫等。任意選擇其中一個分類，即可點選想要看的美術作品，然後進行標記動作。

圖 32是為一個美術作品提供標籤的畫面，在標記功能下方的資訊依序為關鍵詞、標籤、推薦詞、標籤地圖，關鍵詞由系統自動擷取出來的，標籤是使用者提供的，推薦詞則是系統結合前兩者產生的，點選標籤地圖連結即可出現所在類別的標籤地圖。這幅作品名稱是「霞海城隍誕辰」，講述這幅圖以中國遠近法構圖，與「范謝兩將軍」為連作。系統萃取了城隍、誕辰、范謝、將軍等關鍵詞；標籤則有范謝兩將軍、廟會、慶典等；推薦詞提供相關的字詞為構圖、民間、油彩等。

	原件與否	重製
	作品名稱	霞海城隍誕辰
	尺寸大小	56 × 13 cm
	創作日期	1955
	體裁／表現／範圍	繪畫
	技術／技法	水彩
	材質	紙
創作地點	台北	
摘要	<p>這幅《霞海城隍誕辰》，與楊英風作於一九五四年，曾榮獲第九屆「臺灣全省美術展覽會」國畫部入選的《范謝兩將軍》可視為對同一主題的連作。這兩幅作品都是取材於相同的寫生稿，但構法一則以中國畫的遠近法，一則以西方的透視法構圖，兩幅畫恰如題稱，《霞海城隍誕辰》呈現廟會的繽紛景象，《范謝兩將軍》則是靜態地呈現置放於廟前廣場的七爺八爺神祇，呈現熱鬧與祥和迥然不同的氛圍。《霞海城隍誕辰》由於採西方水彩紙尺幅與透視法構圖，明顯較《范謝兩將軍》緊緻；《范謝兩將軍》中靜態置放廣場的七爺八爺神祇，在《霞海城隍誕辰》被截去一半與人群重合，便有如有人群中動態遊走；而人群在兩幅作品中的動作雖然不變，《范謝兩將軍》中人與人之間疏離少互動，在《霞海城隍誕辰》人物卻因為分割重疊多了關連性，構成了參與廟會活動的群體，使畫面因人群的活躍動作，使作品充分表達廟會人聲鼎沸的情境。</p>	
典藏單位	<p>原件由楊英風家屬典藏，本數位作品由財團法人楊英風藝術教育基金會及行政院文化建設委員會國家文化資料庫分別典藏。</p>	
問卷填寫	<p>標記系統問卷A 標記系統問卷B</p>	
<p>標記: <input type="text"/> <input type="button" value="送出"/></p> <p>(連結輸入請用空白鍵隔開)</p> <p>標籤使用說明(一定要看)</p> <p>關鍵詞: 城隍 誕辰 范謝 將軍 人群 動作 景象 西方 靜態 廣場 神祇</p> <p>標籤: 范謝兩將軍 七爺 八爺 廟宇 慶典 民俗 廟會 霞海 油化街 透視法 熱鬧 范謝將軍</p> <p>推薦詞: 構圖 民間 早期 信仰 木柵 油彩 茶座 瑪歌娜</p> <p>標籤地圖 <標籤地圖資料多,點選後請稍待></p>		

圖 32：楊英風美術作品中的關鍵詞、標籤、推薦詞資訊

點選圖 32的任一標籤/關鍵字都會連結到圖 33的搜尋畫面，方便使用者利用標籤搜尋其他相關的美術作品，它除了可以直接由標籤來搜尋資料外，也可以用作品名稱、創作日期、創作地點、材質等。

美術作品 | 老照片 | 日記 | 楊英風論述 | 景觀規劃 | 工作週記 | 回作品集

古厝 標籤 Go

作品名稱	創作日期 / 拍攝日期
1. 台灣農家	1947
2. 霧峰古厝	
3. 傍晚	
4. 台南孔廟	1957

圖 33：楊英風數位美術館標籤搜尋畫面

圖 34是浮雕這個類別的標籤地圖，從地圖中可看到在這個類別被分成 24 類 (Tag-Groups)，從這個類別當中可以發現美術作品內的相關主題，例如：某一些浮雕作品跟獎章、獎座有關，有一些作品是跟菩薩、文殊、普賢，有一些作品是跟主教、教堂、羅馬等內容。





浮雕 Tags:233

標籤地圖 Tags Map

(Tag-Groups: 24) updated: 2008/6/12 下午 03:57:16 [new](#)

■ 美章

[美座](#) (0.59)

[石油](#) (0.83)

中段 (1) 事業 (1) 工人 (1) 菱形 (1)

[電影](#) (0.55)

文藝獎 (1) 東方 (1) 主題 (1) 材質 (1) 文藝 (1)

■ 銅章

[雕刻](#) (0.85)

[紀念章](#) (0.84)

[愛娃波克](#) (0.74)

威尼斯 (1) 水都 (1) 娃波克 (1)

[葛莉爾](#) (0.77)

裘波勒 (1) 雕刻家 (1)

■ 菩薩

[文殊](#) (0.91)

師利 (1) 坐騎 (1) 獅子 (1) 新竹 (1) 普賢 (1) 華嚴三聖 (1) 寶塔 (1)

■ 美展

[方圓](#) (0.78)

[天地](#) (0.83)

金龍獎 (1) 喜悅 (1) 教育部 (1) 中國人 (1) 生命 (1) 新生 (1) 新龍 (1) 結構 (1) 疊紙 (1) 變化 (1) 繪畫 (1)

圖 34：浮雕類標籤地圖部分擷取畫面

肆、問卷評估與結果分析

本章就楊英風數位美術館標記功能之問卷評估調查結果，進行統計分析。第一節為說明問卷設計的方式；第二節為問卷評估項目；第三節為受測者基本資料與網路使用行為分析；第四節為社會性標記應用之需求分析；第五節為標籤、關鍵字、推薦詞內容分析；第六節為社會性標記功能的評估；第七節為社會性標記實作之使用滿意度分析；第八節則為綜合討論本研究所提供的標記功能是否確實符合使用者需求，並針對問卷評估中表現較差的部分進行深入探討。

一、問卷設計

為驗證本研究系統發展之實用性，研究中讓受測者實地瀏覽楊英風數位美術館網站，並對站內藝術品進行標記，再以線上問卷的方式進行評估。評估項目為(1)受測者基本資料與網路使用行為分析；(2)社會性標記應用之需求分析；(3)標籤、關鍵字、推薦詞內容分析；(4)標記功能分析；(5)使用滿意度分析。在受測者選擇方面，前端網頁採公開發佈的方式進行測試，讓受測者基本背景不受到限制，降低受測者對評估結果的影響，也有更多機會讓一般使用者參與本研究之實作。

考量系統為公開發佈的方式進行測試，對於評估流程與評估項目皆有明確的說明圖示，供使用者參考，亦包含標記功能之說明與網頁介面的操作說明。

評估方式分別由 A、B 問卷進行評估，問卷 A 是讓受訪者實地瀏覽楊英風數位美術館內的網站作品，挑選同一類別內的 5 幅作品並建立標籤，利用標記功能瀏覽藝術品，再填寫問卷中的問題。問卷 A 可重複填寫，但每次填寫必須評估不同的標的物，此部分標的物為藝術作品；問卷 B 的主要評估項目為社會性標記應用需求分析及實作滿意度分析，問卷 B 填寫一次即可。本問卷設計流程主要是讓使用者先體驗標記之各項功能，進行問卷 A 的回答，然後再對需求及滿意度進行問卷 B 的填答。本研究之設計網頁介面上可任意瀏覽想觀賞的藝術作品，待受測者實地使用過標記功能後，點選畫面上之問卷 A 或問卷 B 連結，可隨即對楊英

風數位美術館標記功能進行評估。

系統採公開發佈方式進行，評估時程為期三周，問卷調查樣本包含交通大學校內學生及一般上網民眾，其中問卷 A 有 113 份，問卷 B 有 104 份，總共評鑑的美術作品有 186 個，共產生 1088 個標籤，後續幾節即針對回收問卷進行統計分析，並描述評估之結果。

二、 評估項目

本研究所收集的受測者基本資料為系所、年齡、學歷。系統評估分為 A、B 問卷，A 問卷為主要標記功能的評估，B 問卷則針對網路使用行為、標記功能需求及滿意度的評估，詳細評估項目如下：

問卷 A：

1. 標記功能之評估

- (1) 請寫出您所標記的作品名稱?
- (2) 您認為在此幅作品中，系統提供的關鍵詞能代表此作品的程度?
- (3) 您認為在此幅作品中，大眾提供的標籤能代表此作品的程度?
- (4) 您認為系統提供的推薦詞有助於您蒐尋到其他相關聯的作品(點選字詞連結即可得到搜尋結果)?
- (5) 在您對此幅作品標記之後，系統是否產生新的推薦詞?

