

出國報告(出國類別：國際會議)

參加第十六屆亞太數位圖書館國際會議
(The 16th International Conference on
Asia-Pacific Digital Libraries, 簡稱 ICADL)
出國報告

服務機關：國家圖書館

姓名職稱：蔣嘉寧編輯

派赴國家：泰國

出國期間：103 年 11 月 5 日至 11 月 8 日

報告日期：103 年 11 月 12 日

參加第十六屆亞太數位圖書館國際會議

(The 16th International Conference on Asia-Pacific Digital Libraries, 簡稱 ICADL)

出國報告

摘要

2014年第十六屆亞太數位圖書館國際會議 (ICADL 2014)於2014年11月5-7日在泰國清邁召開。本館知識服務組蔣嘉寧編輯於今(2014)年8月10日獲得大會通知，所投稿之英文論文(Collapsing Duplicates at Data Entry: Design Issues in Building Taiwan Citations Index)已獲評審通過，將以長篇論文(Full Paper)形式發表，但必須依大會規定註冊並出席發表口頭報告，以便文章得出版刊登於國外學術期刊Lecture Notes in Computer Science，該期刊被DBLP、ZBI Math/CompuServe、IO-Port、EI、ACM Portal及Scopus等著名資料庫所收錄。

該篇論文目的在闡述建置任何的引文索引，在輸入參考文獻建檔時會不斷地遇到重複的引文資料，以及因為引文格式的不同等原因造成的差異性；因此，國家圖書館同仁建構「臺灣人文及社會科學引文資料庫」將所遇到的問題，因應這種本質上的重複性和差異性，所提出的引文系統的設計理念和功能，如何在達到消除重複的同時，並且能夠保持引文的質量。這項理念和功能的原創性和價值在於：建置臺灣引文資料庫和傳統的書目資料庫有很大不同，國家圖書館的「臺灣人文及社會科學引文資料庫」計畫相關同仁，尋求解決消化大量重複的引文書目，將查重整併工作化入到每天的日常建檔流程，以及應用標準書目將各種變異體引文提升引文的品質，採取的措施能有效的改善問題。

本報告簡介出席 ICADL 2014 國際會議相關事項，並摘記所出席的各項技術專題議程的論文報告，最後提出心得與建議。

參加第十六屆亞太數位圖書館國際會議

(The 16th International Conference on Asia-Pacific Digital Libraries, 簡稱 ICADL)

出國報告

目 次

壹、 ICADL 簡介	4
一、 ICADL 簡介	4
二、 出席 ICADL2014 國際會議的動機和過程	5
三、 申請科技部經費補助出席 2014 ICADL 國際會議	5
貳、 第十六屆亞太數位圖書館國際會議 (ICADL 2014)	6
參、 出席 ICADL 2014 會議專題演講及場次摘記	7
一、 ICADL 2014 會議第一天議程	7
二、 ICADL 2014 會議第二天議程	13
肆、 心得與建議	16
伍、 附錄	18

壹、 ICADL 簡介

一、 ICADL 簡介

亞太數位圖書館國際會議(International Conference on Asia-Pacific Digital Libraries，簡稱 ICADL)，為亞太地區數位圖書館研究人員與專家學者提供了國際性的論壇，涵蓋電腦科學、圖書資訊、資訊系統、教育與線上學習、社會科學等領域。此會議始於 1998 年，由亞太地區不同國家輪流主辦，近年來已成為全球數位圖書館相關研究發表的重要園地。

國際上重要的數位圖書館會議在美洲地區為 JCDL(Joint Conference on Digital Libraries)，歐洲地區為 TPDL(International Conference on Theory and Practice of Digital Libraries)，在亞洲地區則為 ICADL。ICADL 國際會議始於 1998 年在香港 (1998 ICADL) 首先舉辦，每年都由各個亞太國家輪流舉辦，並已成為數位圖書館的重要會議之一。歷屆會議舉辦地點如下：

1. 香港 (ICADL 1998)
2. 臺北，臺灣 (ICADL 1999)
3. 漢城，韓國 (ICADL 2000)
4. 班加羅爾，印度 (ICADL 2001)
5. 新加坡 (ICADL 2002)
6. 吉隆坡，馬來西亞 (ICADL 2003)
7. 上海，中國 (ICADL 2004)
8. 曼谷，泰國 (ICADL 2005)
9. 京都，日本 (ICADL 2006)
10. 河內，越南 (ICADL 2007)
11. 巴厘島，印度尼西亞 (ICADL 2008)
12. 黃金海岸，澳大利亞 (ICADL 2010，併同 JCDL 一起舉辦)
13. 北京，中國 (ICADL 2011)
14. 臺北，臺灣 (ICADL 2012)
15. 班加羅爾，印度 (ICADL 2013)
16. 清邁，泰國 (ICADL 2014)

2015 年 ICADL 會議將會在韓國首爾舉辦，大會提請籌辦單位注意，籌辦這樣的國際會議需要為期兩年的規劃期，需要有力而堅強的數位圖書館研究人員和領導者，優秀的機構的支持，以及強大的地區性興趣的團隊等，這些都對會議的成功是至關重要的。

二、 出席 ICADL2014 國際會議的動機和過程

臺灣曾為 1999 年第二屆 ICADL 會議的主辦國，2012 年再度由國立臺灣大學與中央研究院共同主辦，主要著眼於過去幾十年來，在數位圖書館的發展歷程中，JCDL、TPDL 與 ICADL 這三個重要的國際會議除了相當具有指標性外，亦帶動了全球數位圖書館及相關技術的討論與研究。隨著許多國際性與全國性的數位圖書館計畫相繼展開，關於數位內容的新技術也陸續問世，並對於知識的保存與管理造成深遠影響。

2012 年蔣嘉寧編輯曾奉派參加在臺北舉辦之 ICADL 2012 會議，注意到該項國際會議匯集數位圖書館領域中的研究人員、圖書館員、系統開發商、內容提供商和用戶，分享他們的研究和實務，已經構成了一個學者專家的論壇，透過面對面的討論與意見交流收穫良多。因此，2014 年 1 月 27 日接獲 ICADL2014 年大會徵稿 email 通知，不僅強調“數位人文”也接受提交有關數位圖書館等各方面的論文。因為蔣員目前負責的工作項目之一為協調「臺灣人文及社會科學引文資料庫」相關建置工作，切合數位人文的主題，可以藉此國際會議發表論文，介紹「臺灣人文及社會科學引文資料庫」提高曝光率，並且可以與國際學者專家交流。此外，所有錄用的論文將在 Springer 的計算機科學期刊 Lecture Notes in Computer Science 出版，該期刊被 DBLP、ZBMath/CompuServe、IO-Port、EI、ACM Portal 及 Scopus 等著名資料庫所收錄，也是許多學者專家競相投稿的期刊之一。

