

基於社群問答技術的圖書館 FAQ 問答系統之初探

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology\

郭俊桔

June-Jei Kuo

通訊作者：jjkuo@dragon.nchu.edu.tw

國立中興大學圖書資訊學研究所副教授

Associate Professor, Graduate Institute of Library and Information Science, National Chung Hsing University

【摘要 Abstract】

為提高讀者的滿意度，圖書館通常會提供常見問答（FAQ）以解決讀者的疑問。然而，囿於圖書館人力不足，常見問答的數量並不多，讀者無法找到問題的解答而感到挫折，即使找到想問的問題答案，但 FAQ 的答案內容也可能無法滿足讀者的需求。社群問答技術在自然語言處理領域受到高度關注，藉由此技術，讀者的問題不必由網站經營者自行回答，而可由社群內的網民代為回答，答案經過篩選後，將有用的答案傳給讀者，可提高提問者的滿意度。圖書館是以服務讀者為目的，提供技術服務與讀者服務，針對讀者的問題，除圖書館提供 FAQ 解答外，透過網民提供解答，形成一個圖書館社群。因此，本研究導入社群問答技術，將圖書館常見的問題和答案予以分類，建立圖書館社群 FAQ 問答系統。實驗結果顯示此問答系統比傳統 FAQ 問答系統的讀者滿意度（五分量表）由 2.15 提高到 3.05。

To increase the readers' satisfaction each library usually provides frequent question answering set to answer the related questions. Nevertheless, due to the lack of manpower there are few question

answering pairs in the library website. Thus, the question providers can not usually obtain satisfactory answers. Even worst, there is not such question in the FAQ, they cannot obtain any answer. Those will frustrate the readers not to utilize the library website or libraries. As the community FAQ technology is emphasized in natural language processing era, whenever a question is provided, the answers can be given by not only the website managers but also the members in the same community. Through the quality evaluation for the answers, only the best answers will be provided to the question providers which can definitely enhance their satisfaction. As the purpose of libraries is to provide services for the readers, those services will be very similar. Moreover, as the questions from library readers are very similar, the libraries can be deemed as a community. Thus, the question answering sets among the libraries can be shared to answer the related questions among libraries. In this paper, to tackle the above issue of library FAQ service, by employing the question types, answer types, and community FAQ, technology a library community FAQ system is proposed. Moreover, comparing both traditional and proposed FAQ systems the users' satisfaction (five points scale) can be improved from 2.15 to 3.05.

【關鍵詞 Keywords】

常見問答集、問題型態、答案型態、圖書館社群問答系統

Frequent question answering set; Question types; Answer types; Library community FAQ system

壹、研究背景

問答 (Question answering) 技術旨在使用計算機自動回答使用者利用自然語言提問的問句，是資訊檢索和自然語言處理領域的重要研究內容 (Reddy & Madhavi, 2017; Son & Dung, 2012; 陶幼慧、黃清俊、楊誌欽, 2008; 王怡嘉, 2004)。其中，常見問題 (Frequently asked question, FAQ) 問答系統在現實生活中已被廣泛的利用。基於 FAQ 的問答系統是一種在已有的“問題-答案”對 (Pair) (以下簡稱「問答對」或稱 QA Pair) 的集合中，找到與詢問者輸入的問題匹配的問題句後，將其對應的答案回饋給詢問者之問答式檢索系統 (Jeon, Croft & Lee, 2005; Manning & Schutze, 1999; 陳榮靜、陸韋銘、鮑卓璨, 2009; 許智皓, 2008; 沈天佐、林川傑、陳信希, 2003; 郭宗瀚, 2001)。近幾年由於網際網路的快速發展，使得大多數的公司或組織走向 E 化，並建構所屬網頁，與客戶建立溝通的管道，而線上 FAQ 客服系統逐漸成為公司或組織與客戶之間的重要橋樑 (許智皓, 2008)。

圖書館的參考服務，其互動性與合作性隨著時代的變化而增加，參考館員及讀者所扮演的角色也在轉變中，從資訊的蒐集者、資訊的接收者轉變為資訊的產生者，而參考服務的重點朝向研究、資訊選擇及評估的方向愈趨明顯化 (田代如, 2012)。除了參考資料之外，館員回答過的參考問答，也是重要的參考諮詢資源，館員多半會整理讀者常見問答 (FAQ)，以減輕回答讀者重複詢問類似問題的工作負荷並提高服務效率。知識庫發展之目的就在於可重複利用這些已回答過的參考問題 (簡玉菱、柯皓仁, 2002)。

為了迅速服務讀者和節省參考館員的時間，國內外圖書館的網頁也都有設置常見問題 (FAQ)。本研究以中山大學圖書館、臺灣大學圖書館和臺北市立圖書館網站上常見問題之「圖書館目錄上顯示為在館內的書卻找不到？」或「館藏目錄查到的圖書，為何書架上找不到？」等類似問題為例及其對應的相關解答，進行觀察並綜合歸納須進一步探討的課題。

一、中山大學圖書館

問題：我在館藏目錄上找到一本書，可是在書架上卻找不到耶？

回答：

- (一) 請先確認該書在館藏目錄上的狀態是不是「可流通」。當館藏狀態「Due Date」一欄顯示字樣為：「dd-mm-yy」，如「20-05-2010」表示前一位借閱者可外借至 2010 年 5 月 20 日；顯示「展示期滿」，表示該書正在 3 樓新書展示架或主題書展示架上，待展示期滿後可外借。
- (二) 可能因館藏地不同，而在各個不同樓層或指定之區域（例如：3 樓新書展示架、5 樓韓文資料區、8 樓大本藝術圖書區等），請注意館藏目錄所顯示之館藏地。
- (三) 館藏資料會有被歸錯架或放錯樓層的情形。需要籲請各位讀者多多配合，將資料歸回正確的樓層與架位，或是放置於各樓層的置書車上待工作人員重新上架，以免造成其他讀者的不便。
- (四) 找書時，請前後左右上下大略找找看。也可能有其他人正在使用或尚未歸架，或是置於閱覽桌上、影印區附近、書車上等。在此也籲請使用研究小間者，將書攜入研究小間前請務必依規定辦理借書手續，以免造成其他讀者的困擾。
- (五) 如果真的找不到書時，可到 3 樓櫃檯填寫「圖書搜尋單」(以本校師生為限)，待收到預約通知後再來 3 樓櫃檯借閱就可以了。若一時找不到，我們會持續尋找一段時間，最後若確實無法找到，我們會在館藏目錄上註記「遺失」。遺失的書籍讀者可再行薦購。

二、臺灣大學圖書館

問題：本館館藏目錄查到之圖書，為何書架上找不到？

回答：當您在本館館藏目錄上查到館內有某書，卻在架上找不到時，可能是以下幾種狀況：

- (一) 已被借走：請先確認該書在本館館藏目錄上的狀態是否為「可流通」。當書被借出時，狀態會改成「到期 mm-dd-yy」字樣。
- (二) 館藏地不同：該資料可能存放於總圖各個不同樓層或指定之區域（例如：總圖二樓區分為「人社資料區」、「藝術資料區」、「密集書庫」等區域），亦可能存放於各分館，請注意本館館藏目錄所顯示之館藏地。
- (三) 被上錯架：我們常常發現資料被歸錯架或放錯樓層的情形。這一點需要籲請各位讀者多多配合，資料使用完畢，請置於暫存區，由圖書館工作人員後續歸位即可。找書時，亦請前後左右上下大略找找看。
- (四) 有其他人正在使用或尚未歸架：閱覽桌上、影印室、書車上是這些資料的可能去處。在此我們也特別呼籲使用研究小間者，將書攜入研究小間前請務必依規定辦理借書手續，以免造成其他讀者的困擾。
- (五) 在新書展示架上：最近到館的新書可能仍在展示中，可請館員為您確認。
- (六) 暫時遺失或遺失：找不到書時，可尋求館員的協助，若一時找不到，我們會持續尋找一段時間，確實無法找到，我們會在館藏目錄上註記「遺失」。遺失的書籍讀者可再行薦購。

- (七) 撤架或已註銷：「撤架」或「已註銷」的資料，可能因資料陳舊內容已失時效、或破損不堪修復等因素撤架準備註銷或已銷毀，因此不在架上。
- (八) 如果在總圖書館還是找不到這本書，請洽詢櫃台工作人員協助，填寫尋書單。本館會每日派人尋書，找到書時將通知您至總圖 1 樓流通櫃台（借還書櫃台）取書。若對尋書進度有疑問，請逕洽流通櫃台（Tel:3366-2353，email: tulcir@ntu.edu.tw）。

如果您急需使用，建議先到其他圖書館使用，可透過 MetaCat、NBINet 等國內圖書館聯合查詢系統查詢館藏狀況 http://www.lib.ntu.edu.tw/union_catalog。另外，有關台大教職員工學生到其他館借書的服務內容，請見館際合作服務網頁的「一般館際合作互借證」。

三、臺北市立圖書館（以下簡稱「北市圖」）

問題：為什麼有些書明明標示是在書架上，但卻找不到？

回答：

- (一) 本館日後仍會持續加強書架之圖書整理，然而整齊之書架亦需仰賴讀者的共同維持，本館圖書因採開架式，圖書因讀者隨興取閱而造成歸架錯誤較難完全避免，但本館仍將盡力維持。造成不便，敬請見諒！
- (二) 另有關您找不到館藏目錄查詢顯示在館內的書，可由館員再次為您至架上尋找，或您可以採取預約的方式以便從其他分館調閱該書給您。若您還有任何不了解的地方，歡迎您以電話洽詢（02-27553554#2104）或親至櫃檯詢問，以便本館立即為您說明。