2. 標記工具之評估

- (1) 您認為本研究之標記功能，有助於理解您所標記藝術作品的程度?
- (2) 您認為透過標籤，對您所瀏覽的藝術作品能否分享大眾不同的概念或想法?
- (3) 您認為透過標籤，可搜尋到更多潛藏在網站內的藝術作品?
- (4) 您認為所在類別的標籤地圖，由大眾建立的分類可當成此類藝術作品的導覽?
- (5) 您認為所在類別的標籤地圖能否提供快速瀏覽藝術作品的標籤(關鍵字)?

- (6) 您認為所在類別的標籤地圖能否提升瀏覽所在類別作品的便利性?
- (7) 您認為標籤地圖是否具有實用性?

問卷 B：

1. 網路使用行為

- (1) 請問您接觸網際網路的時間有多久?
- (2) 請問您平均每天使用網路的時間有多長?
- (3) 請問您使用過哪些社會性標記系統?

2. 社會性標記需求評估

- (1) 請問您認為將藝術品或美術作品建置於線上美術館的重要性為何?
- (2) 請問您認為是否需要將社會性標記應用於線上美術館以便讓大眾了解及評判藝術作品的特質?
- (3) 請問您認為是否需要運用 Web 2.0 精神於線上美術館以便讓大眾分享標籤內的語彙?
- (4) 請問您覺得是否需要應用社會性標記，讓使用者可以共同標記網站內容，使分類更貼近使用者?

3. 使用滿意度評估

- (1) 請問您認為楊英風網站的標記系統是否具有實用性?
- (2) 請問您認為將社會性標記應用於數位典藏網站是否有創意?