ICADL 2014 會議論文的提交，重要日期如下：

- 摘要提交截止日期：2014 年 4 月 1 日
- 全文提交截止日期：2014 年 6 月 15 日
- 全文論文錄用通知：2014 年 7 月 15 日
- 最後完稿截止時間：2014 年 7 月 31 日

本次會議大會總共收到 114 篇論文摘要提交審查，其中 79 篇合格接受撰寫論文提交，每篇論文至少由三位評審，平均每篇有 3.73 個評論，經過仔細的評估，大會接受 20 全長篇論文 (23.5%錄取率)，19 短篇論文和 12 篇海報論文。蔣嘉寧編輯依照大會規定提交的全文 10 頁文章，3 位評審皆讚許文章的適切性，並建議容許補充篇幅至 14 頁更能完整敘述，作為完稿提交出版。最終排版後的正式論文為 12 頁的長篇論文(Full Paper)。

三、 申請科技部經費補助出席 2014 ICADL 國際會議

在激烈的論文徵選過程中，蔣嘉寧編輯皆能夠在截止日期前完成大會及論文評審各項相關工作項目要求，最終於 8 月 10 日接獲 ICADL 論文接受通知函(參見附件)，但大會規定作者必須至少 1 人出席，論文方得以出版。由於機會得來不易，館內也無出國預算可支應；因此，蔣員著手爭取科技部補助國內專家學者出席國際學術會議的經費補助。科技部核定補助公函(科部科字第 1030061633 號)於 2014 年 10 月 23 日收文，經館長核定出國並給予公假發表論文。

貳、 第十六屆亞太數位圖書館國際會議 (ICADL 2014)

第十六屆亞太數位圖書館國際會議 (ICADL 2014)，會議在泰國清邁於 11 月 5 日至 7 日舉行，籌辦單位包括泰國的 Chiang Mai University、Khon Kaen University、Suranaree University of Technology 和 Thammasat University。11 月 5 日舉辦 4 項會前的活動，包括輔導學程(Tutorial Session)及研習工作坊(Workshop)需要另外付費報名參加；另外還有特殊對象的活動，包括 Thai Track 及 Program for DC-WIS at ICADL 2014。由於這些活動內容與蔣員工作及研究專長相關度不高，沒有報名參加，僅臚列相關資訊如后。輔導學程(Tutorial Session)包括：(1)社會網絡探勘於數位圖書館中的應用 (Social Network Mining for Digital Library Application)，及(2)CopyCat & MyCat：論文和學術著作剽竊檢查和管理系統 (CopyCat & MyCat: Thesis and Academic Work Plagiarism Checking and Management System)。(2)研習工作坊(Workshop)包括：。(3) Thai Track 係以泰國當地相關議題為主的會前研討會；(4) DC-WIS 係以博士生和資訊系所合作議題的會前研討會(Doctoral Consortium at ICADL 2014 and International Workshop on Global Collaboration of Information Schools WIS 2014)。

ICADL 正式會議於 11 月 6 日至 7 日舉行。2014 年 ICADL 的主題《數位圖書館研究和實踐的興起》(The Emergence of Digital Libraries Research and Practices)，強調“數位人文”，因為它是一個新出現的問題，並且亞太數位人文科學方面是相當具有獨特性和排他性，因此在國際論壇分享特別有價值。會議議程包括宣讀技術性文章、專題演講、工作坊及海報演示，內容圍繞下列議題：例如數位圖書館的基礎設施、機構、詮釋資料、內容、服務、數位化保存、系統設計、科學數據管理、工作流程、實施、界面設計、人機互動、績效評估、可用性評估、典藏發展、知識產權、隱私權、電子出版、文件類型、多媒體、社會/體制和政策問題、使用者社群、以及其他相關的理論課題等。

參、出席 ICADL 2014 會議專題演講及場次摘記

一、ICADL 2014 會議第一天議程

1. Plenary Keynote Speech : Building digital archive systems for historians. Dr. Jieh Hsiang, Director of the Research Center for Digital Humanities, National Taiwan University, Taiwan

會議於 11 月 6 日正式開始，開幕式之後，首場專題演講由國立台灣大學數位人文研究中心主任項潔教授專題演講 Building digital archive systems for historians。主持人為筑波大學教授杉本重雄博士(Dr. Shigeo Sugimoto)介紹過去自 1998 年來與項潔教授有數面之緣，了解他在臺灣數位典藏及圖書館發展為少數開創者之一。

項教授演講的題目為 Building digital archive systems for historians，以他多年參與臺灣數位典藏及圖書館發展，以及投入數位人文相關研究之後，10 多年來與臺灣史的學者合作，史學家的研究在於如何將歷史事實線索聯綴起來；此外，史學研究資料搜尋著重在查全率(recall)以及關係(relationships)，亦即景況或是前後關係(contexts)。項教授一直思考如何利用數位工具從文本中淬煉出供史學研究的線索，讓原本在史籍中翻找事實線索的廢時繁重的工作，利用文獻之詮釋資料加上文本萃取，能夠更快速有效的得到事實和線索，並且增加 context 的系統功能補強現行的檢索機制，讓史學研究得到方便者的資料搜尋工作。

項教授的演講先簡單說明研究臺灣史很困難，在於臺灣的歷史缺中央政府和中央檔案，且歷史檔案資料散佚。他所掌握到的數位典藏包含接手文化部的文化遺產計畫(明清時期臺灣政府的上諭和奏摺)，及臺中圖從日本政府檔案中數位化的地方契書等，並依據這些典藏作全臺地毯式的調查與合作，稱「檢到籃子裡便是菜」的策略，以及項教授過去建置的淡新檔案等。所開發系統的概念稱為，景況或是前後關係(contexts)，從定義上來看就是將史事實線索聯綴起來拼湊還原史實真像。項教授指出系統可以提供的 contexts 包括：文件中的文本景況(textual contexts within documents)、可以解釋演繹的景況(interpretable contexts)，但是無法處理無形的景況(intangible contexts)例如不是資料庫中相關而是和外界環境相關的景況，以及文化的景況(cultural context)、社會的景況(social context)。所以，項教授強調系統功能並不能取代學者做研究，目的在提供用數位工具從文本中淬煉出供史學研究的線索，利用系統可以產出很多很多的線索，透過新興資訊科技如自動提詞(人、地、物等)、詞頻分析、關鍵詞後分類查詢、GIS、視覺化工具等，補強現行的檢索機制，增強史學研究資料搜尋經驗。這個演講的心得是項教授累積多年經驗，接觸臺灣史學研究者，了解使用資料的習慣，需要突破資訊檢索工具精確檢索(precision)不能滿足人文學者遍搜群籍的模式，深思怎樣的工具可以協助數位人文學者更快速有效的得到事實和線索，所提出的系統功能。