觀察上述三間圖書館 FAQ 服務的範例，存在著如下所列現象及待探討的課題：

- (一) 由於問答题數普遍不多（一般都不會超過 20 題），且僅止於由讀者自行瀏覽，並沒有建置自動 FAQ 查詢系統，造成讀者無法由 FAQ 中得到回答或滿意的答案。各圖書館網站的常見問答集具有互補性，然而目前每個網站只能搜尋自己的常見問答集而無法利用其他網站所提供的常見問答集。因此，如何在圖書館社群中共用 FAQ 是值得探討的課題。
- (二) 由於製作常見問答的人員，依其經驗與知識提供的答案詳細程度不一，例如，北市圖和臺灣大學圖書館的回答，可以發現如果北市圖的讀者同時可以獲得臺灣大學圖書館的解答的話，可以進一步了解答案和提高其滿意度。導入社群問答 (Community question answering, CQA) 技術將可有效達成目的 (Liu et al., 2008)。
- (三) 即使是類似問題，各館提供答案的內容與方式也是五花八門，使用傳統資訊檢索的關鍵字搜尋，例如：TFIDF 方式 (Manning & Schutze, 1999)，不容易自動匹配到提問的答案。例如：「網頁上書籍狀態明明就是『在館內』，可是去書架找卻找不到」、「我在館藏目錄上找到一本書，可是在書架上卻找不到耶？」等的關鍵字差異造成無法匹配。
- (四) FAQ 問題的答案內容可分為可分享 (Reusable) 和不可分享 (Non-reusable) 兩部分 (Liu et. al., 2008)。可分享的答案為可提供給類似問題利用的答案，例如，在架上找不到圖書時的答案；不可分享的答案則是無法再利用的答案，是屬各圖書館特有的資訊，例如各圖書館的連絡電話、樓層的服務等。當要分享答案給其他圖書館時，如何區分可分享和不可分享的答案至為重要。

- (五) 如果使用傳統關鍵詞抽取技術，針對「Who was Abraham Lincoln?」和「Who killed Abraham Lincoln?」兩問句所抽取的關鍵詞都是「who」和「Abraham Lincoln」被判斷為一樣而導致抽出錯誤的答案。由於上示兩問句的語意焦點 (semantic focus) 不一樣，造成其答案分別為「A former president of USA」和「John Wilkes Booth」(Hovy, Gerber, Hermjakob, Lin, & Ravichandran, 2001)。因此，如何導入「問題詞」、「企圖詞」(Intention words) 和「相關詞」正確分類問句的「語意焦點」也是重要的課題。

因此，本論文為了解決上述五個課題，遂導入社群問答、問句和答案分類、問題詞、企圖詞、相關詞與語意焦點等技術，提出導入社群問答技術之圖書館 FAQ 問答系統。

貳、相關文獻

一、社群問答技術

隨著網際網路技術的普及和發展，網路已成為日常生活中不可或缺的部分，如網路購物、網路交友、發送及接收推文等，更多時候則是透過網路獲取資訊。隨著 web 2.0 時代的到來，網際網路服務已從單向的發佈資訊朝向雙向互動的發展，社群問答技術便是其中一種新型的應用，為網際網路的知識分享提供嶄新的平台，並為傳統的問答系統開啟另一扇窗。傳統的問答系統，藉由關鍵字匹配，傳回大量網頁供選擇，給使用者帶來極大不便。然而，社群問答技術係使用一個開放、互動的網路平台，透過使用者的參與，利用網路使用者的集體智慧，排除不可靠的答案後，提供問題的直接答案，如 Yahoo! Answer¹、Quora²、百度知道³等搜尋引擎。

¹ 詳細資料參見網站 <https://tw.answers.yahoo.com/>。

² 詳細資料參見網站 <https://www.quora.com/>。

³ 詳細資料參見網站 <https://zhidao.baidu.com/>。

現今，社群問答技術已受到相關領域的重視，例如 TREC(Text Retrieval Evaluation Conference)⁴之 Liveqa 評比 (Agichtein, Carmel, Pelleg, & Harman, 2015; Agichtein, Ben, Harman, Nyberg, & Pinter, 2017)、電子商務應用(Guo, Na, Hou & Li, 2017)以及仿真陳述型 (Factoid) 問答系統 (Pithyaachariyakul & Kulkarni, 2018) 等。社群問答系統構成如圖 1 所示 (蔡黎、周光有、趙軍，2011)。提問者首先提問問題，如果問答對 (QA Pair) 歷史紀錄 (既有的回答庫，或稱 QA Archives) 有相同問題或類似問題的紀錄，則直接將該問題的答案傳回給提問者。如果沒有，則將該問題上傳到網路上，由其他網路使用者回答 (社群回答)，再由提問者選擇最佳的答案，直到提問者關閉問題，同時將這個問答對 (QA Pair) 保存到問答對歷史紀錄 (既有回答庫) 中。

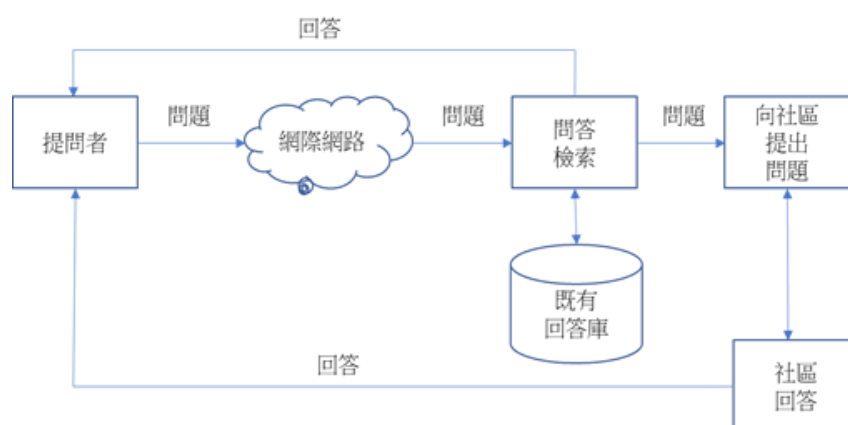


圖 1 社群問答系統構成 (蔡黎、周光有、趙軍，2011)

因此，問答是社群問答系統的核心，實現提問和回答兩方面的功能，一般主要由問題分析、問題檢索、答案處理三大模組組成，如圖 2 所示。「問題分析」模組通過對問題語法、語意等進行剖析和分類，取得提問的主題資訊。「問題檢索」模組從大量問答對 (QA Pair) 歷史紀錄中找到與提問者相關或相近的問題 (Jeon, Croft & Lee, 2005; 斐婧、包宏，2009; 李晨、巢文涵、陳小明、李舟軍，2011)。其中，問題相關性計算是「問題檢索」模組的關鍵所在。「答案處理」模組

⁴ 詳細資料參見網站 <https://trec.nist.gov/>。

則是對檢索出的答案進行排序以及質量評估，從而將最符合提問者意圖的答案傳回給提問者。

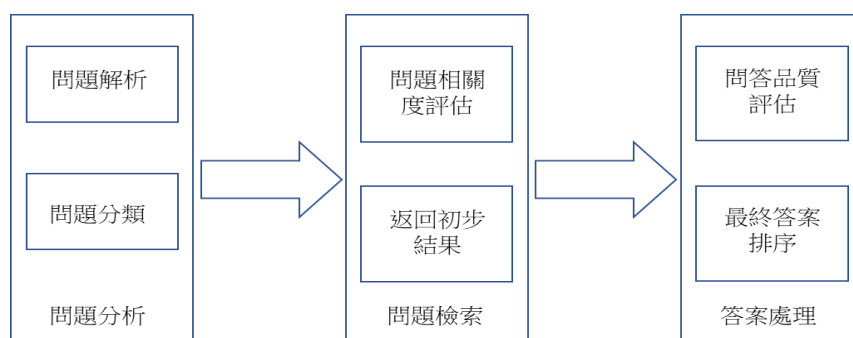


圖 2 問答系統構成圖（蔡黎、周光有、趙軍，2011）

二、問題詞、企圖詞、相關詞與語義焦點

「問題詞」(Question words) 為類似英文疑問句中的 Wh-terms，中文的告訴、有沒有、何時、多少等疑問詞外，問句中還有「想」等其他的疑問詞，例如：「我想搭太魯閣號班車」的「想」。「企圖詞」為顯示問句語意的主要辭彙，如交通類問句的航班、商務艙、訊息、來回等，例如：「我想搭中華航空來回」的「來回」。然而，由於問句語意存在歧異性，使得「企圖詞」仍不足以判斷問句之語意焦點（或稱「問句型態」）(Semantic Focus)，因此，除了從問句中選取「問題詞」、「企圖詞」之外，再從「問題詞」和「企圖詞」周圍選取「相關詞」(Related words)，用以輔助判斷該問句的語意焦點，例如：「我想搭中華航空來回」的「中華航空」和「搭」。Kuo、Lin、Chen、Kao 與 Lin (2002) 提出利用問題詞和企圖詞前後詞彙的個數選取相關詞，例如：選取一個字詞等，然而，實驗結果發現此方法無法準確輔助判斷出語意焦點，反而降低準確率。因此，為了提高擷取相關詞的正確率，本研究使用人工標記方式，選取每個問句的相關詞，輔助判斷該問句的語意焦點。

