

# 科技部補助專題研究計畫成果報告 期末報告

## 提升成員知識分享意願之虛擬社群知識審核模式(I)

計畫類別：個別型計畫  
計畫編號：MOST 104-2221-E-343-001-  
執行期間：104年08月01日至105年07月31日  
執行單位：南華大學資訊管理學系

計畫主持人：楊士霆

計畫參與人員：大專生-兼任助理人員：段正有  
大專生-兼任助理人員：蘇品仔  
大專生-兼任助理人員：卜書婷  
大專生-兼任助理人員：張婉瑄  
大專生-兼任助理人員：廖偉哲

報告附件：出席國際學術會議心得報告

中華民國 105 年 10 月 26 日

中文摘要：既有虛擬社群運作機制因資訊爆炸且群眾智慧品質控管不易落實，致使相似內容同時存在過多差異大之評論，於此，知識需求者為取得正確知識，將須投入相當之時間本於評估群眾智慧之品質，以及社群成員可參考價值，再者，由於知識正確性評估所面臨之挑戰，將致使部分知識貢獻者於未知狀況下傳遞錯誤之資訊，而當錯誤資訊持續累積至一定程度，且逐步影響整體群眾智慧之品質時，將降低知識貢獻者分享知識之意願。

有鑑於上述問題，本研究乃發展「提升成員知識分享意願之虛擬社群知識審核模式」，並劃分為「虛擬社群問答契合度解析」，以及「成員參考度解析」兩核心模組，以群眾智慧之知識凝聚及知識分享為切入點，發展並設計一套適用於虛擬社群之方法論。於前者中，本研究結合領域性關鍵詞彙集合概念、文章內容核心資訊萃取、文章相似度判定，以及詞彙語意分析等技術，以判定問答品質、問答內文及語意契合度；於後者中，本研究則整合頻繁項目集為基分群法、向量空間模型及語意關係距離解析等技術，以解析得成員領域知識可參考程度。

承接於本研究發展之模式，為確認本方法論於實務應用中之可行性，本研究乃建構一套以網際網路為基礎之虛擬社群知識審核系統。此外，為驗證本系統績效，本研究將以虛擬社群真實問答資料作為驗證資料之樣本，並以當中之精華知識文章，作為領域訓練資料樣本，針對「問答契合度解析」與「成員參考度解析」進行獨立驗證，確保兩相議題間之驗證結果不相互影響。另一方面，本研究亦以案例為導向探討整合兩核心模組，所建構之「虛擬社群知識審核系統」，於實務情境中之應用與管理意涵，並於最終質化探討之結果中，討論得本系統具備相當程度之實質管理效用。

整體而言，本研究發展之模式與系統，於社群管理者層面，將為其提供標準化且具公正性群眾智慧品質之管理指標，以提升群眾智慧管理之實質效益，於知識需求者層面，則降低其篩選知識時所需花費之成本，且同時提升其獲取所需且正確知識之機率，減少錯誤資訊再傳遞之行為，最終將藉以增加實具品質群眾智慧之累積，促使更多社群成員知識分享之意願，以活絡虛擬社群之利用率，進而催化虛擬社群持續性之整體發展。

中文關鍵詞：知識管理、虛擬社群、知識分享、知識審核、問答契合度解析、成員參考度解析

英文摘要：In this information explosion era, the quality of collective intelligence is not easy to control. Knowledge demanders are required to invest an amount of time in evaluating the collective intelligence quality of virtual community to acquire correct knowledge. In addition, some knowledge providers might provide poor quality information in this circumstance and overall collective intelligence quality would be affected. Among that, there are two issues need to be discussed. Firstly, the question and answer suitability cannot be guarantee and secondly, member's (knowledge provider) reference degree cannot be determined. Therefore, to improve the accumulation of

quality of collective intelligence and encourage more members to share knowledge in virtual community, this paper develops a Virtual Community Knowledge Examination model including Question and Answer Suitability Determination (QASD) module and Member's Reference Degree Determination (MRDD) module to enhance collective intelligence quality. In QASD module, this paper employs keyword extraction, article's critical information extraction, article similarity determination and semantic analysis techniques to determine the suitability of Q&A content. In MRDD module, this paper integrates frequent itemsets-based clustering, vector space and semantic analysis technologies to evaluate the reference value of knowledge providers. Finally, in order to demonstrate applicability of the proposed methodology, a web-based system is also established based on the proposed model. Furthermore, a real-world case is applied to evaluate the proposed model. As a whole, this paper provides an approach for virtual community to efficiently examine knowledge to facilitate knowledge sharing intention.