2. Session : B1 (2 Full Paper)主題是關於 Linked Data and Knowledge Sharing

接下來參加的第一場技術專題 Session : B1 共發表 2 篇長篇論文(Full Paper)，第一篇是由 Chariya Nonthakarn, Rathachai Chawuthai and Vilas Wuwongse 三位作者的論文發表，題目是 An Application Profile for Linked Teacher Profiles and Teaching Resources. (Full)。這篇論文是屬於 Linked Data and Knowledge Sharing 主題，說明為了教學資源管理系統設計和開發的一個應用規範，重點在如何將教師簡介和教學資源資訊鏈接在一起。他們是以都柏林核心應用規範的新加坡框架為基礎設計開發系統，這個框架用來設計可容許最大互操作性的詮釋資料架構。作者他們所設計的應用規範整合了五個實體 -- 教師簡介、課程教材、學術著作、事件和獎項，這些都被認為對教師本身、他們的學術貢獻和成就而言是重要的資訊。透過使用鏈接開放數據 (LOD) 的原則來呈現這個應用規範，教師就能夠分享他們彼此的教學資源，更進一步還可以合作。這個應用規範已經啟用，並且經過評估是否教師的要求，發現該應用規範確實具有實用性，並且對於詮釋資料架構的設計，是一個有效的機制。

第二篇是由 Knowledge Sharing by Students: Preference for Online Discussion Board vs Face-to-Face Class Participation. 作者為 Shaheen Majid, Pang Yang, Huang Lei and Guo Haoran. (Full)。本篇論文在研究學生參與面對面線上討論的喜好，以及發掘哪些是激發或阻礙知識共享的因素。作者利用問卷調查新加坡兩所公立大學 149 位研究生進行他們的研究，結果發現關於知識共享，多數學生 (82.5%) 偏好課堂上面對面的討論，而且多數人一致認為的因素包括：授課老師的鼓勵、挑戰性的議題，以及來自於老師和同學發人深省的提問，這些因素都很可能會激勵他們參與。至於激勵參與網上討論的因素包括：更多的時間來組織思路、減少緊張、並更有彈性地存取線線討論區的資訊。阻礙面對面參與的因素包括：不好意思發表和無趣的話題。另一方面，線上討論主要的障礙為界面設計不好用和時間上的限制。

3. Session : A2 (1 Full, 3 Short Papers)主題是關於 Digital Repositories and Tools

第一篇論文 Quality Assurance Tool Suite for Error Detection in Digital Repositories. 作者為 Roman Graf and Ross King. (Full)。本文作者指出自動掃描圖像典藏的數位化的工作流程很容易出錯，需要進行品質保證的工作。他們開發的自動化的品質保證工具，旨在檢測可能的品質問題，提供文獻影像圖片集的處理。所開發的工具主要的貢獻在於單一套件中組合了多種圖像處理工具，支援各種不同的錯誤檢測方案。

該工具套件包括：(1) 火柴盒工具：利用 SIFT 特徵提取技術，採近乎重複方法精確的檢測文獻影像圖片集。作者此處所提到的 SIFT 特徵提取技術很有意思。SIFT 英文原文為 Scale-invariant feature transform(簡稱 SIFT)，一般翻譯為尺度不變特徵轉換，是一種電腦視覺的演算法，其專利擁有者為加拿大英屬哥倫比亞大學。SIFT 用來偵測與描述影像中的局部性特徵，它在空間尺度中尋找極值點，並提取出其位置、尺度、旋轉不變數，

應用範圍包含物體辨識、影像追蹤和動作比對等。局部影像特徵的描述與偵測可以幫助辨識物體，基於這些特性，它們是高度顯著而且相對容易擷取，在母數龐大的特徵資料庫中，很容易辨識物體而且鮮有誤認。(2) 手指的檢測工具：對於在掃描過程中，不小心出現在從數位化圖像集時出現的手指影像要如何處理呢？作者指出，這個工具使用的處理技術包含邊緣檢測、本地圖像資訊萃取、和透過掃描品質的推理運算所作的分析，可以利用這個工具作自動檢測誤入圖像中的手指影像。(3) 裁剪錯誤檢測工具：一般影像發生裁剪問題的檢測，例如文本偏移到圖像頁邊界、不需要的多餘頁邊界、或是從前頁誤植過來不想要的文本。

這項工作的另一重要貢獻是清楚定義了數位化圖像集的品質保證工作流程，並且能夠自動執行錯誤檢測。所提出的工具套件檢測出影像掃描數據中描述的錯誤，並提出了他們額外的人工分析和收集清理的結果，提出評估數據和類似的性能和精度特性的統計概覽。該分析的結果證實了作者原先的假設，即一個自動化的方法能夠檢測到的錯誤，而且其結果具有可靠的品質，從而使得對大型數字化項目的質量控制是可行和可負擔得起的過程。蔣員問作者為何會有重複和手指的影像，了解到他們進行大量掃描時，由於掃描器速度很快，很容易有重複的頁數，而且翻頁時手指也會遮住部分頁面，而即便是這個方法去除了手指的影像，但也同時去除了下面被遮住的頁面部分。對於這樣的損失，他們也沒有多餘的人力再重新掃描補救這個頁面，他們只能無可奈何的接受這項事實。