三、問句和答案分類

針對問句對應的答案進行標記與分類，依據答案內容予以標記「可再利用」與「不可再利用」。例如：北市圖「若您還有任何不了

解的地方，歡迎您以電話洽詢（02-27553554#2104）或親至櫃檯詢問，以便本館立即為您說明。」其中，所提供的電話號碼僅能於特定圖書館（北市圖）使用，因此，此內容無法作為其他圖書館的參考答案，必須標示為「不可再利用」。然而，臺灣大學圖書館「閱覽桌上、影印室、書車上是這些資料的可能去處。在此我們也特別呼籲使用研究小間者，將書攜入研究小間前請務必依規定辦理借書手續，以免造成其他讀者的困擾。」答案內容所描述的情況與圖書館館內之相關規則，為所有的圖書館通用，則可標示為「可再利用」。圖 3 為 Liu et al. (2008) 對於社群問答系統答案的分類。

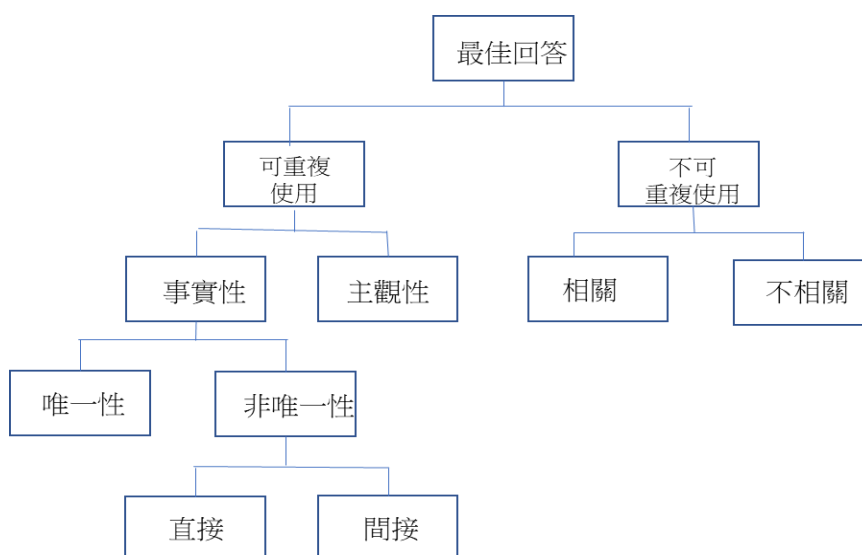


圖 3 答案分類 (Liu et al., 2008)

本論文導入社群問答技術分別解決 FAQ 題數不足 (課題一)、問題回答質量不足 (課題二)，以及嘗試使用人工標記是否可以共用解答以期克服相類似問題之解答無法完全共用 (課題四)，此外，導入疑問詞、企圖詞、與相關詞以解決同義詞造成問句無法匹配 (課題三) 和無法正確判定問句之語義焦點 (課題五)，以期提高讀者使用 FAQ 服務的滿意度。

參、系統架構與處理模組

現今圖書館的常見問答系統屬於 web 2.0，其主機由圖書館提供，資料則由網站主或使用者輸入，因而造成如上所述常見問答系統的種種研究課題。社群問答系統除本身問答集歷史紀錄的答案外，還可以獲得社群其他相關成員的協助，使得提問者可以得到期待或滿意的答案，補強目前 FAQ 問答系統的不足。由於圖書館提供的服務包含技術服務、讀者服務及相關管理服務（例如，館藏管理、空間管理等），提問者的問題自然針對上述面向提出問題。再者，圖書館館員皆有一定的專業素養（王梅玲、劉濟慈，2009），對於 FAQ 問題回答的品質有一定的水準，不會出現如傳統之社群問答系統必須審查答案品質的課題，可以得到事半功倍之效果。因此，為解決上述圖書館之常見問答系統的課題，本研究研擬導入社群問答技術。

本研究的系統架構圖如圖 4 所示，分成五大模組：1.語料蒐集與人工標記模組、2.分類器訓練模組、3.歷史資料紀錄模組、4.答案擷取與摘要模組、5.歷史資料紀錄更新模組，分述於下。

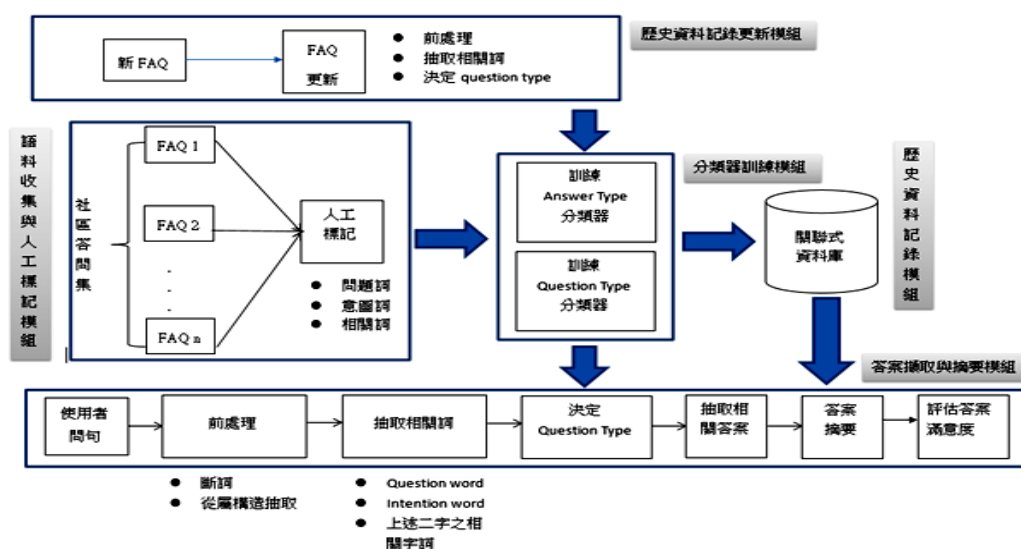


圖 4 系統架構圖

一、語料蒐集與人工標記模組

由網際網路上使用人工或自動方式(例如,網路蜘蛛)(Pranav & Chauhan, 2015)蒐集國內圖書館的常見問答集,並且針對所有問句,使用美國史丹佛大學自然語言處理實驗室⁵或中研院中文詞知識庫小組⁶所提供的免費工具(例如,中文斷詞和中文從屬構造剖析)產生斷詞結果和對應的從屬構造(dependency structure)。接著,聘請校內有修習過語言學的工讀生,參照中文斷詞結果和從屬構造對同一問答對(QA pair)分別進行問題和答案的標記。問題相關的標記有語意焦點、問題詞、企圖詞和相關詞彙等。

二、分類器訓練模組

Huang、Thint 與 Qin (2008) 使用標題詞、語意辭典擴充詞進行問句分類,實驗結果得到的分類正確率介於 89%至 89.2%之間,然而,僅考慮到標題詞及其語意擴充詞,正確率仍有待提升的空間。賴育昇、李坤霖與吳宗憲(2000)則將問句分成「意圖區段」、「關鍵詞區段」兩個部分,意圖區段傳達使用者主要的意圖,關鍵詞區段則包含問句中所有的關鍵詞,問句語意的比對將建立在「意圖區段」、「關鍵詞區段」各自的語意比對之上。賴育昇等人(2000)採用向量空間模型比較問句中的關鍵詞與 FAQ 樣本的答案,經實驗驗證,平均正確答案的排名從第 12.04 名提升到第 2.91 名,且使得前 10 名的召回率(Recall ratio)由 78.06%提升到 95.11%。Kuo 等人(2002)針對交通類問答集歸納出如表 1 所示的問句型態,於論文中亦指出同一相關詞的語意會隨著不同領域而改變,因此,若不重新訓練問題型態,將會影響分類效果。

⁵ 詳細資料參見網站 <http://nlp.stanford.edu/>。

⁶ 詳細資料參見網站 <http://ckip.iis.sinica.edu.tw:8080/>。

表 1 交通類提問之語意焦點 (Kuo et al., 2002)

語意焦點	說明
LEG	query on flight number, for example, 告訴我大約在下午五點十分以後飛離多倫多到鹽湖城的班機。
SELECTION	whether (yes/no) question, for example, 這個航班有沒有商務艙？
PRICE	query on price, for example, 商務艙票價大約貴多少？
DESCRIPTION	the answer is a descriptive sentence, for example, 告訴我舊金山機場的陸上轉運訊息。
CONSTRAINT	it is not a query, but describes the user's constraint when purchasing a ticket, for example, 兩個航班都必須在下午兩點左右到達。
TIME	query on time, for example, 謝謝！這 838 航班是幾點的？
LOCATION	query on place, for example, 列出 838 航班來回於什麼機場。
OTHER	the query cannot be categorized, for example, 哇！好多班次啊！

因此，本研究使用經過人工標記之語料，包含問題詞、企圖詞、相關詞及其對應的語意焦點，利用各種資料探勘方法訓練問句語意焦點分類器，並找出效果最佳的分類器，資料探勘方法如決策樹 (decision tree)、kNN、SVM 等分類演算法。