英文關鍵詞：Virtual Community, Knowledge Management, Knowledge Share, Q&A Suitability, Reference Degree Determination

# 一、報告內容

## 1. 研究動機與目的

虛擬社群盛行的現今，社群成員透過開放式之網路社群以發佈與回覆等行為，進行主題式之討論藉以群聚相關之知識，而因虛擬社群之蓬勃發展，如「批踢踢實業坊」及「Yahoo 奇摩知識+」等，如此實具規模性與熱絡性之虛擬社群，將促使更多社群成員知識分享之意願，以活絡社群成員間交流與知識流通之行為，進而提升虛擬社群整體發展與造就群眾智慧累積 (Kim、Zheng 與 Gupta (2011))。

然而，群聚大量之知識固然容易，但針對群聚實具品質之知識而言，實則存在相當之挑戰，因資訊共享之便利性，促使大量知識相繼湧入虛擬社群，而針對數量龐大群眾智慧之管理而言，多數虛擬社群中之管理者以人工方式移除違規之發言，對於低品質群眾智慧之散播難以進行管控，是故，群眾智慧品質管理與衡量之任務，將仰賴於社群成員間自發性之評論及評分，而成員衡量內容之標準將溯根於己身知識專業能力、討論串中之評論，以及對知識貢獻者專業能力之了解，如此將因成員認知之不同而無公定之標準 (Tsai 與 Pai (2013); Moskaliuk 等人 (2012))，因此，知識需求者仍須透過蒐集、深入了解與學習領域專業知識等過程，方以對資訊可參考價值進行相關評估，但如此將因對領域知識了解程度之不足，而產生低品質之判定，亦將因須蒐集眾多資訊，而致使時間與評估知識正確性成本之提升，故當社群成員持續性地傳遞低品質資訊時，就知識需求者而言，將因蒐集到低品質資訊而產生額外評估知識之成本，就知識貢獻者言之，將降低社群成員持續分享知識之意願，而綜觀於對整體虛擬社群之影響，將導致社群中群眾智慧整體品質之下降，進而影響虛擬社群之發展 (Hung 與 Cheng (2013); Fang 與 Chiu (2010))。綜合上述，目前虛擬社群群眾智慧運作機制 (即虛擬社群群眾智慧運作之 AS-IS 模式)，如圖 1 所示，本研究乃將問題彙整並列點如下：

1. 知識需求者不易判定群眾智慧可採納程度：雖虛擬社群存在大量群眾智慧供知識需求者參考，但因資訊爆炸且群眾智慧品質控管不易落實等因素，致使相似內容同時存在許多差異大之評論，於此，知識需求者為取得正確知識，將須投入相當之時間本於評估群眾智慧之品質及社群成員可參考之價值 (Tsai 與 Pai (2013); Moskaliuk 等人 (2012))。
2. 知識貢獻者持續分享知識之意願不易維持：由於資訊正確性評估所面臨之挑戰，將致使部分知識貢獻者於未知狀況下傳遞低品質之資訊，當低品質資訊持續累積至一定程度並影響整體群眾智慧之品質，將影響知識貢獻者分享知識之意願 (Hung 與 Cheng (2013); Fang 與 Chiu (2010))。

有鑑於上述問題，本研究乃建構「提升成員知識分享意願之虛擬社群知識審核」模式，以審核虛擬社群群眾智慧品質，由於管理者管理社群成員之原則性及公平性，將影響成員間相互之信任關係，進而改變分享知識之意圖 (Fang 與 Chiu (2010))，故本研究乃提供具公正性虛擬社群問答契合度，以及成員參考度衡量指標，藉以改善既有虛擬社群群眾智慧運作機制中存在之問題。當中，本模式乃包含「虛擬社群問答契合度解析模組」，以及「虛擬社群成員參考度解析模組」，於問答契合度解析之面向，主要以發問者「發問標題」、「發問內容」與「討論串所屬領域」，以及回答者「回答內容」等資料做為分析之基礎，以解析得目標虛擬社群討論串提問與回答之整體問答契合度得分；於成員參考度解析之面向，主要以虛擬社群成員之「歷史發文」為基礎，以解析得目標虛擬成員之領域知識可參考程度。



圖 1、虛擬社群群眾智慧運作既有模式



圖 2、虛擬社群群眾智慧運作期望模式



綜上所述，知識貢獻與蒐集行為之產生，將可溯根於自我效能、認知相對優勢、認知相容性，以及互惠規範 (Chen 與 Hung (2010))，是故，本研究發展「提升成員知識分享意願之虛擬社群知識審核」模式，並透過對「虛擬社群問答契合度解析模組」與「虛擬社群成員參考度解析模組」之應用，以自動化之方式同時衡量與評估得群眾智慧之品質，以及知識貢獻者可參考之程度。於知識貢獻者層面，將因本研究所提供公正性成員參考度之衡量，提升其自我效能及認知相對優勢；於知識需求者面向，將因本研究具備相當正確性問答契合度之解析，提供予知識需求者相容於其認知之分析結果；對社群管理者而言，其將可參考對應數據，進行相關管理措施之制定 (期望模式，如圖 2 所示)。

## 2. 文獻回顧

本研究所涉及之研究主題乃包括「虛擬社群成員行為之探討」及「虛擬社群知識內容之探勘」等兩大研究方向，以下即針對此兩項主題之相關研究進行文獻回顧及探討。

### 2.1 虛擬社群成員行為之探討

對於虛擬社群成員行為之議題而言，本研究針對「虛擬社群知識凝聚行為」及「虛擬社群知識應用領域」進行相關文獻探討，以期觀察並探討得包含虛擬社群知識凝聚，以及領域知識應用等更具全面性虛擬社群成員之行為。