第二篇論文 *Comparative Study of Digital Repositories: A Case Study of DESIDOC Journal of Library & Information Technology*。作者為 Alka Bansal and Dipti Arora. (Short)。本篇論文大致上論述 *DESIDOC Journal of Library & Information Technology (DJLIT)* 性質是一個具備同儕審查、提供開放存取的國際性期刊。2002 年初，為了提供 DJLIT 的每一篇文章的全文和詮釋資料給使用者，利用紐西蘭開發的開放性軟體-- GSDL (綠寶石數位圖書館) 建置。DJLIT 數位圖書館所有的卷期資料，當時只透過內部網路(intranet)提供遍布全國的使用者使用，但沒有在網際網路上。2007 年出，DJLIT 使用加拿大 Public Knowledge Project (PKP)所開發的開放性 OJS 軟體，不僅將出版流程自動化，並且儲存文章的全文和詮釋資料，進而典藏所有的卷期作成數位檔案，並且可開放網際網路存取使用。由於同樣的物件同時可以在兩個平台尚存取，因此試圖研究這兩個資料庫的主要功能，包括建檔過程、搜尋和檢索等，以了解是否 OJS 所建置的資料庫可以取代 DJLIT 數位圖書館，而不會影響任何計有的功能，以及了解有 OJS 資料庫有哪些功能需要修改。蔣員過去在加拿大英屬哥倫比亞大學修業期間，曾經安裝過 GSDL (綠寶石數位圖書館)並且提出評估效益的經驗；2002-2009 年參與加拿大 Public Knowledge Project (PKP)研究計畫，為開放性 OJS 軟體的創始開發小組成員，對於這篇論文特別有興趣。可惜作者位能夠出席發表。

第三篇論文 *Personal Digital Libraries: Keeping Track of Academic Reading Material* 作者為 Mohammed Al-Anazi, Annika Hinze, Nicholas Vanderschantz, Claire Timpany and Sally Jo Cunningham. (Short)。本文討論追蹤學術閱讀材料的選擇加上引入個人化數位圖書館

工具可能的方案。作者不僅結合開放性軟體 Zotero 和 Greenstone(綠寶石數位圖書館)並且擴充其功能，讓使用者可以很容易地下載學術閱讀材料，並且納入到作者所開發的組合系統中。作者利用開發 Zotero 開發 Currently Reading、Have Read、To Read 三項功能納入 Greenstone 介面上，提供使用者利用。作者利用這樣的雛形系統進行一項小型的使用者易用性研究(usability study)，只有 4 人參與以日記加上訪談的研究方法，結果有三：(1)功能太繁複；(2)閱讀 PDF 時的註記要方便；(3)希望每次進入文件時游標停在最後閱讀的地方。

第四篇論文 Altmetrics for Country-Level Research Assessment. 作者為 Hamed Alhoori, Richard Furuta, Myrna Tabet, Mohammed Samaka and Edward A. Fox. (Short)。學術交流和科學的版圖刻正不斷地發生變化，政策制定者和研究資助機構也試著尋找方法來評量研究的綜效性(comprehensive impact of research)並吸取其他國家的研究經驗。最近，altmetrics 已被用於評量研究活動的更廣泛的影響。在本文中，作者的 altmetrics 研究植基於發掘國家層面的影響，而發現 altmetrics 所支援的研究評量可以適用於所有的國家。作者比較 altmetrics 和其他幾個傳統的指標，發現國家層面 altmetrics 和出版品數量、引用的次數、h 指數和國內支出總值 (GDP) 佔研究和開發 (GERD) 的比例之間有顯著關係。特別是，作者還發現 2010 年和 2014 年間發表的文章數量的 altmetrics 明顯地逐年增加。

4. 下午的第三場技術專題 Session : A3 (1 Full, 3 Short Papers)都是關於 Data Management and Classification 的主題。

第一篇論文 Cost Evaluation of CRF-Based Bibliography Extraction from Reference Strings. Naomichi Kawakami, Manabu Ohta, Atsuhiko Takasu and Jun Adachi. (Full)。本篇文章敘述數位圖書館的有效利用仰賴書目資料庫的持續維護。特別是，學術論文的參考文獻中充滿了有用的書目資訊，例如作者姓名和文章的標題。因此，作者提出使用條件隨機場 (CRF) 的鑑別式機率模型，條件隨機場 (conditional random fields, CRFs)，是一種鑑別式機率模型，是隨機場的一種，常用於標注或分析序列資料，如自然語言文字或是生物序列。作者利用條件隨機場 (CRF) 從論文的參考書目資訊裡自動提取字符串的方法。然而，想要訓練 CRF 實現高的提取準確精度，至少要使用幾百條引用字符串。為了縮短訓練進程，作者的研究建議可以使用活性採樣和類訓練數據，以減少訓練數據量。此外，作者也提出她們實驗相關訓練費用的評估。

第二篇論文：Finding Co-occurring Topics in Wikipedia Article Segments. Renzhi Wang, Jianmin Wu and Mizuho Iwaihara. (Short)。作者指書維基百科是全球最大的線上百科全書，其中的文章，形成豐富的知識和語義資源。相同的主題在不同的文章指出，這些條款都涉及到彼此的話題。找到這樣的共現的主題是有用的，可以提高查詢和聚類的準確度，並且還可以對比相關的文章。現有的主題定位工作和主題相關的檢測是基於長期發生。在作者的研究中，討論整合利用潛藏式狄利克雷分配 (LDA) 存在的第細分潛在主題，以

檢測相關的話題。作者還研究如何段上的近似，從段排序和超鏈接產生的，應當納入話題檢測和校準。實驗數據證明作者的方法可以找到和區分三種共現的類型。潛藏狄利克里分配的主要精神是將每一篇文件的潛藏主題機率分布以一個彈性的方式定義，將每一篇文件所對應的潛藏主題的機率分布視為一個隨機變數，而這個隨機變數是以一個狄利克雷分配來描述，因而得以對任何一篇未見過的文件，透過狄利克雷分配中取樣出屬於這一篇文件的潛藏主題機率分布，而產生出該文件中的可能主題詞。**潛藏式狄利克雷分配 (LDA)** 的計算方法可用於預測未見過的文件中其潛藏語意的分析，以及所運用的模型參數，不會伴隨著訓練語料中文件數量的增加，而呈現線性成長。近年來，以潛藏狄利克里分配為題材的研究與日俱增，被廣泛的使用於各種領域之中，例如影像處理(Image Processing)、圖樣辨識(Pattern Recognition)、資訊檢索(Information Retrieval)等。在語言模型上，陸續有學者提出以潛藏狄利克里分配為基礎，將更多資訊或更複雜的模型架構引入，使得新的模型可以更加符合需求或更符合實際資料狀況。