三、歷史資料紀錄模組

本研究經由初步的「實體－關係」分析之後，以提問句「網頁上書籍狀態明明就是「在館內，可是去書架找卻找不到？」及其相關答案內容「發生這種現象的原因有很多種，遇到此狀況，請您先確認系統顯示該書的『館藏地』是否為您所在的圖書館，並查看鄰近的書車。」為例，其關連式資料庫資料表見圖 5，由示意圖 (a)、(b)、(c)、(d) 組成。

問題號碼	問題描述	問題詞	企圖詞	語意焦點	組織代號
:	:	:	:	:	:
3	網頁上書籍狀態明明就是「在館內」，可是去書架找卻找不到？	找不到	Description	25	
:	:	:	:	:	:

圖 5-(a)問題表 (資料表 A, 織代號: 25 為銘傳大學)

問題號碼	相關詞
:	:
3	書
3	館內
3	書架
:	:

圖 5-(b) 相關詞表 (資料表 B)

問題號碼	答案	答案分類
:	:	:
3	發生這種現象的原因有很多種，遇到此狀況，請您先確認系統顯示該書的「館藏地」是否為您所在的圖書館，並查看鄰近的書車。	可再利用
3	若仍找不到，您亦可以至櫃檯填寫圖書協尋單，由我們為您找尋。或者協助您透過調閱其他分館圖書、或館際合作方式取得。發生這種現象的原因有很多種，不過若是您發現了，請務必告訴我們，我們會記錄下來並繼續追查，若長期找不到，可能會以遺失	可再利用

處理。
: : :

圖 5-(c)答案表 (資料表 C)

識別碼	問題詞
1	CONSTRAINT
2	DESCRIPTION
3	LEG
4	LOCATION
5	OTHER
6	PRICE
7	SELECTION
8	TIME

圖 5-(d)語意焦點清單 (資料表 D)

圖 5 關聯式資料庫相關表格示意圖

在應用程式方面，本研究利用 SQL 查詢語言設計擷取符合輸入問題表之問題號碼的應用程式，以及利用 ACCESS 表單功能設計人機介面，提供使用者查詢。

四、答案擷取與摘要模組

當使用抽取相關答案應用程式得到的答案數量愈多時，自行瀏覽相關答案的提問者愈容易因為抓不到重點而感到挫折；雖然文件自動摘要已是自然語言處理 (Natural Language Processing, NLP) 的重要標的之一，但其對於所擷取出來的語句之可讀性、前後文之連貫性，一直難有重大突破，提問者仍然需要透過自行瀏覽的方式，取得較為符合的答案內容。文件摘要系統的優勢在於能自動將多筆的檢索結果以簡短摘要的形式呈現，如此節省使用者逐筆瀏覽檢索結果的時間，亦減少大量不必要的資訊傳遞，提升網路傳輸的效率 (Kuo & Chen, 2008; Liu et al., 2008)。

由於本研究蒐集的圖書館常見問答集語料約六千筆問答對 (QA Pair)，加上答案內容重複率不高，故答案擷取與摘要模組保留原圖書

館的問答對，針對各個圖書館相似問題的答案內容，刪除「不可再利用」的答案，並保留「可再利用」的答案，最後以簡短摘要的形式呈現。

五、歷史資料紀錄更新模組

如果有新的問答對產生時，進入歷史資料紀錄更新模組。首先取出問句的問題詞和企圖詞，利用從屬構造剖析器取出相關詞，再經過分類器訓練模組得到問句和答案的類別，之後存入歷史資料紀錄更新模組對應的資料表中，以供後續查詢利用。本模組為未來進階功能。

肆、實驗結果與討論

一、問句和語料的蒐集與標記

本研究使用人工的方式蒐集國內 165 所大學院校圖書館於網站設置的常見問答集，包含常見問題及對應的答案，共 6,804 題。經過人工檢視，刪除有題目但沒有答案的問題後，剩下 6,680 筆問答對。針對所有問句，使用美國史丹佛大學自然語言處理實驗室所提供的免費中文斷詞工具產生斷詞結果。接著，聘請校內有修習過語言學的工讀生，對同一問答對分別進行問題和答案的標記，針對問句的標記包含問題詞、企圖詞、相關詞、語意焦點（見附錄一），而答案的標記包含可再利用（Reusable，以「Ru」表示）、不可再利用（Non-Reusable，以「NRu」表示）。「可再利用」（Ru）指答案內容為其他圖書館皆可利用或通用，而「不可再利用」（NRu）指答案內容僅適用於特定圖書館，無法於其他圖書館通用，例如：「請問 2F 的廁所在哪裡？」為「不可再利用」。圖 6 是使用 Excel 記錄的語料分析結果示意圖，紅字部分是由語料建置人員加入實際語料未出現之重要相關詞。

集，由字串屬性值的完整集合確定。執行篩選程式之後，使用多項式單純貝氏分類器類別進行分類，且建立分類模型。

(二) 分類結果 (正確率)

1. 內部測試 (Inside Test) : 97.8144% (6534/6680)
2. 四折交叉驗證 (4 fold validation) : 71.1527%

(三) 分類錯誤的原因

不一致為分類錯誤的主要原因。檢查分類錯誤的題數過程中，發現類似的問題描述、挑選作為問題詞、企圖詞或相關詞皆相同，但人工標記卻將其標記為不同的語意焦點，造成分類錯誤。

其餘原因是語意焦點模糊所造成。使語意焦點的分類模糊之原因，與進行人工語料標記階段有關，該階段主要由聘請的兩位工讀生標記，而後由兩位兼任助理做檢查，雖然在開始標記之前，已有討論八個語意焦點的分類定義（見附錄一），但給予語意焦點的標準仍因人而異，使得在訓練問題之語意焦點分類器時，出現語意焦點模糊或不一致的情形發生。參照附錄一之語意焦點，說明如下：（並非每一個語意焦點皆有模糊的情況）

1. LEG 相關

特定書名、雜誌名稱、軟體名稱、資料庫名稱、系統名稱、聯盟（如：全國學術電子資訊資源共享聯盟、CONCERT）皆屬語意焦點 LEG，『博碩士論文』、『授權書』、指定的校區名稱亦歸類到語意焦點 LEG。除了上述相關詞彙，造成不一致的詞彙尚有含『證』之相關詞彙，例如：借書證、校友證等，由於其名稱多樣化，也不夠明確，因此有可能未分類到語意焦點 LEG。

2. SELECTION 相關

當問題詞為「？」，但問題描述明確是屬於語意焦點 SELECTION，卻無法被正確分類時，解決方法是標記時在相關詞新增「可以」或「有」，以提升分類正確率。

3. PRICE 相關

「逾期」、「罰款」、「繳」、「賠」等詞彙多被分類到語意焦點 PRICE。然而，檢查分類錯誤題目的過程，觀察到當問題描述同時有兩個(含)以上語意焦點時，分類結果為語意焦點 PRICE 的機率較高，如問題號碼 3219「逾期了會不會罰錢？」雖然此句包含 SELECTION、PRICE 兩個語意焦點，正確分類結果應為語意焦點 PRICE。

4. TIME 相關

「開放」、「時間」、「延長」、「借期」等詞彙多被分類到語意焦點 TIME，由於這些與時間相關聯的詞彙出現的頻率高，而影響到其他語意焦點的判別，尤其是當問題詞為「？」時，此情況更為明顯，造成語意焦點模糊的狀況發生。

5. LOCATION 相關

依定義說明，語意焦點 LOCATION 為地點屬性，原先人工語料標記階段只有考慮到實體位置，並未考量到虛擬位置(如詢問資料庫在圖書館網頁的位置)，於檢查分類錯誤題目的過程，觀察到分類器除了將實體位置分類到語意焦點 LOCATION 之外，亦包含虛擬位置，因此，在人工標記階段，必須再將虛擬位置相關的相關詞標記為語意焦點 LOCATION。例如：問題號碼 3054「二手書交換平台：不知道這平台連結可以放在圖書館首頁嗎?...」此題應被分類到語意焦點 LOCATION，表達二手交換平台在網頁的放置位置。

三、歷史資料紀錄模組之建立

本研究為了將語料分析結果供線上問答系統之測試者使用，經過實體關係模型（ERM）分析、關連式模型（RM）轉換和正規化（Normalization）分析後，使用 ACCESS 資料庫管理系統將語料分析結果匯入資料庫。另外，為了方便給測試者使用資料庫查詢，建立了圖 7 所示的資料庫表單。圖 7 為查詢功能之 SQL 程式示意圖。