#### (I) 虛擬社群知識凝聚行為探討

##### (A) 個體成員知識分享意圖

針對虛擬社群持續性之發展言之，成員自發性參與關鍵因素之探討尤為關鍵，因此，Fang 與 Chiu (2010) 分別針對成員與管理者角色，進行自發性分享知識意圖之探索。再者，Tsai 與 Pai (2013) 提出整合遠、近端自發性分享影響因子及調節變項之探索模型，於定性面，該研究藉訪談歸納得成員「接納性 (接納不同意見)」、「樂於助人程度」、「參與性」，以及「社群資訊可用性」為遠端影響因子，而成員之「年齡」、「過去行為」與其「社群威望」為調節變項；於定量面，該研究根據問卷與社群中之歷史資訊，歸納得成員之「滿意度」及「可識別性」為近端影響因子，綜上而論，該研究認為成員自發性之參與，可歸咎於對社群之整體滿意度及其於社群中之可識別性，亦可溯根於成員己身之態度與背景、其他成員之熱絡度，以及社群中資訊之可用性。

其次，Zhao 等人 (2013) 以「非自覺性意象」及「自覺性意象」進行深入探討，兩者將相互影響以產生「聯合作用」，於此，該研究蒐集問卷並進行量化分析與討論，從中探討得管理者應以新成員為目標，進行獎勵機制策略之訂定，令其感受貢獻知識所獲致之利益，提升其參與社群時間及貢獻知識頻率，著手「自覺性意象」之於「非自覺性意象」之轉換，促使成員自發性之貢獻，進而為虛擬社群提供持續性發展之契機。最後，Zheng 等人 (2013) 整合資訊系統成功模式，以及資訊系統從後採用研究概念，藉以探討社群成員持續分享知識之動機，該研究歸納得管理者應將社群視為一資訊系統，持續管控知識品質，且同時提供成員相當品質社群平台之服務，並以提升成員需求及滿意度作為管理策略施行之方針，藉以維繫成員持續使用社群之動機，進而為虛擬社群注入永續發展之動力。

另一方面，Hung 與 Cheng (2013) 結合科技準備度與科技接受模型等理論，以探討各構面對成員知識分享意圖之影響，該研究認為知識分享意圖於成員自身之層面中，主要取決於個人對於知識所抱持之樂觀度與創新度中，而於成員分享之知識內容本質層面中，則取決於該知識之易用 (或易理解) 及有用性上。Chen 與 Hung (2010) 提出一套以「社會認知理論」為主，且以「社會交換理論」為輔之整合型探索模型，以針對專業型虛擬社群中成員分享、蒐集知識，以及提升虛擬社群運作效能影響因子進行相關洞悉。甚之，Jin 等人 (2013) 以期望確認理論為核心架構，提出一套整合知識分享因子與自我效能之模型，該研究認為貢獻者持續分享之關鍵因素，將溯根於分享時他人對其之評價，以及己身對收穫之認知，而上述因素亦提升貢獻者知識自我效能及確認感，使其獲致相當之滿足感，進而提升持續分享知識意願。而 Kim、Zheng 與 Gupta (2011) 認為虛擬社群服務提供者須注重之關鍵乃維持成員之社群參與度，另一方面則須透過相關策略之實行，如獎勵、成員推薦、文章推薦，以提升社群成員之善意性、社交技巧性及創造力。

再者，Kim、Song 與 Jones (2011) 結合社會認知與目標設定理論，提出服務提供者應建構相對容易使用之社群環境，以直覺且簡便化方式呈現原為龐大且繁雜之資訊，提供予成員良好使用者體驗，藉以提升其自我效能及愉悅性，並同時降低所須耗費時間，使成員願意設置及達成認知目標，進而提升整體虛擬社群知識間之運用及流動狀況。最後，Yan 等人 (2013) 認為管理者應透過對外在環

境因子之調整，面向社群成員自覺愉悅性及焦點關注度之提升，進行策略性之規劃，進而於最終提升社群成員之創造力，藉以為虛擬社群之發展性提供更進一步之貢獻。

## (B) 群體成員群聚智慧

過去研究較少針對問答平台中群眾智慧對使用者帶來之效益進行整體化探索，因此，**Blooma 等人 (2013)** 以知識內容、認知過程及社交等使用者群聚知識觀點，進行定性層面探討，該研究發展之整合型量表為相關研究，帶來後續衡量績效（或影響力）之指標，且探索得關鍵效益之提升因子，亦提供企業知識分享平台管理之參考策略。接著，**Moskaliuk 等人 (2012)** 以維基百科知識分享環境為基礎，塑造模擬實驗環境，並邀請多名使用者進行知識之協同創建，該研究發現虛擬社群知識分享之環境中，若存在中度之相似與高度之相對知識，群眾智慧形成之效率及品質將達最高點，此外，該研究亦發現群眾智慧係由個體使用者，以及社會網路間相互之動態外化與內化過程形塑而得。