第三篇論文：Parameter-Free Imputation for Imbalance Datasets. 作者為 Jintana Takum and Chumpol Bunkhumpornpat. (Short)。在資料探勘與機器學習的領域中常應用分類運算，當訓練資料集中的目標類別分部越均勻時，通常會有比較好的分類效能；然而在實務應時中，卻經常遇到目標類別分佈不平衡的問題。通常分類運算會傾向將所有訓練資料預測為多數類別，卻忽略少數類別，而少數類別的預測效能通常是決策者重視的部分。作者指出分類不平衡的運算方式旨在改善少數類別的精確度的問題，而數值填補是設定數值屬性的缺值填補方式，以代替丟失的值。傳統上，不平衡運算和數值填補都是分開處理。此外，傳統的方法填入少數類別值對於處理不平衡數據集是不夠的。作者在本文中提出了一個新的方法：利用無參數填補運算的方式處理不平衡數據集，作者說明他們的算法忽略平均情況，以避免過度擬合的問題。因此，對不平衡數據集的實驗結果顯示，他們的數值填補用於分類不平衡的作法優於其它技術。

第四篇論文：An Automatic Library Data Classification System Using Layer Structure and Voting Strategy 作者為國立中興大學圖書資訊所教授郭俊桔(June-Jei Kuo)博士發表。(Short)。本文論述了傳統的一本書分層分類系統的問題，採用不同的分類之間的互補特性而設計一種投票策略，提出了一個兩層的圖書分類體系。此外，作者利用從一所大學圖書館館藏的學位論文和書店的電子書的書目集作為訓練和測試語料庫。首先，將論文和書籍的分類代碼是設定為標竿。每個論文由各種詮釋資料項目組成，例如題名、作者、目次、摘要或參考文獻等。作者利用了這些組合進行研究，為了理解所有組合的分類效果，並且提出最佳組合的建議。結果發現從摘要和目次中提取的特徵用於文件分類最有用。另一方面，為了獲得最佳的分類性能，為了建構兩層的圖書分類系統，作者研究結合不同的分類因子，並且提出最佳組合的建議。

5. BEST PAPER AWARD

第一天(11月6日)下午 Best Paper Award,大會宣布獲獎論文的名單及評選的過程。本次會議大會總共收到 114 篇論文摘要提交審查,其中 79 篇合格接受撰寫論文提交,每篇論文至少由三位評審,平均每篇有 3.73 個評論,經過仔細的評估,大會接受 20 長篇論文 (23.5%錄取率),19 短篇論文和 12 篇海報論文。獲獎論文的產生是從入選的 20 篇長篇論文中,依照評審意見經過修正後,再由所有的評審投票選出 5 篇優秀論文,再由其中選出 1 篇最佳論文。紐西蘭的學者 Haley Littlewood, Annika Hinze, Nicholas Vanderschantz, Claire Timpany and Sally Jo Cunningham 等人共同撰寫獲得最佳論文,論文題目為 A Log Analysis Study of 10 Years of eBook Consumption in Academic Library Collection,蔣嘉寧的論文為獲選的 4 篇優秀論文(outstanding paper)之一,並獲獎狀如附件二圖 2。

6. 海報發表場次(Poster Session)

每一篇海報論文只有 5 分鐘的發表,作者代表必須很快的把主題說清楚。總計有 9 篇。

- Poster 1 Calligraphy Character Synthesis from Small Sample Set. Kai Yu and Zhenming Yuan.
- Poster 2 Institutional Repositories in Thai Universities. Namtip Wipawin and Aphaporn Wanna.
- Poster 3 A Comparative Study of Key Phrase Extraction for Cross-Domain Document Collections. Supaporn Tantanasiriwong, Choochart Haruechaiyasak and Sumanta Guha.
- Poster 4 Development of a Book Recommendation System to Inspire “Infrequent Readers”. Shuntaro Yada.
- Poster 5 A Comparison of Dimensionality Reduction Algorithms for Improving Efficiency of PromoRank. Metawat Kavilkrue and Pruet Boonma.
- Poster 6 Library Book Recommendations Based on Latent Topic Aggregation. Shun-Hong Sie and Jian-Hua Yeh
- Poster 7 Query Formulation for Action Search by Bootstrapping. Yoshinori Kitaguchi, Hiroaki Ohshima and Katsumi Tanaka.
- Poster 8 Transforming Publication List of LIS-TW in Author Identification Services to Open Linked Data for Mobile Application. Chao-chen Chen, Hong-shiu Liang and Shun-hong Sie.
- Poster 9 Documentation and Dissertation of Kotagede’s Local Wisdom by Kotagede Heritage Library: The Best Practice. Nur Cahyati Wahyuni, Lista Rantika and Imam Zakaria.

二、 ICADL 2014 會議第二天議程

第二天的專題演講及場次摘記

1. Plenary Keynote Speech : Digital Archive Project “Michinoku-Shinrokuden” of The 2011 Great East Japan Earthquake Disaster Dr. Akihiro SHIBAYAMA Institute of Disaster Science, Tohoku University, Japan

第二天的專題演講，由日本東北大學災害科學研究院的 Akihiro SHIBAYAMA 博士演講，題目為：數位檔案計畫” Michinoku-Shinrokuden” 2011 年東日本大地震災害。這個東日本大地震的數位檔案項目計畫 “みちのく震録伝” (Michinoku-Shinrokuden)由東北大學主導，結合約 102 所日本公共圖書館、工業界、政府、私人企業和學術界的機構協作下，收集在東日本大地震中所有的記憶、記錄、案件、知識等，Michinoku 其實指得是宮城、福島、岩手等受災區，約寬 150 公里、長 450 公里的區域，當時震幅達到 9.0，上下震盪約 8 公尺。一開始介紹 2011 年 3 月 11 日 14:46 分大地震當時民眾錄下的影像，這個項目，以該地震的中心受災地區擴及周邊距離，記錄歷史災難的東日本大地震。東北大學主導之下，收集了從各種角度的記憶和記錄，包括錄影、文字記錄、錄音、照片等等，從案例研究的基礎上去發現資料，到目前為止已經過了約 3 年，該中心仍然透過災區的公共圖書館每周招集資料蒐集會議，並展開跨部門的研究，以協助東日本大地震實際災難狀況的闡釋(例如災難前後景況的比對)和提供重建的知識(例如利用 laser measurement)作 3D 的重建。而且，日本的著名藝人、音樂家也發表朗讀和音樂會喚起民眾不要忘記這場觸目驚心、刻骨銘心的大災難，民眾記憶和曾經發生的感人的故事。這些努力是為了發展學術的措施和巨大災難的管理，可以採取的措施，重點是未來發生在東海，東南海和南海地震時能夠及時掌握與因應。最終，SHIBAYAMA 教授說這項大數據工作其實目的在記取教訓、傳承知識，希望未來能防患於未然。