三、 受測者基本資料與網路使用行為分析

1. 基本資料分析

(1) 受測者系所

此受測者系所分佈所佔比例以電機資訊學院及人文社會學院為最大，理工學院、管理學院、其他各佔 13%~15% 比例，剩下的則佔很少數，如圖 35 所示。

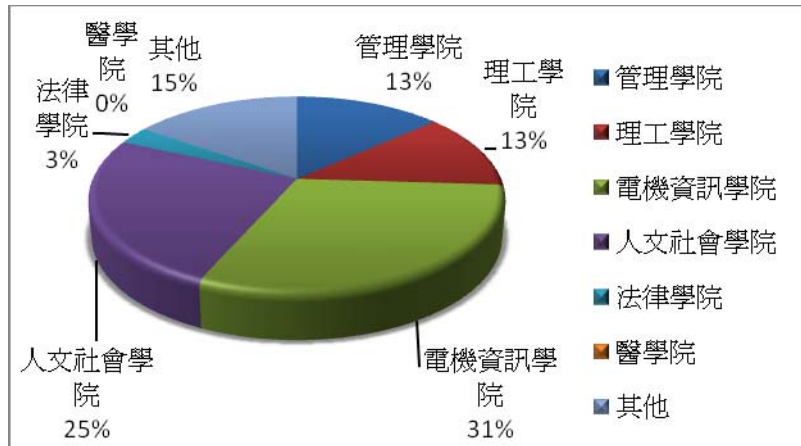


圖 35：受測者系所分佈圖

(2) 受測者年齡

此受測者超過一半的比例(55%)為大學生的年齡層，24~28 歲佔 24%，29~35 歲則減少至 16%，其他的年齡層則佔少數，如圖 36 所示。

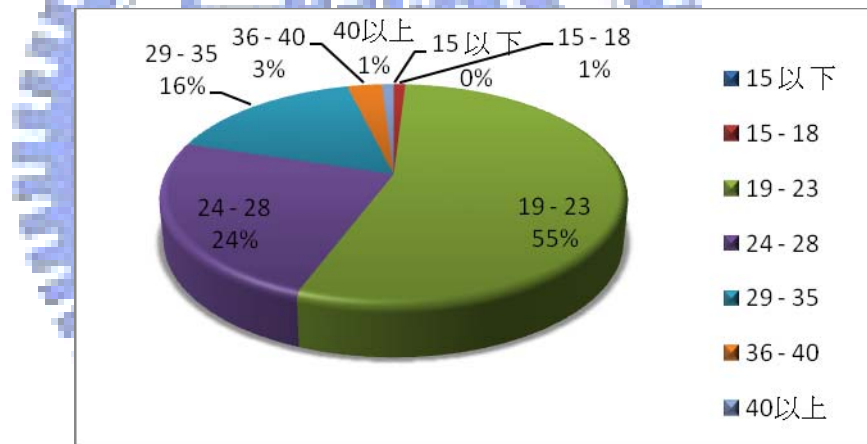


圖 36：受測者年齡分佈圖

(3) 受測者學歷

此受測者學歷超過一半為大學生(61%)，接近四成為研究所學歷，其他則佔少數，如圖 37 所示。

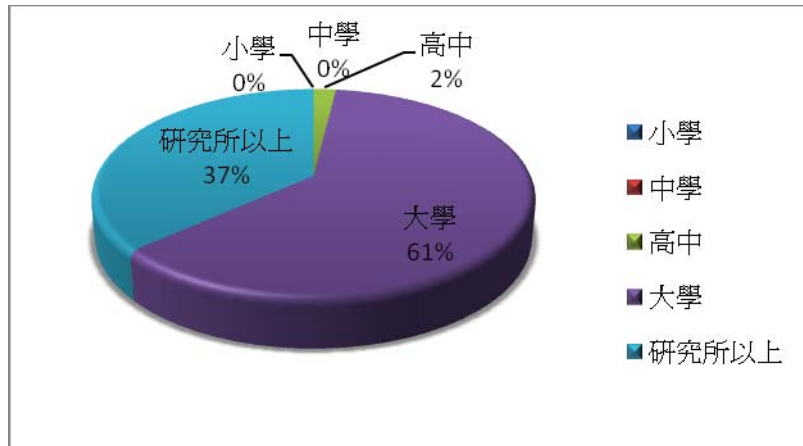


圖 37：受測者學歷分佈圖

2. 網路使用行為分析

(1) 接觸網路的資歷

由圖 38 可以發現受測者之接觸網路資歷有九成具五年以上資歷，四至五年只有 7%，四年以下則只有 3%。

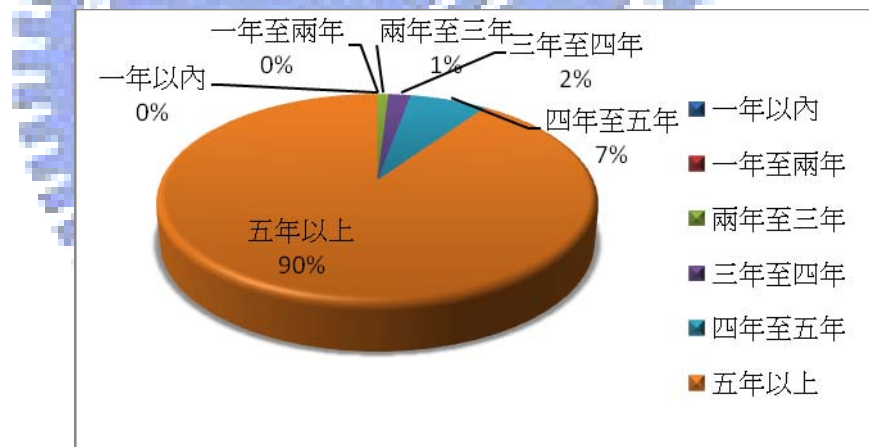


圖 38：接觸網路的資歷分佈圖

(2) 每日使用網路的時間

達五成受測者每天使用網路的時間為 3~5 小時，5~7 小時者佔 12%，1~3 小時則佔 14%，而在 1 小時以內的有 24%，如圖 39 所示。

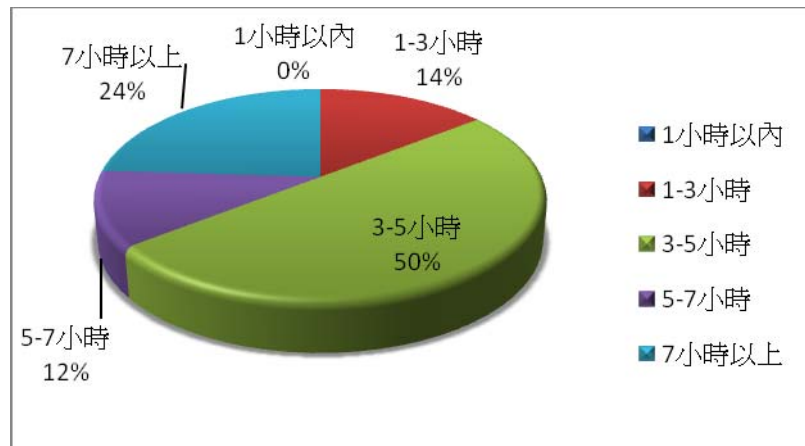


圖 39：每日使用網路的時間分佈圖

(3) 使用過的社會性標記系統

由圖 40可看出較常使用的社會性標記系統為YouTube、Flickr、Del.icio.us、Yahoo!分享書籤，大多是比較知名當紅的網站，國內的書籤網站有HEMIEDEMI、Myshare、PChome Search，各占 6%或 7%，而其他國外的書籤網站則較少被使用者採用，僅 Furl有 4%，其餘有使用的也只佔 1%或 2%。

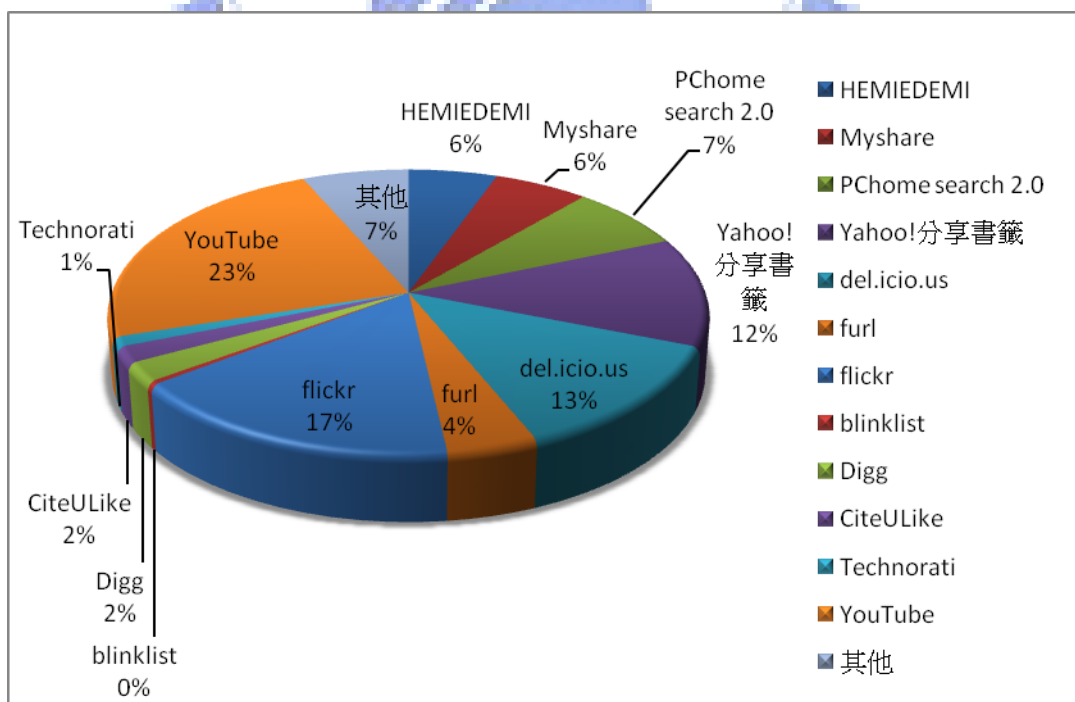


圖 40：使用過的社會性標記系統分佈圖

四、 社會性標記應用之需求分析

需求分析評估社會性標記是否需要應用於線上美術館，總共詢問四個問題，依序為「藝術品或美術作品建置於線上美術館的重要性」、「社會性標記是否需要應用於線上美術館，以便讓大眾了解及評判藝術品的特質」、「是否須將Web 2.0 精神應用於線上美術館，以便讓大眾分享標籤內的語彙」、「是否需要讓使用者可以共同標記網站內容，使分類更貼近使用者」。圖 41顯示社會性標記應用的需求，重要及非常重要涵蓋了大多數受測者的建議，少數受測者覺得普通，幾乎沒有受測者認為不重要及非常不重要。

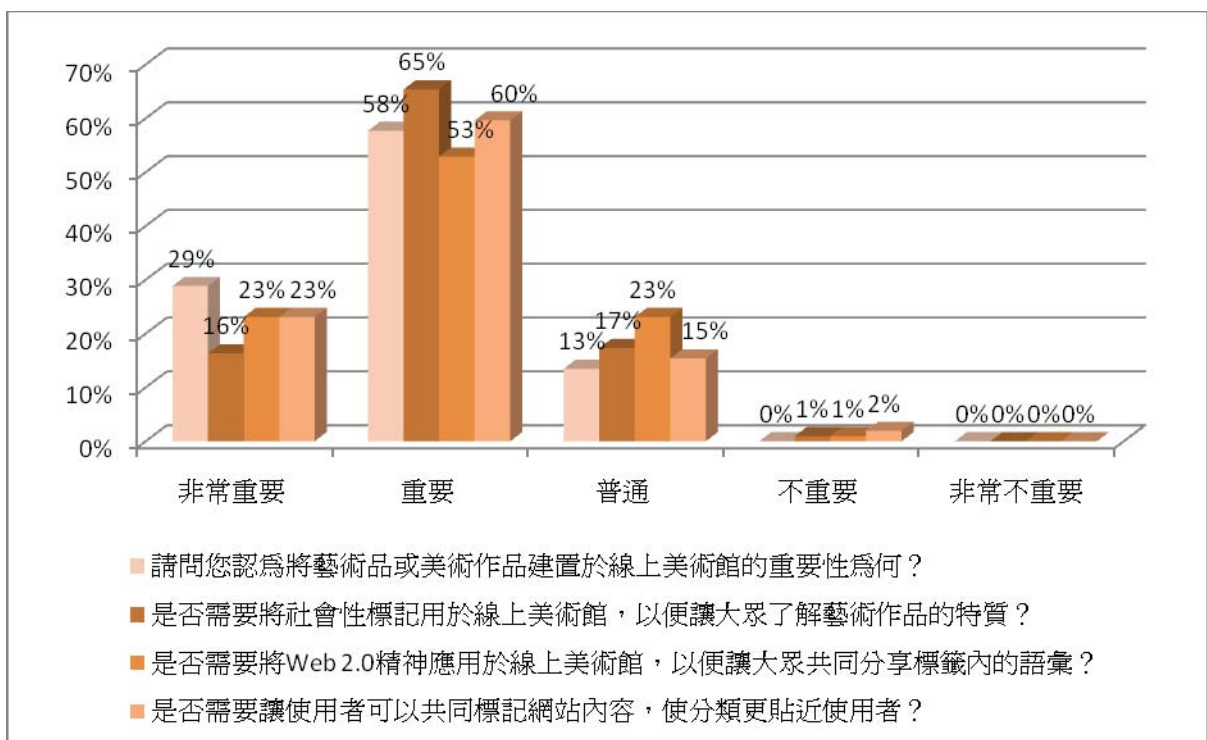


圖 41：受測者對於社會性標記應用之需求分析

五、 關鍵詞、標籤、推薦詞內容分析

本節是探討當受測者在同一類別內任選五樣藝術作品並加註標籤時，評估系統關鍵詞、大眾提供的標籤是否代表所瀏覽藝術品的程度，並評估系統產生推薦詞是否能幫助搜尋其他作品。問卷中對所瀏覽作品詢問了三個問題，依序是「系統提供的關鍵詞能否代表此作品」、「大眾提供的標籤能否代表此作品」、「系統提供的推薦詞是否有助於蒐尋其他相關聯的作品」。以下分別以楊英風數位美