其次，為使社群探測之結果擁有語意特性以更具參考性，**Zhao 等人 (2012)** 整合社群個體分群及鏈結分析技術，發展以主題為導向之社群探測模式，以完成社群探測，而相關探測得之結果將可提供予協同學習、專家尋找等研究參考，藉以充分運用於群眾智慧探勘領域之發展。而 **Liu 等人 (2013)** 以重建為知識關係網路，並將重建過程中之知識關係網路視為趨勢演進圖，提供予決策者參考，進而使其可針對趨勢之演進與方向訂定相關決策目標。

另一方面，傳統分群演算法針較難以語意方式進行分群，故 **Agarwal 等人 (2010)** 提出一套以群眾智慧為基部落格文章分群演算法，透過對動態與即時更新群眾標籤資訊之運用，聚類得具備群眾智慧結果，以同時解決既有分群演算法對不具結構及參考資訊較少文章，所面臨語意相似解析之問題。

而為了探討群眾標籤對使用者行為造成之影響，**Cress 等人 (2013)** 提及群眾智慧將影響受測者標記行為，亦將觸發偶發性之自主學習，而學習行為將與群眾智慧結果進行雙向回饋，以修正得更為正確之標籤資訊（群眾智慧）。而 **Huang 等人 (2012)** 發展以語意標籤為基之群眾註記系統，藉此改善傳統以群眾自訂關鍵字為基之標籤註記系統，未考量知識模糊涵義議題，所致巨量搜尋結果問題。

另一方面，**Huang 等人 (2014)** 透過對社交網路概念之借鑑，建置使用者間社交網路之拓樸，將其應用於推薦機制中，除拓展並延伸群眾標籤資源之應用外，更同時活絡社群中使用者間之互動行為。再者，**Hong 與 Scardamalia (2014)** 年針對「Knowledge Forum」虛擬社群進行實驗研究，將此社群之知識文章劃分為兩資料集合，透過對「Analytic Toolkit (**Teplovs (2008)**)」詞彙分析工具之應用，以知識文章中之關鍵字為基礎，該研究以量化方式進行探討並從中發現，相似知識關鍵項目適合應用於知識自動化之評估、學習型虛擬社群知識整合之應用，以及標準化群眾知識之產生，提供予相關虛擬社群研究參考，將可應用於群眾智慧分析及自動化群眾知識之產生。

## (II) 虛擬社群知識應用領域探討

### ● 教育領域

為即時獲取學生之學習狀況，**Tobarra 等人 (2014)** 蒐集學習型虛擬社群中學生討論過程之資料，解析得學生討論主題之偏向性，並以「標題網路」之方式呈現標題間之相關性，將相關結果提供予教育單位參考，進而提供除面對面授課以外獲取學生問題之管道。再者，**Dasalu 等人 (2014)** 認為多（跨）學科學習小組之成立乃學習型虛擬社群須關注之重點，該研究建議群聚得最佳化多學科學習小組，而如此學習小組之劃分將能促進成員間多樣化知識之交流，進而凝聚得更具價值之群眾智慧。而為了於網際網路中提升教師之專業能力，**Tseng 與 Kuo (2014)** 認為必須維持教師自發性參與社群之活動力，方以進行社群實踐，藉此活絡成員間知識交流行為，進而提升教師之專業能力。

### ● 醫療領域

為協助領域專家管理龐大之討論資訊，**Huh 等人 (2013)** 發展醫療討論文章分類方法，可針對文章進行自動化之分類，並呈現予領域專家參考，於此，各領域之專家即可選擇與其領域性專業契合之討論串，提供相關醫療資訊之回覆，藉以提升整體回答之品質與效率，進而改善巨量討論資訊增加所致之影響。其次，**Alali 與 Salim (2013)** 整合科技接受及資訊系統成功模型，發展多維度虛擬社群實踐成功模式，以探討提升醫療領域知識分享行為之因素。

再者，針對於雲端醫療服務導入社交網路概念對使用者之影響而言，少有研究由統計理論之層面進行探討，有鑑於此，**Sato 與 Costa-i-Font (2013)** 以問卷方式蒐集大量資料，並以機率單位模型 (Probit Model) 進行解析，以分析得受測者對使用社交網路型雲端醫療服務之可能性。甚之，**Ba 與 Wang (2013)** 發展一套結合量化研究概念與統計之衡量方法，藉以探討成員於健康資訊虛擬社群中自發性分享之因



素，並衡量其分享之動力。

## ● 其他領域

為了以動態與即時方式針對社交媒體進行趨勢之分析，Stieglitz 與 Kaufhold (2011) 建構以公共社交媒體為導向之自動化全文分析系統，藉以探測公眾議題之成長趨勢。為改善資訊超載之問題，Liu 等人 (2014) 針對問答型虛擬社群發展問答主題知識推薦模式，該研究即根據成員與群體偏向性之關係，推薦符合其偏向主題知識，除改善成員擷取知識時所面臨知識正確性之問題外，更降低成員獲取所需知識之時間成本。再者，為拓展社交網路結構中消費者關係鏈結性之應用，Wang 等人 (2013) 藉由社交網路分析與資料探勘技術之整合，發展消費者興趣偏好探索與搜尋推薦系統。此外，有鑑於僅以單一應變中心進行決策，實難為災害規劃相對完善之應變措施，Li 等人 (2014) 建置以虛擬社群概念為基協同資訊共享與整合系統，以整合並管理各地區緊急事件之資訊，藉由社群化及雲端進行跨地區緊急應變措施策略協同規劃，以訂定更為完善之緊急應變措施。