2. Session : A4 (3 Full Paper)主題是關於 Scholarly Document Repositories

第一篇論文 OA Policies and the Sustainability of Digital Libraries of Scholarly Information.作者為 Gobinda G. Chowdhury. (Full)。Chowdhury 教授最近出版了一本書，這篇文章是抽取自書中的內容而發表。作者介紹學術交流的近況，由各國政府機構和科研資助機構(澳洲、美國 NIH、英國、美國、英國的機構典藏計畫主持之機構 JISC 等)所推出的 OA (開放存取) 的政策入手，討論了具體的 OA 政策及其執行如何影響了學術資訊的交流，而且從機構典藏的經濟、社會和環境的與續性來看，與其零散的建置機構典藏，耗時費力又不討好，建議成立集中式的全國性機構典藏。蔣員茶敘時與 Chowdhury 教授討論機構典藏的定義和起源，機構典藏企圖建立各校自己的鏈結、點閱、下載等排行數據作為評鑑指標，以及「提早公開、更多使用者」的現象是否至今仍然成立等相關議題。

第二篇論文 *Evaluating the Academic Performance of Institutions within Scholarly Communities*. 作者為 Lili Lin, Zhuoming Xu, Yuanhang Zhuang and Jie Wei. (Full)。作者指出對於學術機構的評估，現今的作法多半是進行同儕訪視評估或採用文獻計量指標評比。有鑒於，同儕訪視評估動用許多人力且耗時費日，而現有的文獻計量方法可能產生偏頗的評估結果，因為計量的結果，到目前為止，沒有可靠的綜效型模式能夠反應出許多學術力的重要因素。因此，作者在這篇論文中討論她們的研究，提出了一個因子圖示的機構排名模式(a factor graph-based institution ranking model)。作者利用機構本身的資訊 (定量和定性資訊) 和學術網路資訊 (例如：協作強度) 進行研究，選擇最知名的美國新聞與世界報導 (U.S. News & World Report) 的調查結果作為基礎事實 (ground truth)，然後將美國的圖書館和資訊科學 (LIS) 研究領域的機構排名進行個案分析，作者認為她們的實驗結果，與現有的文獻計量學方法相比，她們的方法應該可以為機構評估提供一個更好的選擇。但是，現場也有學者提出美國新聞與世界報導 (U.S. News & World Report) 的調查結果應該也有一些質量上的分析產生，是否適合直接拿來作為基礎事實 (ground truth) 有待商榷。

第三篇論文 *Axis-based Alignment of Scholarly Papers and Its Presentation Slides Considering Document Structure*. 作者為 Yuhei Kawakami, Atsuto Nishida, Toshiyuki Shimizu and Masatoshi Yoshikawa. (Full)。作者指出大多數研究人員用 PowerPoint 作為發表論文文稿內容的演示，他們希望利用 PowerPoint 這種大綱式的內容，透過檢索並通過網站瀏覽文件和 PowerPoint，讓閱讀論文的使用快速的吸收作者發表論文中所要表達的重點。作者提出一種方法來比對 PowerPoint 的大綱和各個 slides 的內容相較於論文的段落，在先後次序上和涵蓋內容的相似性。這篇論文在報告他們手動創建一個小的數據集，及所進行的一項實驗，及驗證該方法的有效性。

3. Session : B5 (1 Full, 3 Short Papers) 包含了 Information Retrieval and Search Methods 和 Metadata and Ontologies 這 2 個主題的論文

第一篇論文為蔣嘉寧的長篇論文屬於 Information Retrieval and Search Methods，論文題目為 *Collapsing Duplicates at Data Entry: Design Issues in Building Taiwan Citations Index*. Chia-Ning Chiang. (Full)。蔣嘉寧編輯以英文發表論文--該篇論文闡述建置任何的引文索引，在輸入參考文獻建檔時會不斷地遇到重複的引文資料，以及因為引文格式的不同等原因造成的差異性；因此，國家圖書館同仁建構「臺灣人文及社會科學引文資料庫」將所遇到的問題，因應這種本質上的重複性和差異性，所提出的引文系統的設計理念和功能，如何在達到消除重複的同時，並且能夠保持引文的質量。這項理念和功能的原創性和價值在於：建置臺灣引文資料庫和傳統的書目資料庫有很大不同，國家圖書館的「臺灣人文及社會科學引文資料庫」計畫相關同仁，尋求解決消化大量重複的引文書目，將查重整併工作化入到每天的日常建檔流程，以及應用標準書目將各種變異體引文提升引文的品質，採取的措施能有效的改善問題。

第二篇屬於 Metadata and Ontologies 主題,發表的短篇論文題目是 Transformation of DSpace Database into Ontology. 作者為 Humaira Farid, Sharifullah Khan and Mohammad Younus Javed. (Short)。作者是以本體論的觀點建構軟體專案管理的知識庫系統,他指出本體(ontology)的特性是能夠用來控制獨立的數據源之間的資訊共享和交換。DSpace 的是機構典藏的開放軟體。它採用了一個傳統的資料庫來儲存詮釋資料,並透過詮釋資料提供數位文建相關的詳細描述數據。作者認為 DSpace 開放軟體容許開發一些進階功能,認為可以轉變 DSpace 的資料庫到以本體論的觀點建構的軟體資料庫,進而可以共享和重複利用資訊。但是,現有的轉化系統是不能夠準確地變換的 DSpace 數據庫到本體,因為它的模式是間位模式(meta-schema)。因此,作者提出的作法是發展一個中繼的資料庫,首先從 DSpace 資料庫轉換成以關聯式資料庫模式的中繼資料庫,然後把它轉換為一個本體。該系統已經進入實作階段,結果證實整個的過程是可以正確無誤轉化。

第三篇屬於 Information Retrieval and Search Methods 主題,題目是 Dynamic Web Application for Managing and Searching Antique and Art Object Information. 作者是 Areerat Trongratsameethong and Pongkwan Saejao. (Short)。作者來自計算機科學的背景,但是強調古董和藝術品資訊在考古、歷史、文化的領域中越來越重要。尤其是博物館中,這些寶貴的資料不斷增加,它需要有效的系統來處理可用的知識,提供後代人了解當時的文化。因此,本研究的目的是設計一個資料庫來存儲古董和藝術品資訊,並設計動態網頁來管理和搜尋資訊。該系統由兩個主要部分組成:1) 資料管理和 2) 進階檢索。資料管理部分是給博物館的工作人員來存儲、更新和刪除資料。進階檢索是提供大眾利用關鍵字、器物名稱、類型、材料、時期和藝術風格的結合,以有效地檢索資訊,他們不僅提供布林邏輯運算符“AND”和“OR”,還提供“LIKE”的功能,讓使用者有更寬廣的檢索結果可以參考。