術館中的景觀雕塑類及繪畫類來分析其內藝術品的標籤、關鍵詞、推薦詞。

(1) 景觀雕塑類作品分析

在景觀雕塑類作品中，對於受測者挑選的五樣作品瀏覽，由圖 42 可觀察受測者對於作品內的關鍵詞、標籤、推薦詞給予好及普通的範圍，其中大眾提供的標籤「好」的程度最高，顯示標籤確實能彼此分享對藝術作品的觀感，總體來說，受測者對於這個類別的作品大多給予普通以上的程度，顯示本系統所設計之標籤、關鍵詞、推薦詞都獲得正面以上的評價。

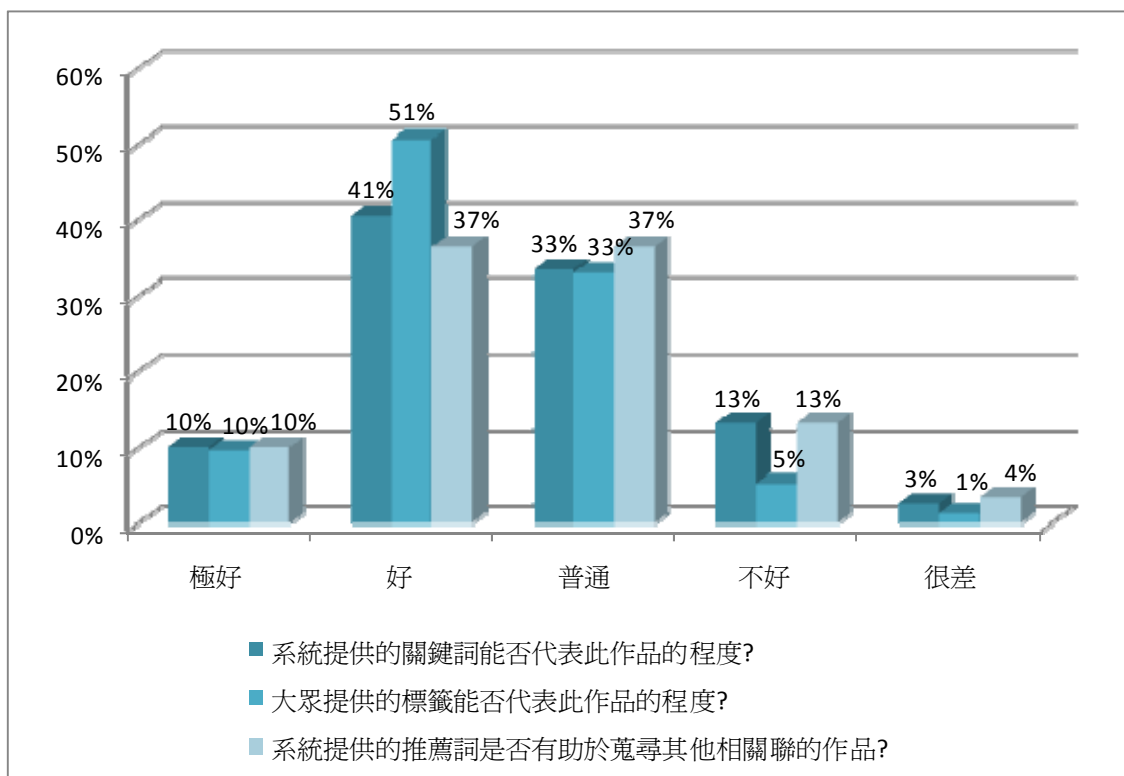


圖 42：景觀雕塑類別作品分析

(2) 繪畫類作品分析

在繪畫類作品中，由圖 43 可觀察受測者對於其內作品之關鍵詞、標籤、推薦詞代表藝術作品的程度給予好及普通的評價，其中大眾提供的標籤「好」的程度也是最高的，其次是關鍵詞，最後是推薦詞，總體來說，雖然受測者對於這個類別的作品大多給予普通以上程度的評價，但是發現在此類別關鍵詞及推薦詞「不好」與「很差」的評價比景觀雕塑類稍高，細節的探討將在

第八節中說明。

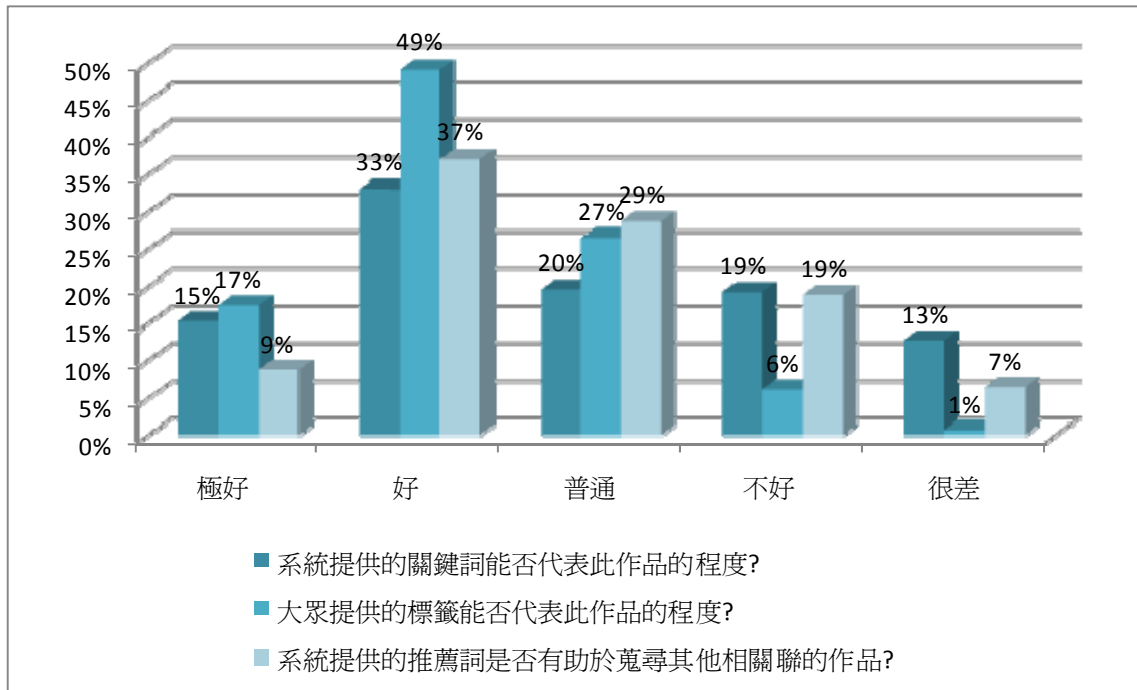


圖 43：繪畫類別作品分析

六、社會性標記功能評估

(1) 標記功能評估

標記功能的評估是由受測者對楊英風數位美術館實際使用標記功能後，所作的評估，主要詢問三個問題探討受測者對此功能使用的感受。三個問題依序為「本研究之標記功能，有助於理解所瀏覽的藝術作品」、「透過標籤，對瀏覽的藝術作品能分享大眾概念與想法」、「透過標籤可否搜尋潛藏在網站內的藝術作品」。由圖 44 可知三個問題受測者大多給予「極好」、「好」、「普通」，而「透過標籤，對瀏覽的藝術作品能分享大眾概念與想法」在「極好」又得到更多的評價，「標記功能有助於理解藝術作品」給予「極好」的程度上則稍為差些，「透過標籤可否搜尋潛藏在網站內的藝術作品」的「極好」的程度則介於以上兩者之間。