## 2.2 虛擬社群知識內容之探勘

針對虛擬社群知識內容探勘技術言之，為能解析虛擬社群知識群聚之程度，本研究以「虛擬社群問答內文契合狀況解析」及「虛擬社群成員個人資訊解析」作為文獻回顧之主軸，探討過去研究以更深入了解相關之做法，探討相關研究之差異，以此為本研究之模式帶來更具全面與完善之發展。

### (I) 虛擬社群問答內文契合狀況解析

#### (A) 問答核心資訊萃取

為了於以問題與回答為導向之虛擬社群中，過濾不具意義之回答並取得契合於問題之答案，Wang 等人 (2013) 根據討論串中使用者所回覆之對象，進行相對應主題分割之動作，蒐集相同對象之回覆內容以成為相關之內文集合，接著，則以人工之方式進行「正面」與「負面」情緒指標之界定；而於答案探測議題中，該研究認為最佳解答，大多屬於「正面」情緒指標，因此，該研究即蒐集具正面情緒指標之回答內容，並藉支持向量機 (Support Vector Machine; SVM) 之運用，訓練相關之答案語料庫，以於討論串中各筆回覆中判定得最佳解答，並提供予知識需求者參考。再者，Ko 等人 (2010) 提出整合邏輯回歸模式之答案排名與合併系統，藉回歸方式預測候選答案成為最佳解答概率，以改善既有研究於篩選答案時較無依據之狀況。

此外，為了萃取各種線上論壇中之內文及提問之解答資訊，並以結構化之方式進行表達，Cao 等人 (2011) 亦於結構支援向量機中，增添更為完善演算與推理之法則，使該研究所提之方法能更廣泛地應用於萃取各種不同形式線上論壇中之內文及解答。甚之，Moreo 等人 (2012) 認為關鍵字之擇選將因領域專業知識，以及擇選標準嚴謹性不足等問題，而致萃取得關鍵性重點資訊偏差之疑慮，因此，該研究發展一套「最小化微型表達 (Minimal Differentiator Expressions; MDE)」演算法，以自動化方式萃取提問內容之關鍵重點。

而為了準確分析得部落格使用者撰寫之文章中，主題與內容代表之外顯內文核心，以及內隱潛在涵義，Tung 與 Lu (2012) 建構一套分析隱含於部落格文章事件、情緒與需求模式，以及情緒對事件與需求之相關性分析，於不具結構性之部落格文章中，準確分析得代表該篇文章之核心事件，以及發文者於文章中所隱含之需求及情緒。再者，Liu 等人 (2011) 發展一套得以解決上述狀況之自動化論壇使用者評論萃取系統，該研究提出兩種萃取方法，分別為「直接萃取法」及「封裝為基萃取法」，當中，前者係應用於多筆不加以類別區分，且具備多個評論討論串之論壇資料，進行相關評論資訊之萃取，而後者則可應用於上述之反例，是故，該研究以上述不同面向為出發點，提供完善之資訊萃取法，以使使用者能於不同狀況中，萃取得最具代表性之評論數據。

此外，有鑑於既有資訊檢索相關研究大多僅以統計式方法解析文章關鍵資訊，Ferrández (2011) 提出一套以詞彙與語句探索為基之資訊檢索模式，以詞彙及語句於文章中之涵義為切入點，進行文章關鍵資訊之萃取與相似度之計算。而 Sorg 與 Cimiano (2012) 提出以維基百科為基跨語言精確語意分析模式，藉由「維度推導」、「概念關係強度」、「內文相關性」與「概念空間」之解析，衡量文章相關性，以獲致跨語言文章間內文與概念整體相關性程度，藉此為文章核心概念萃取領域增添跨語言文章處理能力，並提升文章相似度排名與檢索精確度。

另一方面，因網路蓬勃發展所衍伸的資訊爆炸問題，導致既有虛擬社群逐漸充斥龐大且零散之知識，有鑑於此，Chen 與 Chen (2012) 提出一套專業型虛擬社群之知識探索方法，該研究依序透過「萃取經驗知識」、「驗證知識品質 (經領域專家)」、「建構知識資料」與「推論並彙整知識之領域類別」

等處理過程，於專業型虛擬社群中進行經驗知識之統整，並於最終提供相關知識之彙整索引予知識擷取者，降低其取得所需知識所花費之時間成本。為了縮短電信服務中排解故障與客戶諮詢所耗費之時間成本，Iwashita 等人 (2011) 提出一套語意分析及分類方法，針對過去客戶諮詢所保留之提問內容，進行相關語意特性之解析，並根據解析結果進行相對應問題屬性之定義與分類，而服務提供者將可透過對提問內容所屬類別，以及其所對應之解決方案進行檢索，以更具效率之方式，排解較具複雜度客戶諮詢之問題。