第四篇屬於 Information Retrieval and Search Methods 主題,論文題目是 Information Retrieval skills of Research Scholars of Social Sciences: A Study.作者為 M.R.Murali Prasad. (Short)。這篇論文主要研究在印度海得拉巴地區社會科學學者的資訊檢索技能。這些數據是利用調查問卷收集,總計 201 位社會科學的研究學者填答問卷,詢問他們在進行搜尋資訊的時候,習慣利用的管道,已及當時首選的搜索字詞、概念的理解、關鍵詞搜索、題名的布林邏輯運算符的熟悉度,以及能夠確切的找到參考文獻。作者認為其研究結果真實地反應出社會科學學者的資訊利用行為。

4. Session : A6 (1 Full, 2 Short Papers) 3 篇論文都是關於 User Skills and Experiences 主題的論文。

第一篇論文 The Design of an Information Literacy Game. 作者為 Yan Ru Guo and Dion Hoe-Lian Goh. (Full)。作者指出教育體系中、在工作場所並在日常生活中越來越關注資訊素養 (IL)。學生需要資訊素養,尤其是對大學生特別重要,因為他們在課業上不斷地尋找和匯集多個來源的各種知識,以便專注的消化吸收後撰寫學期論文。鑑於學生的課業

非常忙碌，面對面的討論永遠只限於一小部分的學生，對圖書館員而言，如何利用新興的數位技術提供一個更好的圖書館服務的機會。因此，作者採用參與式的設計方式，設計了一個數位資訊素養線上遊戲。然後根據 Game Flow 模型分析參與設計工作坊 (participatory design workshop) 所收集到的數據，並解釋他們所設計的一個雛形系統的相關功能及報告所觀察到影響學生資訊素養的效果。

第二篇論文 A Multi-lingual and Multi-cultural Tool for Learning Herbal Medicine. 作者為 Verayuth Lertnattee and Chanisara Lueviphan. (Short)。作者指出在泰國，傳統醫學是來自於幾種文化，例如印度傳統醫學 (阿育吠陀)，與中國傳統的中藥。中草藥是泰國傳統醫學的重要組成部分。執業的傳統醫學藥師，需要熟悉草藥的知識；但是，對於醫學學生而言，要在短時間內熟悉的好多種藥材是相當的困難；而且，一種草藥還有多種語言和名字。以蘆薈的名字為例，作者列出泰國有 3 種語文、新加坡有 4 種、寮國有 3 種、中文有 3 種。為了提昇學生有效率地認識中草藥的知識，並符合多元文化學習，作者開發了一個多語言草藥的工具。KUIHerbRx2014 是以草藥為內容的一個網頁學習工具，而且它支援協作學習，用科學的方法提高學生學習多元文化的草藥，包含多種語言和多種文字藥草的名字。所開發的系統功能還包括：收集、提供新的意見、票選意見、提供建議等互動功能，提升學習中草藥的知識和技能。

第三篇論文 Dimensions of User Experience and Reaction Cards. 作者為 Tanja Merčun and Maja Žumer. (Short)。作者指出類似數位圖書館這樣的資訊系統，其成功的因素很多，但是良好的用戶體驗是決定因素。目前這方面研究尚屬於相對較新的領域；因此，構成一個用戶體驗的概念不具有很強的框架，並且也不容易測量。作者報告她利用 Microsoft 公司的反應卡片研究方法，改良以清單方式作研究工具。字卡上的字都是形容詞，她在使用者體驗系統功能操作後，立即請使用者勾選一項反應形容詞，並用質化分析確定有意義的概念，提出此法可以做位進行分析用戶體驗的新方法，並提出未來研究的方向可 (1) 同樣的字卡用於不同的資訊系統進行使用者體驗測試，(2) 增加更多的形容詞對同一個系統進行使用者體驗測試等。

閉幕式

閉幕式簡短而隆重，由杉本重雄教授致感謝詞，泰國主辦單位致詞，即由泰國團隊交棒給下一屆的韓國承辦單位代表，並在大合照互道再見。

肆、心得與建議

- 一、國家圖書館早自行政院推動國家基礎建設(NII)以來，就積極參與遠距圖書服務項目，致力建置許多數位圖書館和數位典藏，甚至數位學習網站為推動臺灣資訊國際化而努力不懈，對於臺灣在國際學術社群的能見度，建議本館應繼續派員參與各項相關會議活動。

- 二、 日本大地震他們收集整理的資料，各方協力建置，希望能夠記錄並教育大家不要忘記 311 地震災難，學習有關於地震災難的教訓，整合資訊為未來防災作準備。對於他們有系統的進行收集的工作、運用先進的電腦技術去呈現災難前後的景象，這種理性的處理方式很佩服，而文化上的軟性訴求，例如音樂會和朗讀等去撫育受傷的心靈也喚起人民不要忘記這個創傷，非常感動。
- 三、 藉由此次參與 ICADL 年會，我們有機會瞭解數位圖書館相關技術的發展概況，也有一些論文是和引文分析相關論文值得參考。
- 四、 觀察本次年會各國與會代表所發表的論文，在質量上都很不錯，收穫很多。台灣在此次會議，開幕式之後即由臺灣大學項潔教授作專題演講，其實臺灣數位圖書館方面發展很不錯，美中不足為文章被收錄的篇數不多；全程會議中有 1 篇長篇論文，1 篇短篇論文，及 2 篇海報論文在研討會上發表，建議能夠有機會呼籲相關學者專家投稿，有更多的論文讓世界各國瞭解台灣數位圖書館方面的發展，分享台灣在數位典藏方面經驗。

伍、 附錄

附件一、ICADL 投稿英文摘要

Collapsing Duplicates at Data Entry: Design issues in building Taiwan Citations Index

Chia-Ning Chiang, National Central Library, Taipei, Taiwan

Purpose - When building citation indexes, citation entries are found duplicated and variant by nature at the time of data entry. This paper's aim is to describe the design concept and functions supported by the Taiwan Citation Index – Humanities and Social Sciences system in collapsing the duplicated entries while also improving the quality of citations.