圖 44 顯示本系統之標記功能確實能幫助使用者理解與互相分享藝術作品的概念及想法，並利用標籤可達到有效蒐尋的目的。

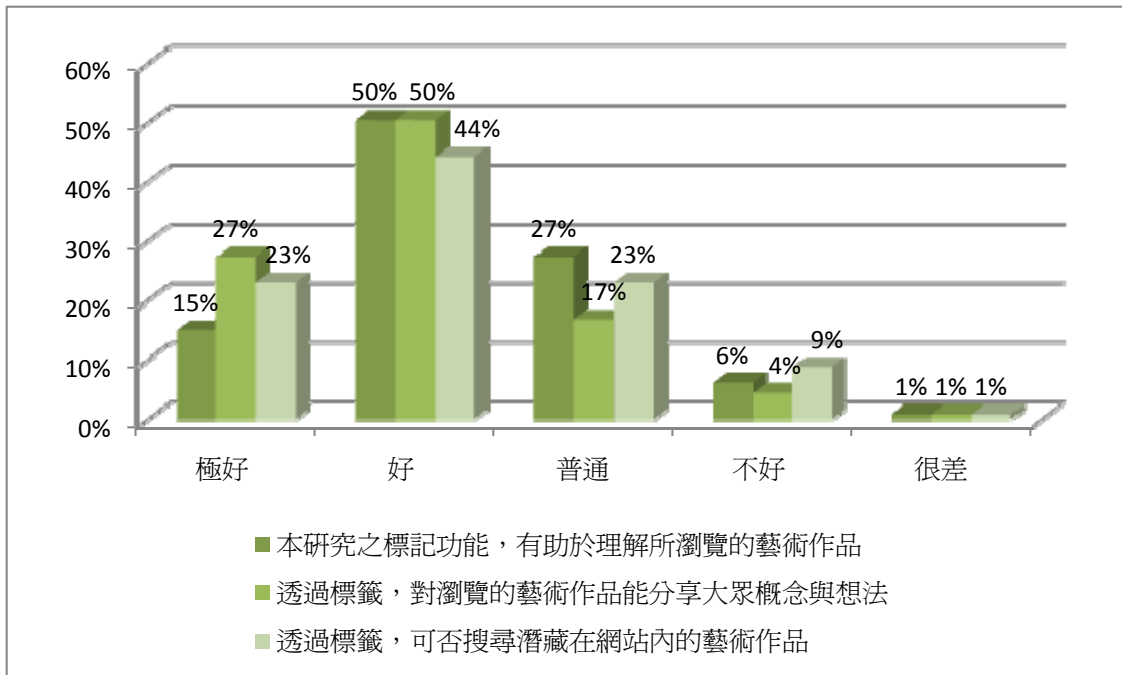


圖 44：標記功能評估分析

(2) 標籤地圖功能評估

標籤地圖是本研究標記系統所提供的另一功能，主要詢問四個問題以評估受測者對標籤地圖使用的感受。四個問題依序為「標籤地圖可否當作藝術作品的導覽」、「標籤地圖可否提供快速瀏覽藝術作品的關鍵字」、「標籤地圖可否提供瀏覽藝術品作的便利性」、「標籤地圖是否具實用性」，由圖 45 可發現標籤地圖功能的評估程度多居於「普通」、「好」及「極好」的範圍，故可知本系統之標籤地圖確實有其助益。

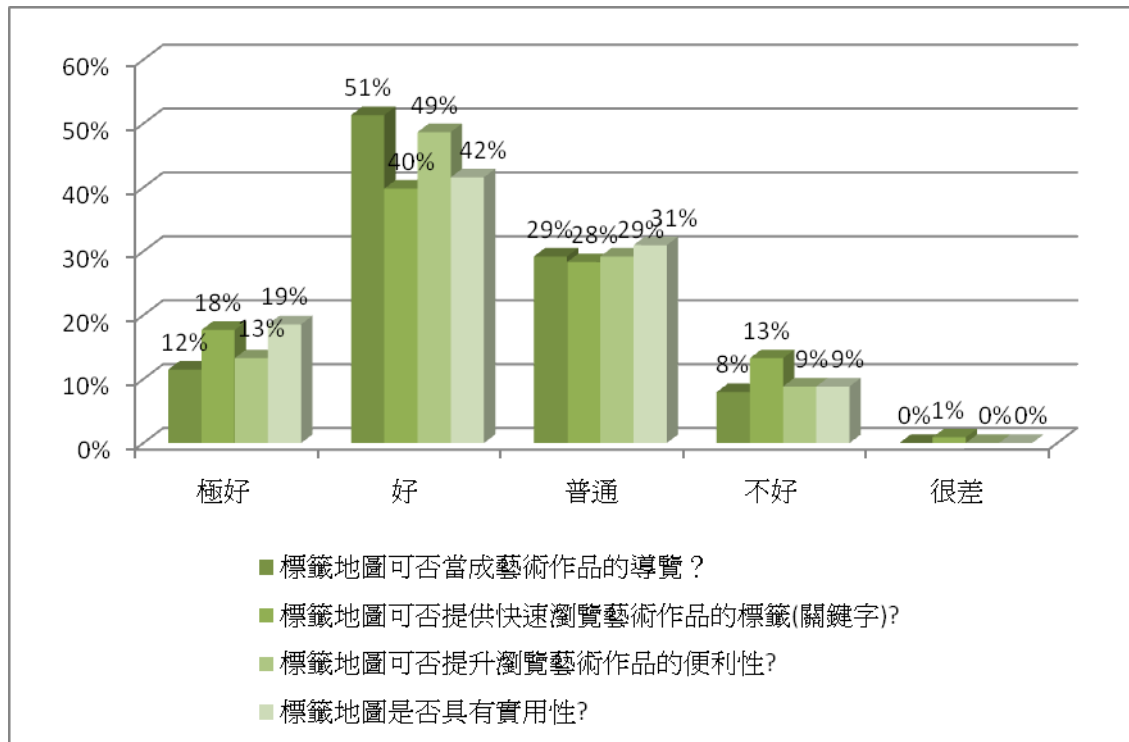


圖 45：標籤地圖功能評估分析

七、 社會性標記實作滿意度分析

(1) 楊英風數位美術館的標記功能是否具有實用性

圖 46 是顯示當受測者實地使用楊英風數位美術館之標記功能後，對於整體實用性的評估，有超過一半的受測者認為此功能具有實用性(59%)，而 9% 的受測者認為非常具有實用性，29% 的受測者則覺得普通，經由調查評估後可發現本研究之標記功能有 68% 的受測者給予本研究正面的評價，並認為本標記功能應用於楊英風數位美術館是具有實用性的。

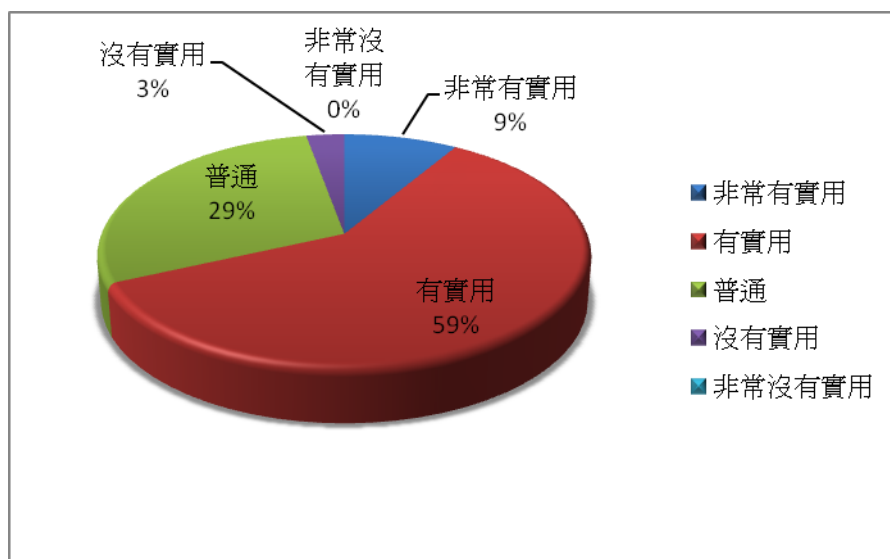


圖 46：受測者評估楊英風網站之標記功能是否具有實用性分佈圖

(2) 楊英風數位美術館的標記功能是否具有創意

圖 47顯示受測者是否覺得楊英風數位美術館之標記功能具有創意，從分佈圖來看，可發現受測者中高達七成的人給予本研究有創意的肯定，而非常有創意的佔 16%，普通及沒有創意的僅佔 14%，由此可知，超過八成的受測者認為本研究標記功能應用之呈現是具創意的，給予高度正面的評價。