## (B) 語句及詞彙相關性分析

詞彙與詞彙，以及語句及語句間語意相關性之分析係自然語言處理中重要的一環，然而，針對於中文語句之解析，Yan 等人 (2009) 運用「賓州中文樹庫」並透過「單純貝氏」、「決策樹」與「最大熵」等分類法，進行訓練資料之匯入，以檢驗上述方法於解析中文詞彙語意相關性之準確性，因此，該研究整合三種分類法之特性，發展整合型中文語意相關性分析模式，不僅止於詞彙與詞彙間語意相關性之分析，該模式亦可解析完整語句間之關聯性，是故，藉由該模式之運用，將可以更具全面（整合三種分類法）之方式，分析並計算得詞彙間及語句間語意相關性之程度。

而為了於不建構語意資料庫之條件下判定詞彙間之關係程度，Cilibrasi 與 Vitanyi (2007) 建構以 Google 搜尋引擎為基之詞彙語意相關度解析方法，並將運算後結果命名為「標準化 Google 距離 (Normalized Google Distance; NGD)」，該研究所提之方法係透過 Google 搜尋引擎，針對兩詞彙之搜尋結果，統計當中之各獨立詞彙，以及同時包含兩詞彙搜尋所得之結果數，以分析兩詞彙於網際網路中所呈現之標準化動態邏輯距離，進行相關詞彙語意關係之解析。有鑑於上者，為使詞彙間關係距離之計算更具精確性，Chen 與 Lin (2011) 修正原有之 NGD 演算法，針對原始研究中對「Google 搜尋引擎所含之總網頁數」未曾進一步探討之層面，提出其觀點並同時建構改良後詞彙關係距離解析方法，即 Google 核心距離 (Google Core Distance; GCD)，以使詞彙關係距離之解析結果更具精確度，此外，該研究亦根據其改良後之演算法提出具體之詞彙關聯程度彙整方式，即跳躍路由 (Hop-by-Hop Routing; HHR) 演算法，並於最終藉 Google 核心距離及跳躍路由演算法之兩相結合，建構出完整之詞彙關係網路 (Word AdHoc Network; WANET)，以此擴增詞彙關係距離之相關應用層面。

甚之，對於自然語言處理之領域而言，語句及短文本語意相似度之解析至今仍實存重要性，有別於一般之研究，Oliva 等人 (2011) 認為一段語句的整體涵意，並非僅取決於語句中各獨立詞彙所蘊含之意義，各詞彙間的排序方式（即語法）亦為需考量之重點，此外，該研究亦根據過去研究中之結果，即不同人對於語法與語意間關聯程度之認定有所差異，以提供語法權重調整之功能。再者，Wenyan 等人 (2010) 提出結合語意及統計資訊短文本相似度解析模式，透過該模式所建置之文本向量亦可運用於其他資訊檢索技術中，如以文本分群技術為例，套用該研究文本向量數據，經實驗後將獲致績效提升之結果。

另一方面，大多數分群技術較少考量詞彙間之語意關係，以致難以準確分群較無結構化且字詞數較少之虛擬社群文章，故 Zheng 等人 (2009) 融合名詞短語檢測之概念於評估多個 WordNet 為基分群演算法之效能，而該研究評估所得之結果，將可為具備語意概念分群演算法之研究帶來語意關係類型選定之參考基準，藉此提升語意分群時之精確度。

而 Karamolegkos 等人 (2009) 針對目前較具代表性之分群演算法，包含階層式、K-means 與 Spectral 等分群法，進行分群效能與效率之評比，於最終評估結果中得知，Spectral 分群法於複雜性與正確性中擁有最佳之績效，而階層式分群法於新成員加入時具備最佳之績效。

此外，為了於句法分析樹中解析得無法直接觀測關鍵字之語意，並同時將此應用擴展至段落、句子、短語與詞彙中，Galitsky 等人 (2012) 發展一套以探勘句法分析樹為基之語意屬性推導模式，首先，該研究藉由邏輯節點間之相互匹配，於句法分析樹中取得不可測關鍵字之語意，且同時為文章、段落、句子、短語及詞彙間之相似度進行排名，於此提供更精確語意相關性之參考數據，並同時為句法分析樹之探勘提供更為廣闊與深入之應用。而 Li 等人 (2011) 提出應用簡要語意解析之快速文章分類模式，有別於既有研究大多以全文本之方式進行解析，該研究探勘核心概念語意相關度之作法，將大幅提升模式運作之效率。

整體而言，就虛擬社群問答內文契合狀況解析議題言之，於問答核心資訊萃取之層面，大多以詞頻為基進行核心資訊之萃取（如：Wang 等人 (2013)；Moreo 等人 (2012)；Cao 等人 (2011)；Sorg 與 Cimiano (2012)），而就少數具備語意概念之相關研究而言，需先以人工方式進行資料之預處理，