圖 7 資料庫表單示意圖

提問者(測試者)使用鍵盤輸入問句之後，經過前處理進行斷詞，比對相關字典分別取出問題詞、企圖詞和相關詞。然後，使用分類器訓練模組訓練出來的語意焦點分類器決定此提問句的語意焦點。接著，進入答案擷取與摘要模組，使用已取出的問題詞、企圖詞和相關詞於歷史資料記錄模組抽取相關答案，並產出答案摘要。如圖 8 所示。

```

SELECT DISTINCT 答案表.問題號碼, 答案描述, 答案分類, 問題表.館稱, 問題表.語意焦點
FROM 答案表, 問題表, 相關詞表
WHERE 答案表.問題號碼 = 相關詞表.問題號碼 AND 答案表.問題號碼 = 問題表.問題號碼 AND 答案表.問題
號碼 IN (
    SELECT 問題表.問題號碼
    FROM 問題表, 相關詞表, 答案表
    WHERE 答案表.答案分類 = 'Ru' AND 語意焦點 = Forms![表單 1]!SF AND 問題詞 = Forms![表單
1]!Question AND 企圖詞 = Forms![表單 1]!Intent AND 問題表.問題號碼 = 相關詞表.問題號碼 AND 問題
表.問題號碼 = 答案表.問題號碼 AND (問題表.問題號碼 IN (SELECT 問題號碼 FROM 相關詞表
WHERE 相關詞 = Forms![表單 1]!R1)
    OR
    (問題表.問題號碼 IN (SELECT 問題號碼 FROM 相關詞表 WHERE 相關詞 = Forms![表單
1]!R2))
    OR
    (問題表.問題號碼 IN (SELECT 問題號碼 FROM 相關詞表 WHERE 相關詞 = Forms![表單
1]!R3))))
ORDER BY 答案表.問題號碼;
    
```

圖 8 查詢模組之 SQL 示意圖

四、摘要產生與測試

答案擷取與摘要模組之答案摘要功能，首先使用圖書館代號取出問題匹配成功的答案後，彙總由其他圖書館提供的相關答案，執行答案摘要的產生與匯出。本研究採用經驗法則（Heuristic Rules）提出如下所述的自動摘要演算法。

(一) 摘要產生步驟

1. 步驟一：依輸入館稱代號將答案分可再利用（Ru）及不可再利用（NRu）：
 - (1)若館稱代號未輸入，查詢結果只會出現 Ru。
 - (2)若館稱代號有輸入，查詢結果可能會出現 Ru 及 NRu 兩類型，也有可能只有 Ru：
 - I. 若兩類型答案皆出現，選取 NRu 類型（捨棄 Ru 類型）；
 - II. 若仍只有 Ru 類型，則同（1）。
2. 步驟二：以企圖詞、相關詞作為判斷依據，判斷每一個答案是否和問題相關，判斷結果分直接相關、間接相關、完

全不相關三種，用於答案摘要僅使用直接相關及間接相關。

- (1) 直接相關：輸入的企圖詞、相關詞有出現在答案裡為直接相關，僅一個詞符合即成立。
 - (2) 間接相關：輸入的企圖詞、相關詞沒有出現在答案裡，參照中文索引典（例如，廣義知網知識本體架構 Hownet⁷）後，卻有回答到問題者為間接相關。例如：企圖詞或相關詞為「電子資源」，答案含「資料庫」。
 - (3) 完全不相關：無法從（1）和（2）中得到與問題之間的關聯性，則為完全不相關。
3. 步驟三：列出步驟二得出的最終結果（不改寫），即答案摘要，如圖 9 所示。

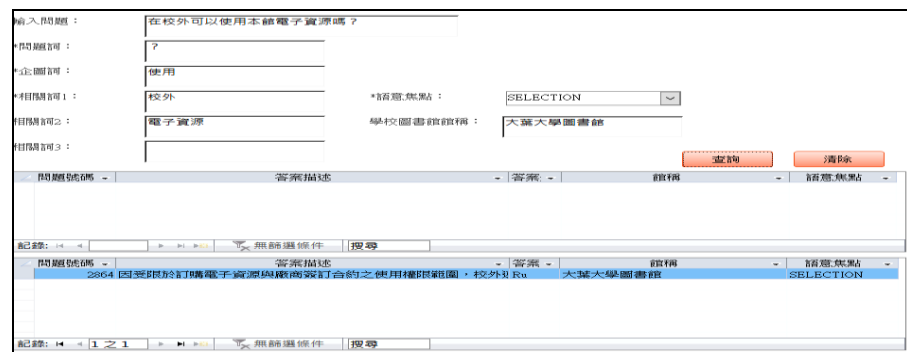


圖 9-(a)指定學校之查詢結果示意圖

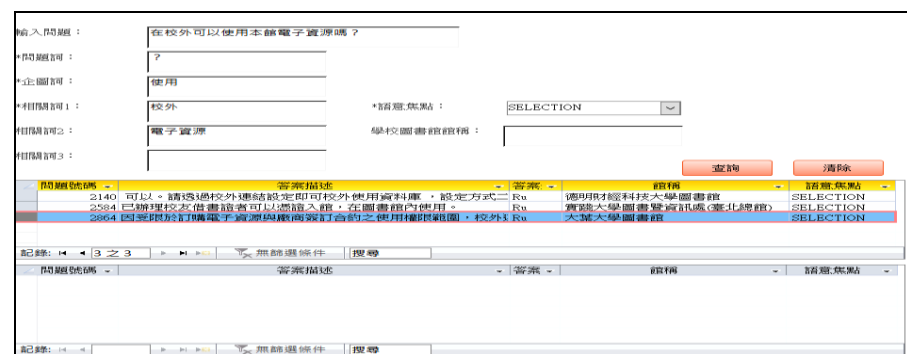


圖 9-(b) 未指定學校之查詢結果示意圖

⁷ 詳細資料參見網站 <http://ehownet.iis.sinica.edu.tw/index.php>。

圖 9 查詢結果示意圖

(二)摘要產生範例

1. 以圖 10 第 3 題為例：在校外可以使用本館電子資源嗎？

學校圖書館館稱輸入「國立中興大學圖書館」，查詢結果為 4 筆 Ru 和 1 筆 NRu。

(1) 選取 NRu 一筆。

(2) 企圖詞「使用」、相關詞「校外」、「電子資源」皆有出現在步驟一選取的答案中，為直接相關。

(3) 得答案摘要：

圖書館為提升服務品質，讓讀者在校外也能使用本館各項線上電子資源，已完成本館「代理伺服器」與「圖書館自動化系統」讀者認證整合，本校讀者可在校外使用電子資源（除四庫全書內聯版 3.0 版外皆不需設 PROXY）。校外連線使用特定資料庫，「古今圖書集成」「四庫全書」1.2 版及 3.0 版校外使用本資料庫若無法連線請參考本校[中興大學校園虛擬私有網路 (VPN)] 之設定 <http://cc.nchu.edu.tw/network-c/sslvpn.html>；「四庫全書」1.2 版進入點選以固定 IP 進入，「四庫全書」3.0 版以 IP:140.120.80.55 登入。

2. 以圖 10 第 12 題為例：我想贈送書刊資料給圖書館，要如何處理？

學校圖書館稱有輸入（「國立中興大學圖書館」），查詢結果為 4 筆 Ru。

(1) 選取 Ru 四筆。

(2) 企圖詞「贈送」、相關詞「書刊資料」及「處理」在被選取的四筆中僅出現於一筆，其他三筆為完全不相關。

(3) 得答案摘要：

影(列)印終於遇效果不良、缺紙、夾紙或機械故障情形，請通知服務櫃檯人員處理。

五、滿意度測試

為了評估本研究之社群問答技術之圖書館 FAQ 問答系統的效果，準備了傳統(直接問題比對)和本研究提出之 FAQ 問答系統兩個系統。隨機挑選不同校的五位研究生(人文領域：A1、A2 和醫藥領域：C1、C2、C3)為測試者，要求提出欲詢問圖書館的六個問句及問句中的問題詞、企圖詞與相關詞，共計 30 個問題(見圖 10)。接著，使用本研究訓練出來的語意焦點分類器得到每個問句的語意焦點。經過 10 分鐘的任務說明，請五位測試者以所提問的 30 題問題，分別測試兩個系統(平均 120 分/人)後，填寫滿意度調查問卷(見附錄二)和進行訪談。

問題號碼	問題描述	問題詞	企圖詞	語意焦點	相關詞(至多三個，請以,隔開)
1	在校外可以使用資料庫嗎?	?	使用	SELECTION	資料庫,校外
2	請問校外人士可以進入本館使用圖書資源嗎?	?	使用	SELECTION	校外人士,進入,圖書資源
3	在校外可以使用本館電子資源嗎?	?	使用	SELECTION	校外,電子資源
4	如何推薦圖書?	如何	推薦	DESCRIPTION	圖書
5	請問本館開館時間為何?	為何	開館	TIME	時間
6	請問如何借用研究小間?	如何	借用	DESCRIPTION	研究小間
7	為什麼經過圖書館的門禁系統時常會嗶嗶叫呢?	為什麼	門禁系統	DESCRIPTION	圖書館,嗶嗶叫
8	使用電子期刊需要注意什麼?	什麼	注意	DESCRIPTION	電子期刊
9	請問要如何退還保險金呢?	如何	退還	DESCRIPTION	保險金
10	圖書館各樓層提供哪些設施及服務?	哪些	提供	DESCRIPTION	設施,服務
11	請問要如何預約DVD??	如何	預約	DESCRIPTION	DVD
12	我想贈送書刊資料給圖書館,要如何處理?	如何	處理	DESCRIPTION	書刊資料,圖書館,贈送
13	要借的書不在架上,怎麼辦?	怎麼辦	借	DESCRIPTION	書,不在架上
14	圖書館冷氣好像又壞了,為什麼常常都會停擺沒有運轉呢?	為什麼	停擺	DESCRIPTION	冷氣,運轉,壞
15	NDDS上顯示「待收件」,這表示可以去圖書館收件了嗎?	?	收件	LEG	待收件,NDDS
16	請問博碩士論文如何取得?	如何	取得	LEG	博碩士論文
17	收到圖書館寄來的電子郵件呈現亂碼怎麼辦?	怎麼辦	電子郵件	DESCRIPTION	亂碼,圖書館
18	如何借用討論室?	如何	借用	DESCRIPTION	討論室
19	請問多媒體中心的開放時間?	?	開放時間	TIME	多媒體中心
20	期刊為什麼不能外借?	為什麼	外借	DESCRIPTION	期刊
21	圖書館可否代購圖書?	可否	代購	SELECTION	圖書
22	期中考有延長開館嗎?	?	開館	TIME	延長,期中考
23	期末考有延長開館嗎?	?	開館	TIME	延長,期末考
24	找不到書時怎麼辦?	怎麼辦	找不到	DESCRIPTION	書
25	如何申請其他學校的博碩士論文?	如何	申請	LEG	博碩士論文,其他學校
26	校外人士可以使用與閱坊嗎?	?	使用	SELECTION	與閱坊,校外人士
27	我查到一本狀態顯示「編目中」的書,可以借嗎?	?	借	SELECTION	編目中,書
28	校外人士可以使用圖書館的電腦嗎?	?	使用	SELECTION	圖書館,電腦,校外人士
29	校外人士可以使用圖書館的網路嗎?	?	使用	SELECTION	圖書館,網路,校外人士
30	逾期罰款?	?	罰款	PRICE	逾期