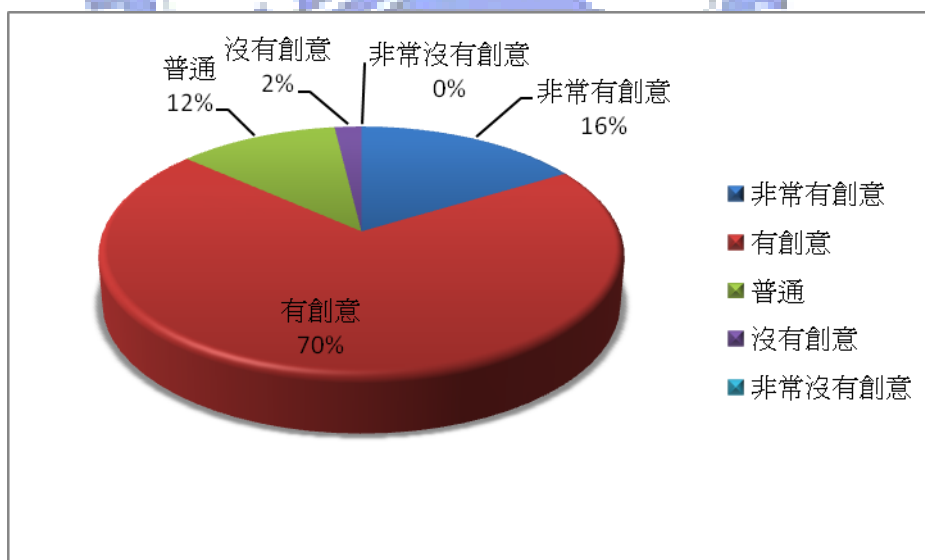


圖 47：受測者評估楊英風網站之標記功能是否具有創意分佈圖

八、 綜合討論

本研究經由受測者評估後，所得到的結果傾向於普通以上的評價，顯示出

標記功能確實可幫助使用者在楊英風數位美術館分享、瀏覽及搜尋資源。在本節中將深入探討關鍵詞、標籤、推薦詞的內容是否確實達到使用者之需求，並針對評價不好或極差狀況的結果深入探討，作為後續改進的依據。

在第五節曾經發現繪畫類中，「關鍵詞能否代表此作品」，不好及很差的評價佔約 30%，「推薦詞是否有助於蒐尋其他相關聯的作品」不好及很差的評價也有 25%，在此查看於繪畫類中，系統無法提供良好關鍵詞及推薦詞的為何種藝術作品。

如圖 49，繪畫類作品「窗外」，被某些受測者於關鍵詞評價很差，推薦詞也多呈現「不好」及「很差」的評價，這幅作品窗外，意思大約是說窗明几淨的畫面放著書本及茶具，呈現畫家日常生活的一面，作品強調悠閒、明快、愉悅等等感受。關鍵詞「畫家」在此作品的摘要出現二次，故系統根據權重必定會將此關鍵詞擷取出來，在此作品中受測者給予的標籤為景物、窗景、桌椅、靜物、悠閒等字眼，而非著重在畫家這個詞彙。

在推薦詞的地方，系統推薦香蕉、遠景、燕尾、古厝、樹等詞彙。若是點選推薦詞香蕉，則會搜尋到「香蕉」這幅作品(如圖 50)。

在香蕉這幅作品中，其內標籤出現「靜物」、「香蕉」，兩幅作品共用的標籤為「靜物」，因此系統會將這兩個標籤的關聯性建立起來，如圖 48，兩者計算出來的標籤相似度為 0.78，所以「窗外」這個作品推薦詞會出現「香蕉」這個詞彙，也就是兩幅作品共同相似的地方是「靜物」這個標籤，故在「窗外」，就會推薦「靜物」階層底下的詞彙給使用者，以搜尋其他作品，其他的推薦詞也是如此產生出來。

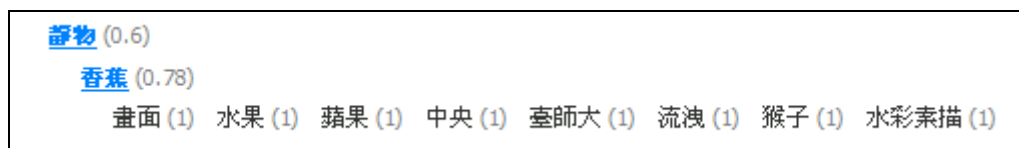



圖 48：靜物與香蕉的階層式結構

不同的使用者對作品感受不同，建立標籤所用到的詞彙也有所不同，雖然依

據演算法產生正確的推薦詞，仍然無法滿足所有使用者的喜好，某些使用者會認為此推薦是不好的。針對此種現象的解決方法之一是建立個人化標籤，讓不同使用者針對自己加註的標籤得到個人化的推薦。另外，關鍵詞是經由系統自動擷取出每一個作品內重要的詞彙，以作為使用者參考及快速瀏覽作品的依據，若是系統擷取出來的關鍵詞與使用者想法有所差異時，應當可讓使用者即時作手動的調整，因為社會性標記的精神著重使用者的互動，以使用者的體驗為依歸，接下來在下一章將說明現階段不足之處，並闡述未來發展方向。



標記:

(連續輸入請用空白鍵隔開)

標籤使用說明(一定要看)

關鍵詞: 畫家.

標籤: 街景.屋內屋外.窗景.水彩.景物畫.窗外.水壺.靜物.寧靜.香.桌椅.茶.生活.風景.悠閒.飲茶.溫暖.午後.2.憶.建築.日常生活.城市.異文化.樓房.千變萬畫.遠望.悠賞.

推薦詞: 香蕉.遠景.燕尾.古厝.樹.

圖 49：繪畫類作品—窗外




標記:

(連續輸入請用空白鍵隔開)

標籤使用說明(一定要看)

關鍵詞: 畫面.中央.

標籤: 水果.香蕉.蘋果.水彩.猴子.臺師大.靜物.流洩.水彩素描.

推薦詞: 生活.

圖 50：繪畫類作品—香蕉

伍、 結論與未來展望

本章總結本論文並說明實作社會性標記於楊英風數位美術館所達成的效益。第一節總結本研究提出的方法及對於數位典藏所產生的應用結果，第二節則闡述本研究未來可發展的方向。

一、 結論

本研究應用了Hsieh等人[13]的方法於數位典藏，並加以改良，結合中文斷詞、資訊檢索於網站資源的詮釋資料，並利用標籤及關鍵詞的共用性及相似度，建立階層式標籤分群，產生推薦詞給使用者，經由受測者評估測試確實能得到很好的結果及很高的使用滿意度。

在研究過程中，回顧先前標籤的研究，大多將使用者給予的標籤直接作群聚演算，而未結合原始數位資源的詮釋資料，故系統於一開始建置時，會造成無法分群或分群稀疏的結果，本研究認為詮釋資料也扮演一個很重要的角色，裡面包含廣泛的資訊，如何快速將詮釋資料轉換為知識讓使用者理解其實是一件很重要的工作。於是本研究針對此點，先將數位資源的詮釋資料作前置處理，利用字詞權重計算擷取出關鍵詞，列於網頁上，當使用者經由關鍵字搜尋或給予標籤時則會隱性調整系統關鍵詞的權重，最後結合使用者標籤與系統關鍵詞建立階層式標籤分群，經由資源間共用標籤的關係，讓標籤與標籤間的連結建立起來，以產生推薦詞。

本研究發現透過標籤的群聚，產生的大眾分類頗能符合使用者需求，並鼓勵人們提供自己想法的標籤使蒐尋更為廣泛及容易，這些目的都是希望能夠挖掘出潛藏於網站內部不易被大眾所發掘的資源，透過標籤之間的串連將搜尋的觸角深及於整體資源的深處。

在呈現標籤資訊方面，本研究提供一個標籤地圖，作為整體資源的導覽。標籤地圖在楊英風數位美術館上是以不同的美術作品類別來分別呈現，因在不同類別的美術作品多有不同描述及特性，故初始設定便以分別呈現的方式，只要進入

網站內選擇某一美術作品加註標籤時，就可以點選顯示所在類別的標籤地圖。標籤地圖的功用是一種輔助概念，用索引的方式輔助使用者搜尋有興趣的資源，並可從中發掘特定類別作品群聚的標籤之主題性。經系統評估顯示，標籤地圖的應用，確實有助於提升瀏覽藝術品的實用及便利性。