或藉語料庫訓練相關資料，方以進行後續之解析（如：Ferrández (2011)；Iwashita 等人 (2011)）；於語句及詞彙相關性分析之層面，相關之研究大多朝向具備語意概念之解析進行發展，而語意之解析主要分為「以語料庫為基」與「以上下文為基」兩種方式，前者需先以語料庫進行訓練方可進行解析，後者則以內文之上下位關係，分析得詞彙或語句之相似性。整體而論，針對人工資料預處理，以及語料庫訓練之部分言之，雖最終解析得之結果具備相當之正確性，但其所需花費之人力及時間成本，將遠大於自動化處理之系統，且以人工進行處理亦可能發生判定標準不一之狀況（如：Wenyin 等人 (2010)；Li 等人 (2011)；Oliva 等人 (2011)），而就以上下文為基之語意分析言之，雖無上述之問題，但針對於虛擬社群中之討論串而言，大多為篇幅較短之片段留言，若以上下文判斷語意將致正確性之疑慮（如：Yan 等人 (2009)；Galitsky 等人 (2012)），因此，為了避免上述之問題，發展一套整合語料庫與上下文之自動化解析，且通用於虛擬社群之方法論係本研究於「虛擬社群問答契合度解析模組」之目標。

## (II) 虛擬社群成員個人資訊解析

### (A) 成員專業領域偏向分析

有鑑於知識需求者難以由個人檔案中獲知分享者專精之知識領域，Vertommen 等人 (2008) 提出多維向量個人檔案萃取模式，藉由對使用者撰寫與閱讀文章之蒐集，從中分析其專業知識領域偏向。其次，Meo 等人 (2011) 提出通用型相似使用者、資源與社交網路之推薦模式，首先，該研究將使用者、資源與社交網路皆視為節點，並以超圖 (Hypergraph) 方式進行表示，次之，該研究針對使用者間之會員資格、友誼關係、歷史發文，以及評價紀錄進行相似度之解析，並以超圖呈現相關使用者個人化之數據，最終，該研究即根據上述數據，藉標準化之方式建置通用型之使用者個人檔案，不以單個社交網路為限，築起全球使用者推薦之管道。

再者，為改善社群中惡意資訊之擴散，Caverlee 等人 (2010) 建置一套社交信任 (SocialTrust) 系統，藉此管理使用者間互動時衍伸之社交資訊。而 Han 與 Chen (2009) 提出一套以模糊分群法為基之本體使用者個人檔案建構模式，該研究由使用者行為紀錄中，蒐集資訊並作為建構個人檔案之依據，並結合「模糊 C 均值 (Fuzzy C-means; FCM)」、「Powell Hestenes Rockafellar; PHR」與「模擬退火 (Simulated Annealing)」等演算法，發展整合型模糊分群法，以模糊方式進行相似使用者之聚類，並藉 PHR 與 OSGP (Optimal Solution under Given Parameters) 之整合型演算法解析分群之結果，以更為清晰之階層本體化方式呈現使用者專業領域之偏向。

另一方面，為了針對知識擷取者推薦與其需求 (興趣) 知識相符合之知識專家，Li 等人 (2012) 發展一套提升線上論壇知識分享效能之社交推薦機制，該研究以社交網路之形式，適當地推薦知識擷取者感興趣或所需之知識，以此提升線上論壇中知識分享之效能。甚之，Ni 等人 (2012) 提出一套以主題為基虛擬社群成員知識偏好判定模式，該研究以潛在狄利克雷分配模型 (Latent Dirichlet Allocation; LDA) 為基，將當中原先僅以單個提問作為主題萃取依據之方式，修改為以成員所有歷史發問為基礎之主題萃取方法，以解決既有研究所面臨因詞彙共同出現數過少，以及未考量知識領域之特性，所導致主題萃取可信度不足之問題。再者，為了進行相似專業背景使用者間之推薦，Maleszka 等人 (2013) 提出一套運用知識整合工具，以及階層式使用者個人檔案結構之協同合作知識成員推薦模式，於最終，該研究將已建構完成個人檔案進行分群，藉以取得具備相似專業背景之使用者群集，於此推薦適切之協同合作成員，以提升知識創新之契機。而 Chen 等人 (2013) 發展個人化知識搜尋與推薦之領域知識本體調適機制，進行與特徵因子相符合知識之推薦，供使用者進行與領域背景相符合專業知識之學習，進而以知識管理之方式提升其專業能力。

此外，為了使成員間群聚實具品質之領域知識，Cantador 與 Castells (2011) 提出一套虛擬社群成員個人檔案之結構化、萃取與成員推薦模式，進行多層次成員間興趣內容之分群，以藉由群聚與推薦相似成員之方法，活絡其分享知識之互動關係，進而提升成員間群聚實具品質領域知識之可能性。

### (B) 成員知識可信度分析

有鑑於過去的知識專家發現系統較少針對一般性虛擬社群進行解析，且未同時考量分享者發文內容及其之可信程度，Wang 等人 (2013) 發展以標題感知為基知識專家發現方法，藉由「線性組合」、「級聯排名」與「定標策略」等運算方法之整合，進行知識專業程度數據之呈現，藉以提供知識擷取者更多知識專業度之參考面向。再者，有鑑於使用者評價紀錄過少，將降低推薦結果可信度，Lai 等人 (2013) 蒐集使用者評價資訊，藉模式化方式生成使用者個人檔案，並分析當中相似程度，以進行