圖 10 滿意度測試的問題集

如表 2 所示，在使用傳統 FAQ 系統上，五位測試者使用傳統系統的平均滿意度分數只有 2.15，顯示對於傳統 FAQ 系統普遍不滿意。

主要的原因是很多問題都沒有或太少答案。然而，使用本研究的社群問答系統，針對 (1) 系統品質、(2) 資訊內容品質、(3) 整體滿意度的平均分數分別為 2.80, 3.15 和 3.10, 總平均為 3.02, 比傳統 FAQ 系統有明顯的進步 (3.02>2.15), 顯示本研究的可行性與進步性。然而，於系統品質方面只有 2.80, 顯示系統整合與人機介面設計，以及結果的呈現仍需整合與改善。

表 2 滿意度測試結果彙總

	題號	A1	A2	C1	C2	C3	項平均	總平均	
傳統	1	2	1	1	2	1	1.4	2.15	
FAQ	2	3	3	2	2	3	2.6		
系統	3	3	2	2	2	3	2.6		
	4	2	2	2	2	2	2.0		
社群問答系統	5	3	1	3	4	3	2.8	2.80	
	6	2	3	3	3	3	2.8		
	7	2	3	3	2	3	2.6	3.02	
	8	3	3	3	3	3	3.0		
	9	3	3	4	4	2	3.2		3.15
	10	3	4	3	3	3	3.2		
	11	3	4	4	3	2	3.2	3.10	
	12	3	3	4	2	3	3.0		
	13	2	3	4	4	3	3.2	3.10	
	14	3	3	3	3	3	3.0		

伍、結論與未來研究方向

一、結論

為提高讀者的滿意度，圖書館提供常見問答以解決讀者的疑問。然而，基於圖書館人力不足，每間圖書館的常見問答的數量通常只有十來個問題，相關解答的多寡也不一。社群問答技術在自然語言處理領域被高度關注，藉由此技術，讀者的問題不必由網站自行回答，而可由社群內的網民代為回答，經過答案品質的篩選後，將有用的答案傳給讀者，提高了提問者的滿意度。

本研究藉由分享各館的常見問答來克服上述課題，只使用輕型資料庫管理系統 (ACCESS) 以及導入社群問答技術、問句和語意焦點

的自動分類和結果呈現等技術提出基於社群問答技術之圖書館 FAQ 問答系統。使用滿意度調查問卷，針對使用傳統 FAQ 和社群問答系統（五分量表），平均滿意度由 2.15 提升到 3.02。

圖書館隨著數位時代的來臨館藏量急速攀升，讀者容易因為找不到書而感到氣餒。另一方面，隨著請求參考服務之讀者數增加，造成參考館員疲於奔命，無法提供優質的參考服務。導入基於社群問答技術的圖書館 FAQ 問答系統的應用實務需求也應運而生。因此，參考館員可以將大多數的參考服務委與本研究問答系統，而將精緻化的優質參考服務提供給讀者。

二、未來研究方向

分析評估結果後，本研究仍有如下的研究方向可繼續努力。

- (一)本研究的摘要模組只使用經驗式法則，為了提高摘要的品質可導入自動摘要技術的問題回答，刪除重複的資料和濃縮回答的長度，以增加讀者的可讀性和滿意度。
- (二)本研究採用語意焦點之分類，使得借期、開放時間等目前被歸在 TIME，會導致語意模糊；另外如權限、借閱冊數、空間使用權等應如何歸類，需要再進一步討論其適當性。
- (三)擴大問卷受測人數，進一步了解與分析讀者滿意度。
- (四)本研究之系統效率相依於 FAQ 問答對之數量，另一個重要課題是如何使用自然語言處理，例如機器翻譯，解決語言差異，導入國外圖書館之 FAQ 問答對。

【誌謝】

科技部計畫 MOST 106-2410-H-005-021 對於本論文的部分經費補助和支持以及研究助理陳雨沛同學和屈亭君同學熱心參與實驗和討論，在此致上由衷感謝之意。

【參考書目】

- 王怡嘉 (2004)。問答系統之查詢擴充-以網路資源為本之非監督式學習策略。未出版之碩士論文，國立交通大學資訊科學學系，新竹。
- 王梅玲、劉濟慈 (2009)。從圖書館價值探討我國圖書館員基本專業能力。《圖書資訊學研究》，4(1)，27-68。
- 田代如 (2012)。公共圖書館參考服務之演變。《臺北市立圖書館館訊》29(4)，50-76。
- 李晨、巢文涵、陳小明、李舟軍 (2011)。中文社區問答中問題答案質量評價和預測。《計算機科學》，38(6)，230-234。
- 沈天佐、林川傑、陳信希 (2003)。以網際網路內容為基礎之問答系統“Why”問句研究。第十五屆自然語言與語音處理研討會，新竹。
- 許智皓 (2008)。運用具學習能力的領域知識本體於線上 FAQ 系統。未出版之碩士論文，朝陽科技大學資訊管理系，臺中市。
- 郭宗瀚 (2001)。FAQ 式中文自然語言資訊檢索建構在 SQL 語法上。未出版之碩士論文，國立暨南國際大學資訊管理學系，南投。
- 陶幼慧、黃清俊、楊誌欽 (2006)。網路論壇 FAQ 知識之自動轉換設計。《資訊管理學報》，13(2)，89-112。
- 陳榮靜、陸韋銘、鮑卓璨 (2009)。運用本體知識於常用問題集之自動回應—以糖尿病為例。2009 年資訊科技國際研討會，臺中市。
- 斐婧、包宏 (2009)。漢語句子相似度計算在 FAQ 中的應用。《計算機工程》，35(17)，46-48。

- 蔡黎、周光有、趙軍 (2011)。社區問答技術。中國人工智能學會通訊, 1(1), 13-18。
- 賴育昇、李坤霖、吳宗憲 (2000)。網際網路 FAQ 中意圖萃取與語意比對之研究, 第十三屆自然語言與語音處理研討會 (ROCLING XIII), 135 ~ 155, 臺北市。
- 簡玉菱、柯皓仁 (2002)。「網際網路時代的參考諮詢服務: 線上合作參考諮詢服務」。國家圖書館館訊, 3 (民 91 年 8 月), 7-13。
- Agichtein, E., Ben, Abacha, A., Harman, D., Nyberg, E., & Pinter, Y. (2017). Overview of the TREC 2017 Liveqa track. *Proceedings of The 26th Text Retrieval Conference (TREC 2017)*, Gaithersburg, Maryland, USA.
- Agichtein, E., Carmel, D., Pelleg, D., & Harman, D. (2015). Overview of the TREC 2015 Liveqa track. *Proceedings of The 24th Text Retrieval Conference (TREC 2015)*, Gaithersburg, Maryland, USA.
- Huang, Z.H., Thint, M., & Qin, Z.C. (2008). Question classification using headwords and their hypernyms. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 927-936.
- Hovy, E., Gerber, L., Hermjakob, U., Lin, C.Y., & Ravichandran (2001). Towards semantic-based answer pinpointing. *Proceedings of the 2001 DARPA Human Language Technology Conference*, San Diego.
- Jeon, J., Croft, W. B., & Lee, J.H. (2005). Finding similar questions in large question and answer archives. *CIKM '05 Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 84-90.
- Guo, H., Na, X., Hou, L., & Li, J. (2017). Classifying Chinese questions related to health care posted by consumers via the Internet. *JMIR Medical Informatics*, 19(6), e220.