針對本研究主要之社會性標記功能，在使用者問卷中，關鍵詞、標籤、推薦詞極好與好的評價約佔 40%~60%，而普通約佔 20%~30%，足以證實本研究內容的精確及有效性。綜觀最後所得之成果，除了改良先前標籤分群演算法，也在標記功能中提出創新的概念，從問卷評估的使用者滿意度及創意上，皆得到高度的評價，顯示本研究發展是具有正面結果的。然而，在研究中仍發現有許多需要進一步改良以滿足使用者需求之處，於下一節中說明未來研究發展方向。

二、 未來發展方向

本研究致力發展一套社會性標記平台於楊英風數位美術館，於現階段獲得正面的評價，但仍存在許多需要精益求精的地方，以下依序提出未來研究之發展方向與應改善之處。

- (1) 現階段的研究並未區分個人化標籤與大眾標籤，故皆可分享所有使用者建立的標籤，雖然得到的搜尋詞彙較為廣泛，但其缺點是並非所有使用者的喜好、感受皆相同，所運用的詞彙也不一樣，他人得到的推薦詞不一定是自己所認同的推薦，故接下來的研究必須建立個人化標籤，有個人化標籤之後便會有個人化標籤地圖，如此可知道不同使用者一定會有不同觀賞喜好以及所使用詞彙的習慣。運用個人化標籤，每個人所得到的標籤群聚效果就會有所不同，有些人可能喜好抽象、水彩、風景、柔和的線條等，有些人可能喜歡強烈對比的、剛硬線條的畫面，也有人會喜歡金屬、光澤、雕塑等作品。再者，個人化標籤也可與大眾標籤結合，建立更廣泛的搜尋與推薦。
- (2) 對標籤本身來說，由於隨著使用者不斷地新增，標籤數量也會愈來愈龐

大，有時使用者會有建立錯誤標籤的可能性，所以必須建立一套標籤管理機制。若是個人化的標籤則可藉由使用者自行管理、新增、刪除或修改，若是透過標籤間的共用性以推薦給其他使用者時，則必須查核是否推薦的詞彙出現頻率或相似度夠高，達到足以推薦給其他使用者的門檻。

- (3) 關鍵詞本身是由系統自動擷取出來的，作為系統剛建置的初始搜尋詞彙，其準確性必須依賴中文斷詞的技術，現階段系統自動擷取出來的詞彙畢竟多有所謬誤，很難達到 100% 的準確性，故未來發展方向可由使用者及時調整關鍵詞的權重，或使用者自行刪除或修改，以符合更精確更具代表藝術品的關鍵詞。
- (4) 本研究標記功能之標籤分群結構是以樹狀圖展開，目的是方便使用者瀏覽及直接點選所需要的標籤連結，礙於技術性及實驗性質的考量，所提供的功能仍有諸多限制，未能達到完全友善的介面，階層結構為靜態呈現，無法讓使用者自行調整。另外標籤的畫面隨使用者想瀏覽再展開即可，畫面上可呈現更友善與互動的介面。
- (5) 許多技術上可改進的空間，都能讓本研究更臻完美。後續的研究可再進一步透過使用者訪談等質性分析，深入了解使用者的需求及對標記系統的觀感，使未來能設計出具有更高優使性(Usability)的介面，以期能創造更完善的解決方案。

參考文獻

- [1]. Begelman, G., Keller, P. and Smadja, F., “Automated tag clustering: Improving search and exploration in the tag space”, *Collaborative Web Tagging Workshop, 15th International World Wide Web Conference 2006*.
- [2]. Chen, K. J. and Bai, M. H., “Unknown Word Detection for Chinese by a Corpusbased Learning Method”, *International Journal of Computational linguistics and Chinese Language Processing*, 1998, vol. 3, pp. 27-44.
- [3]. Chen, K. J. and Ma, W. Y., “Unknown Word Extraction for Chinese Documents”, *Proceedings of COLING*, 2002, pp. 169-175.
- [4]. Chen, K.J. and Kiu, S.H., “Word identification for Mandarin Chinese sentences”, *Fifth International Conference on Computational Linguistics 1992*, pp. 101-107.
- [5]. Chun, S. and Jenkins, M., “Cataloguing by crowd; a proposal for the development of a community cataloguing tool to capture subject information for images”, *Museums and the Web 2005*, Vancouver, Retrieved June 21, 2008, from http://www.archimuse.com/mw2005/abstracts/prg_280000899.html
- [6]. Chun, S., Cherry, R., Hiwiller, D., Trant, J. and Wyman, B., “Steve.museum: an ongoing experiment in social tagging, folksonomy, and museums”, *Museums and the Web 2006*, Albuquerque, New Mexico.
- [7]. Cuene, J., “Web 2.0: Evolution towards a read/write platform”, Retrieved June 21, 2008, from <http://www.weiterbildungsblog.de/archives/000863.html>
- [8]. Despres, C. and D. Chauvel., “Knowledge management(s)”, *Journal of knowledge management*, 1999, Volume 3 (Issue 2).
- [9]. Hammond, T., Hannay, T., Lund, B. and Scott, J., “Social bookmarking tool (1) – A general review”, *D-Lib Magazine*, 11(4), Retrieved June 21, 2008, from <http://www.dlib.org/dlib/april05/hammond/04hammond.html>
- [10]. Han, J. and Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2000.
- [11]. Hend, S. A. and Hugh, C. D., “Folksonomies versus automatic keyword extraction: an empirical study”, *Proceedings of IADIS Web Applications and Research 2006*, Southampton, United Kingdom.
- [12]. Hinchcliffe, D., “Web 2.0 definition updated and Enterprise 2.0 emerges”, *Web 2.0 Conference 2006*, Retrieved June 21, 2008, from <http://blogs.zdnet.com/Hinchcliffe/?p=71>
- [13]. Hsieh, W. J., Lai, W. S. and Chou, S. C., “A Collaborative Tagging System for learning resources sharing”, *Current Developments in Technology-Assisted Education, FORMATEX*, Badajoz, Spain, 2006, pp.1364-1368. http://ls.hket.com/UserFiles/Image/121/Chapter_121.doc
- [14]. Jain, A.K. and Dubes, R.C., *Algorithms for Clustering Data*, Prentice Hall, 1988.

- [15]. Lee, D. L., Chuang, H. and Seamons, K., "Document ranking and the vector space model", *IEEE Software*, 1997, Vol.14 (2), pp. 67-75.
- [16]. Li, G. C., Liu, K. Y. and Zhang, Y. K., "Identifying Chinese Word and Processing Different Meaning Structures," *Journal of Chinese Information Processing*, 1988, Vol. 2, pp. 45-53.
- [17]. Liang, N. Y., "Knowledge of Chinese Word Segmentation," *Journal of Chinese Information Processing*, 1990, Vol. 4, pp. 42-49.
- [18]. Ma, W. Y. and Chen, K. J., "A bottomup Merging Algorithm for Chinese Unknown Word Extraction", *Proceedings of ACL workshop on Chinese Language Processing*, 2003, pp. 3138.
- [19]. Mathes, A., "Folksonomies-cooperative classification and communication through shared metadata", Retrieved June 21, 2008,
from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.pdf>
- [20]. Merholz, P., "Metadata for the masses", Retrieved June 21, 2008,
from <http://www.adaptivepath.com/ideas/essays/archives/000361.php>
- [21]. Millen, D., Felinberg, J., Kerr, B., "Dogear: social bookmarking in the enterprise", *Proceedings of the SIGCHI Conference on Human Factors in Computing*, Montréal, Québec, Canada, April 2006.
- [22]. Millen, D., Yang, M., Whittaker, S. and Felinberg, J., "Social bookmarking and exploratory search", *Proceedings of the Tenth European Conference on Computer Supported Cooperative Work*, Limerick, Ireland, September 2007.
- [23]. Nie, J. Y., Brisebois, M. and Ren, X., "On Chinese text retrieval", *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, August 1996, pp. 225-233.
- [24]. O'Dell, C., "A Current Review of Knowledge Management Best Practices", *Business Intelligence*, London, 1996.
- [25]. O'Reilly, T., "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", *COMMUNICATIONS & STRATEGIES*, 1st quarter 2007, no. 65, p. 17.
- [26]. Oslon, S., "Yahoo faces flickr backlash", *Cnet News.com*, Retrieved June 21, 2008,
from http://news.com.com/Yahoo+faces+Flickr+backlash/2100-1038_3-5844383.html
- [27]. Salton G., A. Wong, and C. S. Yang, "A Vector Space Model for Information Retrieval", *Journal of the American Society for Information Science*, 1975, Vol. 18(11), pp.613-620.
- [28]. Salton G., *Automatic Text Processing*, New York: Addison Wesley, 1989, pp. 313-373.
- [29]. Salton, G. and McGill, M.J., *Introduction to Modern Information Retrieval*, New York: McGraw-Hill Co, 1983.