Design/methodology/approach - Large amount of duplication sets, taken from the legacy of THCI and TSSCI, are loaded in a separate pool in the TCI system, while more duplicates added online from everyday data entry. Finding efficient and effective solutions to harness everlasting duplicates on a daily basis is the major challenge. This paper describes the analyses in data, database structure, data entry process, and batch processing functions. The analyses showed that the composition of manual editing and re-applying better quality record during the collapsing process, as well as an overnight batch process of merging all collapses is inevitable.

Findings –The findings showed that the system needs to support functions for collapsing citations and improving record quality on-the fly. Four design issues were identified. First, a working area is needed to support on-the-fly collapsing of identical citations found in both duplication sets and online data entry. In addition, a batch processing to merge all collapses performed at the end of a day is required. Second, the identification of standard record (i.e., the Periopath Index record) that has better quality than any other duplicates is required. Third, an Apply-Standard function needs to be devised for controlling the variant citations. At the time of collapsing, data entry person would search for variants, identify the standard record, select identical citations, and click the Apply-Standard to point the variants to the standard record. The variant records are all pointed to the standard one, and displayed in the form and content of the standard thereafter.

Practical implications - The design not only accommodates the collapse in two different areas in the TCI system. In a nutshell, search, select, apply, collapse, aggregate collapses are major functions needed in achieving the ultimate goal for efficiency and effectiveness.

Originality/value - The paper shows that building a citation database like the Taiwan Citations Index is very different from conventional bibliographic databases. This paper describes the initiatives taken in finding an efficient way to digest the large amount duplicate into a day-to-day process, as well as an effective way to apply the standard record in place of variant citations for the betterment of citation quality.

附件二 來往信件摘錄

一、 徵稿獲得評選錄取為長篇論文的信件

來源: ICADL2014 <icadl2014@easychair.org>
收信: Chia-Ning Chiang <cnchiang@ncl.edu.tw>
標題: ICADL2014 notification for paper 19
日期: Fri, 18 Jul 2014 18:31:08

Dear authors,

We are happy to inform you that your paper has been accepted to appear as a FULL paper in the proceedings of ICADL2014 conference.

We have enclosed the reviewer comments for your use at the end of this message. Please take the opportunity to improve your submission based on the provided feedback. Especially please pay attention to the paper formatting (according to Springer's LNCS guidelines) and to the grammatical correctness of your paper.

The final, camera-ready version of the paper must be submitted by AUGUST 8th, 2014 (hard deadline). Your final full papers may be 10-14 pages long. Detailed preparation and submission instructions are available at http://icadl2014.cmu.ac.th/page-paper_submission/

Also, please remember that early bird registration for ICADL2014 ends on August 15th, 2014. For a paper to appear in the conference proceedings at least one co-author must register and attend the conference. The way to register is explained at: http://icadl2014.cmu.ac.th/page-registration_information/

We are looking forward to seeing you in Chiang Mai,

Sincerely,
ICADL2014 PC Chairs

二、徵稿獲得評選錄取 email

來源: Kulthida Tuamsuk <icadl2014pc@gmail.com>
收信: cnchiang <cnchiang@ncl.edu.tw>
標題: ICADL2014-Certificate of Paper Acceptance
日期: Sun, 10 Aug 2014 12:40:57
附檔: Certificate_of_ICADL_2014_Paper_19.pdf(146 k)

Dear Chia-Ning Chiang,

Congratulations!

Attached please find a Certificate of ICADL2014 Paper Acceptance for your submitted paper no. 19.

Thank you for your contribution to ICADL2014. See you in Chiangmai on November 5-7 2014.

Best regards,
Kulthida Tuamsuk
PC co-chair

附件二 論文錄取證書、獲獎證書



Certificate of Paper Acceptance

This is to certify that the following paper has been reviewed and accepted for an oral presentation at **ICADL 2014** - The 16th International Conference on Asia-Pacific Digital Libraries, which will be held on 5-7 November 2014 at Chiangmai, Thailand.

Author(s): Chia-Ning Chiang

Title: Collapsing Duplicates at Data Entry: Design Issues in Building Taiwan Citations Index.

The authors are cordially invited to participate and present your paper at the conference. All papers of the authors who have registered and presented at ICADL 2014 will be published in **Springer's Lecture Notes in Computer Science Proceedings** which is indexed in SCOPUS.

Please be sure to register by August 15, 2014.

Date of Issue: August 10, 2014.

Kulthida Tuamsuk

Dr. Kulthida Tuamsuk
ICADL 2014 Program Co-chair
Email: icadl2014pc@gmail.com

圖 1 論文錄取證書



圖 2 蔣嘉寧(Chia-Ning Chiang)論文獲獎證書

附件三 ICADL 會議網站、論文集與大會手冊等相關資料

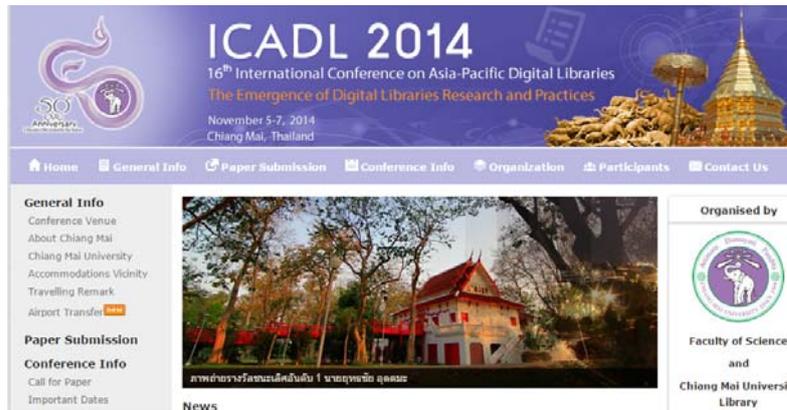


圖 3 2014 ICADL 會議網站

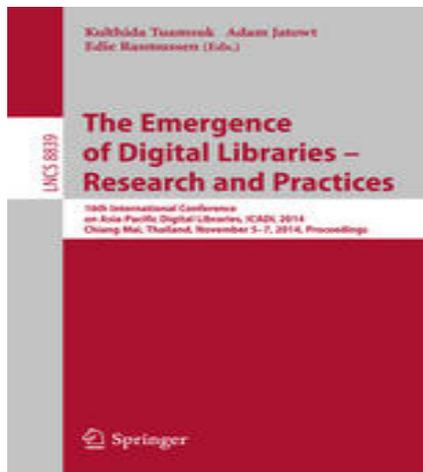


圖 4 大會出版的論文集



圖 5 蔣嘉寧發表論文