- Kuo, J.J., & Chen, H.H. (2008). Multidocument summary generation: Using informative and event words. *ACM Transactions of Asian Language Information Processing (TALIP)*, 7(1).
- Kuo, J.J., Lin, K.K., Chen, H.H., Kao, C.H., & Lin, B.S. (2002). Question type classification and its application to a question answering system. *Proceedings of 2002 IEEE International Conference on Systems, Man and Cybernetics*, Hammarnet, Tunisia.
- Liu, Y.J., Li, S.S., Cao, Y.B., Lin, C.Y., Han, D.Y., & Yu, Y. (2008). Understanding and Summarization Answer in Community-based Question Answering Services. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 497-504.
- Manning, C., & Schutze, H. (1999). *Foundation of Statistical Natural Language Processing*. The MIT Press.
- Pithyaachariyakul, C., & Kulkarni, A. (2018). Automated question answering system for community-based questions. *Proceedings of the 32th AAI Conference on Artificial Intelligence (AAAI-18)*, 8131-8132.
- Pranav, A., & Chauhan, S. (2015). Efficient focused web crawling approach for search engine. *International Journal of Computer Science and Mobile Computing*, 4(5), 45-551.
- Reddy, A., & Madhavi, K. (2017). A survey on types of question answering system. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 19(6), 19-23.
- Son, D.T., & Dung, D.T. (2012). Apply a mapping question approach in building the question answering system for Vietnamese Language. *Journal of Engineering Technology and Education: The 2012 International Conference on Green Technology and Sustainable Development (GTSD2012)*, 380-386.

附錄一：語意焦點說明

語意焦點	語意焦點與其他標記語料為相輔相成的效果。如：該問句若是沒有找到『疑問詞』，則利用語意焦點來判斷問答的類型/內容。
LEG	含有明確的人事物。例如：特定的某本書。
SELECTION	具選擇性，例如：可否、是否、有沒有。
PRICE	與價格有關，例如：(逾期) 罰款。
DESCRIPTION	該問句屬於敘述性質。
CONSTRAINT	指帶有條件限制。
TIME	時間屬性。
LOCATION	地點屬性，例如詢問報章雜誌的位置。
OTHER	該問句無法歸類上述七項（盡量少用）。

附錄二：滿意度調查問卷

壹、傳統 FAQ 系統

此部分問題欲瞭解參與者對於目前各圖書館網頁附上的「常見問題」之使用經驗。請您就下列的描述，勾選符合的選項。

題目	非常滿意	滿意	普通	不滿意	非常不滿意
1. 使用任一圖書館網頁提供的「常見問題」，所附上的答案或回覆之完整性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. 使用任一圖書館網頁提供的「常見問題」，所附上的答案或回覆之可靠性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. 使用任一圖書館網頁提供的「常見問題」，所附上的答案或回覆之正確性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. 使用任一圖書館網頁提供的「常見問題」，所附上的答案或回覆之及時性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

貳、社群答問系統

此部分問題欲瞭解參與者使用本計畫設計的問答系統之後，對系統的滿意度。請您就下列的描述，勾選符合的選項。

一、系統品質

5. 圖書館 FAQ 問答系統操作與使用的容易性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. 圖書館 FAQ 問答系統運作的可靠性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. 圖書館 FAQ 問答系統運作的穩定性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. 圖書館 FAQ 問答系統的反應時間。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

二、資訊內容品質

9. 圖書館 FAQ 問答系統輸出資訊內容之完整性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. 圖書館 FAQ 問答系統輸出資訊內容之可靠性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. 圖書館 FAQ 問答系統輸出資訊內容之正確性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. 圖書館 FAQ 問答系統輸出資訊內容之及時性。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

三、整體滿意度

13. 您對本計畫圖書館 FAQ 問答系統之滿意程度。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. 本計畫圖書館 FAQ 問答系統有發揮其應有之效益與功能。	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

參、參與者回饋 (自由作答)

此部分提供給參與者填寫針對本計畫圖書館 FAQ 問答系統的意見回饋。

本問卷至此已全部結束，感謝您的填寫。

最後提醒您再次檢查是否有漏答之處，感謝您的協助。

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology

June-Jei Kuo

Corresponding author: jjkuo@dragon.nchu.edu.tw

Associate Professor, Graduate Institute of Library and Information Science, National Chung-Hsing University

Abstract

To increase the readers' satisfaction each library usually provides frequent question answering set to answer the related questions. Nevertheless, due to the lack of manpower there are few question answering pairs in the library website. Thus, the question providers can not usually obtain satisfactory answers. Even worst, there is not such question in the FAQ, they cannot obtain any answer. Those will frustrate the readers not to utilize the library website or libraries. As the community FAQ technology is emphasized in natural language processing era, whenever a question is provided, the answers can be given by not only the website managers but also the members in the same community. Through the quality evaluation for the answers, only the best answers will be provided to the question providers which can definitely enhance their satisfaction. As the purpose of libraries is to provide services for the readers, those services will be very similar. Moreover, as the questions from library readers are very similar, the libraries can be deemed as a community. Thus, the question answering sets among the libraries can be shared to answer the related questions among libraries. In this paper, to tackle the above

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

issue of library FAQ service, by employing the question types, answer types, and community FAQ technology, a library community FAQ system is proposed. Moreover, comparing both traditional and proposed FAQ systems the users' satisfaction (five points scale) can be improved from 2.15 to 3.05.

Keywords: Frequent question answering set; Question types; Answer types; Library community FAQ system

SUMMARY

Introduction

The modes of interaction and collaboration in library reference services are changing with time, as are the roles of reference librarians and readers. In addition to reference information, such librarians keep track of questions they have answered and compile a list of frequently asked questions (FAQs), which serves as a crucial reference source that reduces the frequency of repeatedly answering the same questions and enhances the service efficiency. To provide rapid reference services to readers and save the time of reference librarians, the websites of libraries in Taiwan and worldwide feature FAQ sections. However, the following phenomena and topics merit further exploration regarding library FAQ:

(a) A FAQ list usually contains few questions (≤ 20 questions in general) and is browsed by readers on their own. The lack of an automatic FAQ inquiry system often prevents readers from finding satisfactory answers to their questions.

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

(b) Observing the frequent question answering sets of various libraries indicates that the compiled questions are complementary. Therefore, they should be better utilized to further improve reader satisfaction.

(c) The answers provided by different libraries to the same question vary greatly, and the manner in which the question is answered also differs. The use of keywords in conventional inquiry systems is ineffective in automatically finding the corresponding answer.

(d) Answers to FAQ are divided into reusable and non-reusable. When sharing FAQ answers with other libraries, how to distinguish reusable answers from non-reusable answers is essential.

(e) How to introduce question words, intention words, and related words to accurately classify the semantic focus of a question is another crucial topic.

To address these five topics, this study proposed a community question answering (CQA) technique applicable to library FAQ systems.

Methods

CQA techniques have received notable attention in the field of natural language processing. Using such techniques, website managers no longer need to answer website visitors' questions personally and can instead rely on other online community members to provide answers. After such answers are screened and selected, those that are considered informative are sent to askers and thereby enhance the user satisfaction. The goals of libraries are to provide technical and reader services. To address readers' questions, libraries provide FAQ lists and compile answers to such

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

questions from Internet users to form an online community. A system for managing such communities encompasses the following modules:

(a) Language data collection and manual labeling module: The FAQ lists provided by the libraries of 165 higher education institutions in Taiwan were collected manually to compile 6,680 question–answer pairs. Next, the question and answer of each pair were labeled. Specifically, labels were placed on the question words, intention words, related words, and semantic focus of the question. The answer was labeled as either “Ru” or “NRu”, which denoted a reusable or non-reusable answer, respectively.

(b) Classifier training module: The manually labeled language data were used to train question semantic focus classifiers using the decision tree, k-nearest neighbor, and support vector machine (SVM) algorithms, thereby identifying the optimal classifier. The results demonstrated that the SVM classifier was most accurate. Accordingly, the classifier was used in the subsequent classification of semantic focus.

(c) Historical data record module: The language data analysis results needed to be tested by online FAQ system users. To achieve this, the language data were subject to entity–relationship model, relationship model, and normalization analyses, and then the ACCESS data management system was used to input the analysis results into a database.

(d) Answer retrieval and summary module: Library code numbers were used to retrieve matched answers, and then the relevant answers retrieved from all libraries were compiled to generate and output an answer summary.

Results

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

Five participants were briefed on the experiment objectives for 10 minutes before they used 30 questions to test two systems; the participants spent 120 minutes, on average, to complete the test. Subsequently, they completed a satisfaction survey (5-point scale) and participated in interviews. The conventional FAQ system received a mean satisfaction score of 2.15, indicating that the participants were dissatisfied with the system. By contrast, the proposed CQA system received a mean score of 2.80, 3.15, and 3.10 in the aspects of system quality, information content quality, and overall satisfaction, respectively, yielding an overall mean score of 3.02. Said score was notably higher than that of the conventional FAQ system (2.15).

Conclusions and Suggestions

In this study, FAQ lists of various libraries were compiled and shared to address difficulties encountered in library reference systems. ACCESS, a light database management system, was employed in combination with the CQA technique, automatic classification of questions and semantic focus, and answer presentation technique to create a CQA-based library FAQ system. A satisfaction survey was used to compare the conventional FAQ system and proposed CQA system and revealed that the proposed system improved the mean satisfaction score from 2.15 to 3.02.

In the current era of digitalization, the volume of library archives is expanding rapidly, which frustrates the readers when they cannot find specific books in a massive archive. The increasing number of readers requiring reference services easily exhausts reference librarians and prevents them from providing adequate services. To address this problem,

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

CQA-based library FAQ systems have been created. The proposed system can be used by reference librarians to refine relevant services and satisfy the need of readers.

ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT

王怡嘉 (2004)。問答系統之查詢擴充-以網路資源為本之非監督式學習策略。未出版之碩士論文，國立交通大學資訊科學學系，新竹。

【Wang, Yi-Chia (2004). *Web-Based Unsupervised Learning to Query Formulation for Question Answering*. (Unpublished master's thesis). College of Computer Science NCTU, Hsinchu (in Chinese)】

王梅玲、劉濟慈 (2009)。從圖書館價值探討我國圖書館員基本專業能力。《圖書資訊學研究》，4(1)，27-68。【Wang, Mei-Ling, & Liu, Chi-Tzu (2009). A Study of the Core Competencies for Taiwan's Librarians from the Library Values Perspective. *Journal of Library and Information Science Research*, 4(1), 27-68 (in Chinese)】

田代如 (2012)。公共圖書館參考服務之演變，臺北市立圖書館館訊 29(4)，50-76。【Tien, Tai-Ju (2012). The Transforming of Reference Services in Public Libraries. *Bulletin of the Taipei Public Library*, 29(4), 50-76 (in Chinese)】

李俞鋒 (2019)。基於 Google Dialogflow 技術框架之 FAQ 對話機器人系統設計。未出版之碩士論文，國立臺北科技大學資訊工程系，新竹。【Lee, Yu-Fong (2019). *Design of a FAQ Chatbot System Based on Google Dialogflow*. (Unpublished master's thesis). Department of Computer Science and Information Engineering, Taipei Tech, Hsinchu (in Chinese)】

李晨、巢文涵、陳小明、李舟軍 (2011)。中文社區問答中問題答案質量評價和預測。《計算機科學》，38(6)，230-234。【Li, Chen, Chao, Wen-

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

- Han, Chen, Xiao-Ming, & Li, Zhou-Jun (2011). Quality Evaluation and Prediction for Question and Answer in Chinese Community Question Answering. *Computer Science*, 38(6), 230-234 (in Chinese)】
- 沈天佐、林川傑、陳信希 (2003)。以網際網路內容為基礎之問答系統“Why”問句研究。第十五屆自然語言與語音處理研討會，新竹。
- 【Shen, Tian-Zuo, Lin, Chuan-Jie, & Chen, Hsin-Hsi (2003). A Study of Why Questions in Web-Based Question Answering System. *ROCLING XV*, Hsinchu (in Chinese)】
- 許智皓 (2008)。運用具學習能力的領域知識本體於線上 FAQ 系統。未出版之碩士論文，朝陽科技大學資訊管理系，臺中市。【Hsu, Chih-hao (2008). *Leading the Domain Ontology with Learning Ability in Frequently Asked Questions System*. (Unpublished master's thesis). Department of Information Management CYUT, Taichung (in Chinese)】
- 郭宗瀚 (2001)。FAQ 式中文自然語言資訊檢索建構在 SQL 語法上。未出版之碩士論文，國立暨南國際大學資訊管理學系，南投。【Kuo, Chung-Han (2001). *FAQ Like Chinese Natural Language Information Retrieval Based on SQL*. (Unpublished master's thesis). NCNU Information Management, Nantou (in Chinese)】
- 陶幼慧、黃清俊、楊誌欽 (2006)。網路論壇 FAQ 知識之自動轉換設計。資訊管理學報，13(2)，89-112。【Tao, Yu-Hui, Huang, Ching-Chun, & Yang, Chih-Chin (2006). Designing an Automatic FAQ Abstraction for Internet Forum. *Journal of Information Management*, 13(2), 89-112 (in Chinese)】
- 陳宜琳 (2019)。國立臺灣師範大學圖書館參考諮詢機器人建置與評估。未出版之碩士論文，國立臺灣師範大學圖書資訊學研究所，臺北市。【Chen, Yi-Lin (2019). *Development and Evaluation of a Chatbot for Reference Service in National Taiwan Normal University*

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

Library. (Unpublished master's thesis). Graduate Institute of Library & Information Studies NTNU, Taipei (in Chinese)】

陳榮靜、陸韋銘、鮑卓璿 (2009)。運用本體知識於常用問題集之自動回應—以糖尿病為例。2009 年資訊科技國際研討會，臺中市。

【Chen, Rung-Ching, Lu, Wei-Ming, & Bao, Jhuo-Tsan (2009). Yunyong benti jihshih yu changyong wuntiji jih zihdong hueiying - yi diabetes weili. 2009 International Conference on Advanced Information Technologies, Taichung (in Chinese)】

裴婧、包宏 (2009)。漢語句子相似度計算在 FAQ 中的應用。《計算機工程》，35(17)，46-48。【Pei, Jing & Bao, Hong (2009). Application of Chinese Sentence Similarity Computation in FAQ. *Computer Engineering*, 35(17), 46-48 (in Chinese)】

歐陽彥隆 (2001)。電信服務業務 FAQ 查詢系統之研製。未出版之碩士論文。國立交通大學資訊科學學系，新竹。【OuYang, Yen-Long (2001). *Study and Implementation of A Dialogued-Based Query System for Telecommunication FAQ Services*. (Unpublished master's thesis). College of Computer Science NCTU, Hsinchu (in Chinese)】

蔡黎、周光有、趙軍 (2011)。社區問答技術。《中國人工智能學會通訊》，1(1)，13-18。【Cai, Li, Zhou, Guang-you, & Zhao, Jun (2011). Shechyu wunda jishu. *Chinese Association for Artificial Intelligence Newsletter*, 1(1), 13-18 (in Chinese)】

賴育昇、李坤霖、吳宗憲 (2000)。網際網路 FAQ 中意圖萃取與語意比對之研究，第十三屆自然語言與語音處理研討會 (ROCLING XIII)，135 ~ 155，臺北市。【Lai, Yu-Sheng, Lee, Kuen-Lin, & Wu, Chung-Hsien (2000). Intention Extraction and Semantic Matching for Internet FAQ Retrieval Using Spoken Language Query. *ROCLING XIII*, 135 ~ 155, Taipei (in Chinese)】

簡玉菱、柯皓仁 (2002)。「網際網路時代的參考諮詢服務：線上合作參考諮詢服務」。《國家圖書館館訊》，3 (民 91 年 8 月)，7-13。

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

- 【Chien, Yuh-Ling, & Ke, Hao-Ren (2002). Wangjiwanglushihdai de tsankao zihsyunfuwu : sianshang hezuo tsankao zihsyunfuw. *National Central Library News Bulletin*, 3 (2002, August), 7-13 (in Chinese)】
- Agichtein, E., Ben, Abacha, A., Harman, D., Nyberg, E., & Pinter, Y. (2017). Overview of the TREC 2017 Liveqa track. Proceedings of The 26th Text Retrieval Conference (TREC 2017), Gaithersburg, Maryland, USA.
- Agichtein, E., Carmel, D., Pelleg, D., & Harman, D. (2015). Overview of the TREC 2015 Liveqa track. Proceedings of The 24th Text Retrieval Conference (TREC 2015), Gaithersburg, Maryland, USA.
- Huang, Z.H., Thint, M., & Qin, Z.C. (2008). Question classification using headwords and their hypernyms. Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, 927-936.
- Hovy, E., Gerber, L., Hermjakob, U., Lin, C.Y., & Ravichandran (2001). Towards semantic-based answer pinpointing. Proceedings of the 2001 DARPA Human Language Technology Conference, San Diego.
- Jeon, J., Croft, W. B., & Lee, J.H. (2005). Finding similar questions in large question and answer archives. CIKM '05 Proceedings of the 14th ACM international conference on Information and knowledge management, 84-90.
- Guo, H., Na, X., Hou, L., & Li, J. (2017). Classifying Chinese questions related to health care posted by consumers via the Internet. *JMIR Medical Informatics*, 19(6), e220.
- Kuo, J.J. and Chen, H.H. (2008). Multidocument summarization: Using Informative and Event Words. *ACM Transactions of Asian Language Information Processing (TALIP)*, Vol.7, No.1.
- Kuo, J.J., Lin, K.K., Chen, H.H., Kao, C.H., & Lin, B.S. (2002). Question

A Preliminary Study of Library Frequently Asked Question and Answering System Based on Community QA Technology / Kuo

- type classification and its application to a question answering system. Proceedings of 2002 IEEE International Conference on Systems, Man and Cybernetics, Hammarnet, Tunisia.
- Liu, Y.J., Li, S.S., Cao, Y.B., Lin, C.Y., Han, D.Y., & Yu, Y. (2008). Understanding and Summarization Answer in Community-based Question Answering Services. Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), 497-504.
- Manning, C., & Schutze, H. (1999). Foundation of Statistical Natural Language Processing. The MIT Press.
- Pithyaachariyakul, C., & Kulkarni, A. (2018). Automated question answering system for community-based questions. Proceedings of the 32th AAI Conference on Artificial Intelligence (AAAI-18), 8131-8132.
- Pranav, A., & Chauhan, S. (2015). Efficient focused web crawling approach for search engine. International Journal of Computer Science and Mobile Computing, 4(5), 45-551.
- Reddy, A., & Madhavi, K. (2017). A survey on types of question answering system. IOSR Journal of Computer Engineering (IOSR-JCE), 19(6), 19-23.
- Son, D.T., & Dung, D.T. (2012). Apply a mapping question approach in building the question answering system for Vietnamese Language. Journal of Engineering Technology and Education: The 2012 International Conference on Green Technology and Sustainable Development (GTSD2012), 380-386.