- [30]. Schauer, B., "Experience attribute: Crucial DNA of Web 2.0", Retrieved June 21, 2008, from <http://www.adaptivepath.com/publications/essays/archives/000547.php>
- [31]. Seldow, A., "Social tagging in K-12 education: folksonomies for students folk", Retrieved June 21, 2008, from http://mrseldow.gradeweb.com/custom/Social_tagging_in_K12_Education_Seldow_4_3_06.pdf
- [32]. Shirky, C., "Many 2 Many: A Group Weblog on Social Software", August 2004, Retrieved June 21, 2008, from <http://many.corante.com/archives/2004/08/25/folksonomy.php>
- [33]. Sproat, R. and Shih, C., "A Statistical Method for Finding Word Boundaries in Chinese Text", *Computer Processing of Chinese and Oriental Languages*, 1990, pp.336-351.
- [34]. Trant, J., "Exploring the potential for social tagging and folksonomy in art museums: proof of concept", Retrieved June 21, 2008, from <http://www.archimuse.com/papers/steve-nrhm-0605preprint.pdf>
- [35]. Trant, T., "Social Classification and Folksonomy in Art Museums: early data from the steve.museum tagger prototype", *ASIST-CR Social Classification Workshop*, November 4, 2006.
- [36]. Tung, Y. H., "Upgrade to Web 2.0", Retrieved June 21, 2008, from
- [37]. Wal, T. W., folksonomy presentation at *Online Information Conference 2005* in London, Retrieved June 21, 2008, from <http://vanderwal.net/essays/051130/folksonomy.pdf>
- [38]. Weinberger, D., "Taxonomy to tags: From trees to piles of leaves", Release 1.0, 23(2), 1-33. Retrieved June 21, 2008, from http://www.hyperorg.com/blogger/misc/taxonomies_and_tags.html
- [39]. Wiig, K. M., *Knowledge Management Foundations: Thinking about Thinking – How People and Organizations Create, Represent, and Use Knowledge*, Arlington, TX: Schema Press, 1993.
- [40]. Wikipedia, "Social bookmark", Retrieved June 21, 2008, from http://en.wikipedia.org/wiki/Social_bookmarking
- [41]. Wikipedia, "分眾分類法", Retrieved June 21, 2008, from <http://zh.wikipedia.org/wiki/Wiki>
- [42]. Yu, T.Z. and Chien, L.F., "Automatic organization of user-generated tags from the Web" *Master Thesis of Nation Taiwan University*, 2006.
- [43]. 卜小蝶, "Folksonomy的發展與應用", 2007年, <http://www.lib.ncku.edu.tw/journal/16/1.htm>
- [44]. 卜小蝶, "淺談社會性標記之意涵與應用", 論文發表於淡江大學圖書館舉辦之「Web 2.0 與圖書館」研討會, 臺北市, 2006年12月。

- [45]. 中文斷詞系統簡介說明，<http://ckipsvr.iis.sinica.edu.tw/>
- [46]. 中文斷詞服務說明，<http://ckipsvr.iis.sinica.edu.tw/>
- [47]. 李士傑，“Wiki, 社會性網路軟體與知識管理”，2006年，<http://ilyagram.org/blog/present/blogwiki-km.pdf>
- [48]. 馬偉雲，“未知詞擷取作法”，中研院詞庫小組，<http://blackjack.iis.sinica.edu.tw/uwextract/document.html>
- [49]. 陳克健、陳正佳、林隆基，“中文語句的研究—斷詞與構詞”，中央研究院技術報告，1986年，TR-86-006。
- [50]. 陳品均，“Web 2.0 應用服務策略行動之研究-以 Yahoo!、Google、MSN 為例”，臺灣大學商學研究所碩士論文，2005年。
- [51]. 陳稼興、謝佳倫、許芳誠，“以遺傳演算法為基礎的中文斷詞研究”，資訊管理研究，2000年，第二卷，第二期，頁 27-44。
- [52]. 黃可樂，“維基魅力大探索”，udn數位文化誌，http://mag.udn.com/mag/dc/printpage.jsp?f_ART_ID=14802
- [53]. 詹智凱，“以詞的關聯性為基礎的文件自動分類”，國立臺灣科技大學資訊管理碩士論文，2000年。
- [54]. 數位時代，“完全解讀Web 2.0”，http://www.bnext.com.tw/LocalityView_153
- [55]. 數位時代，“為藝術品下標籤”，http://www.bnext.com.tw/LocalityView_6545
- [56]. 鄧兆旻，“Social Tagging火紅新網路商機逐漸成形”，數位時代，http://mag.udn.com/mag/digital/storypage.jsp?f_ART_ID=89332
- [57]. 盧澤宇譯，“什麼是Web 2.0”，Intelligent Times，http://www.itmag.org.tw/magazine/article_single_138.htm
- [58]. 儲三翔，“學習歷程分析之學習參數改良機制實作與評鑑”，國立花蓮師範學院學習科技研究所碩士論文，2005年。
- [59]. 龔仁文等，“Web 2.0 創新應用案例集：科技化服務新趨勢”，財團法人資訊工業策進會，2006年。

附錄一、中研院平衡語料庫詞類標記集

簡化標記	對應的CKIP詞類標記 ¹	
A	A	/*非謂形容詞*/
Caa	Caa	/*對等連接詞，如：和、跟*/
Cab	Cab	/*連接詞，如：等等*/
Cba	Cbab	/*連接詞，如：的話*/
Cbb	Cbaa, Cbba, Cbbb, Cbca, Cbcb	/*關聯連接詞*/
Da	Daa	/*數量副詞*/
Dfa	Dfa	/*動詞前程度副詞*/
Dfb	Dfb	/*動詞後程度副詞*/
Di	Di	/*時態標記*/
Dk	Dk	/*句副詞*/
D	Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	/*副詞*/
Na	Naa, Nab, Nac, Nad, Naca, Naeb	/*普通名詞*/
Nb	Nba, Nbc	/*專有名稱*/
Nc	Nca, Ncb, Ncc, Nce	/*地方詞*/
Ncd	Ncda, Ncdb	/*位置詞*/
Nd	Ndaa, Ndab, Ndc, Ndd	/*時間詞*/
Neu	Neu	/*數詞定詞*/
Nes	Nes	/*特指定詞*/
Nep	Nep	/*指代定詞*/
Neqa	Neqa	/*數量定詞*/
Neqb	Neqb	/*後置數量定詞*/
Nf	Nfa, Nfb, Nfc, Nfd, Nfe, Nfg, Nfh, Nfi	/*量詞*/
Ng	Ng	/*後置詞*/
Nh	Nhaa, Nhab, Nhac, Nhb, Nhc	/*代名詞*/
I	I	/*感嘆詞*/
P	p*	/*介詞*/
T	Ta, Tb, Tc, Td	/*語助詞*/
VA	VA11,12,13,VA3,VA4	/*動作不及物動詞*/
VAC	VA2	/*動作使動動詞*/
VB	VB11,12,VB2	/*動作類及物動詞*/
VC	VC2, VC31,32,33	/*動作及物動詞*/
VCL	VC1	/*動作接地方賓語動詞*/
VD	VD1,VD2	/*雙賓動詞*/
VE	VE11, VE12, VE2	/*動作句賓動詞*/
VF	VF1, VF2	/*動作謂賓動詞*/
VG	VG1, VG2	/*分類動詞*/
VH	VH11,12,13,14,15,17,VH21	/*狀態不及物動詞*/
VHC	VH16, VH22	/*狀態使動動詞*/
VI	VI1,2,3	/*狀態類及物動詞*/
VJ	VJ1,2,3	/*狀態及物動詞*/
VK	VK1,2	/*狀態句賓動詞*/
VL	VL1,2,3,4	/*狀態謂賓動詞*/
V_2	V_2	/*有*/
DE	/*的、之、得、地*/	
SHI	/*是*/	
FW	/*外文標記*/	

¹ 斜體詞類，表示在技術報告#93-05中沒有定義，即後來增列的。