相似使用者群組之聚類，其次，於個體面中，根據使用者參與狀況分析其影響程度，以計算得其可信度，於群組面中，則以群組成員對文章之共同評價狀況為依據，計算得評價資訊可參考程度，最終，即整合上述數據，取得混合型評價資訊可採納之程度，藉此解決因評價紀錄過少，而衍生推薦結果可信度不足之問題。甚之，Liu 等人 (2013) 提出一套整合成員個人檔案、成員評價與知識類別鏈結分析之問答平台專家查找模式，該研究蒐集使用者回答內容與評價等歷史數據，並藉中研院 CKIP 斷詞系統與 TF-IDF 之兩相結合，進行類別相似度彙整之資料預處理，接著，即藉該研究所提之模式分析得成員知識績效整體得分，以及成員可信度得分。

另一方面，Daud 等人 (2010) 認為傳統專家尋找模式，未考量時間因子之作法，將致時效性問題，進而衍伸正確性疑慮。Schall (2012) 提出一套透過社群成員活動與上下文鏈結測量之專業程度排名模式，該研究認為社群成員之可信度、活動水平與預期訊息量為衡量專業度之指標。再者，為了準確鑑別得社交網路中之意見領袖，Li 等人 (2013) 以學習型線上社群為基發展社交網路意見領袖衡量之混合型系統，該研究針對發文數量與發文期間進行標準差之運算，以取得使用者執著度，爾後，針對執著度進行發文類型與回覆類型之統計，以取得集中度；最後，將全部數據予以彙整及整合，藉以取得使用者作為意見領袖之得分。甚之，Lykourantzou 等人 (2010) 提出一套以維基百科概念為基提升群眾知識品質之自適調節系統，該研究藉「分類」、「使用者互動」、「整體風險」、「推薦」與「協同合作」等五項品質指標之自動化衡量，以提供管理者系統風險管理之指標，並於達到相對應風險等級時以「警示匯報模組」提醒管理者，建議其針對相應之部分進行人為之管理，以避免因某些成員的濫用，爾或是無適時管理所致知識品質之下滑，進而影響其他成員使用系統之意願。

此外，有鑑於人際關係間之信任於社交網路中扮演重要之角色，Zolfaghar 與 Aghaie (2012) 提出整合上下文及結構型資料使用者信任度評估方法，以取得使用者之可信度，建置使用者間信任關係拓樸，以作為網站管理員管理使用者間之依據及標準。其次，Kim 與 Ahmad (2013) 認為過去研究著重於可信度之衡量，而較忽視於「不可信程度」對使用者信任度之影響，該研究將證據理論應用於知識發布者可信度之評估中，發展一套混合型可信度評估模式，提升其於網路社群中擷取正確知識效能。

再者，有鑑於網路社群中使用者討論之議題，經常充斥兩極化之評價，以致知識需求者須花費龐大時間於評估知識可信度，Kim 與 Phalak (2012) 根據其他使用者與目標發布者間評價之關聯性，以分析其於社群中之影響性，計算得整體可信度，供知識需求者作為知識內容評估之依據，進而減少所需耗費評估知識可信度之時間成本。而 Korovaiko 與 Thomo (2012) 針對使用者於社群中之背景資料，諸如發布之文章、評論，以及對其他使用者或產品之評價進行蒐集，並將使用者間信任之相關性訂定並劃分成「類別評價相似性」、「類別偏好相似性」與「類別評論相似性」等類型，同時計算與解析得使用者間之信任相關性及隱含信任度，藉社交網路概念建置各使用者間信任相關性之連結，以提供知識需求者參考。

整體而論，經由上述之文獻回顧，本研究歸納得，針對於成員之解析可面向成員個人檔案、歷史使用行為，以及成員間互動關係等資訊進行蒐集並予以深入分析，就成員歷史使用行為言之，多數解析成員歷史發文之研究僅以詞頻之方式進行解析，未考量詞彙間語意之特性，由於專業型虛擬社群文章較具結構性與辨識性（專有名詞較多），故於此類型之社群較易適用，而於一般性之虛擬社群中，將因文章較不具結構性且辨識性較低之因素而影響解析之正確性（如：Vertommen 等人 (2008)；Li 等人 (2012)；Wang 等人 (2013)；Liu 等人 (2013)；Zolfaghar 與 Aghaie (2012)；Kim 與 Phalak (2012)），此外，就成員互動關係言之，甚少研究針對於此進行整合性之解析（如：Chen 等人 (2013)；Daud 等人 (2010)），且考量成員互動關係之研究，大多僅著重於成員與成員間之關係，少有研究針對成員與群眾智慧間之關係進行探討（如：Meo 等人 (2011)；Caverlee 等人 (2010)；Kim 與 Ahmad (2013)；Han 與 Chen (2009)），然群眾智慧亦實存價值，是故，為設計一套更具全面性之方法論，本研究將藉探討所得之結論，進行相關之改善，以發展一套同時具備詞頻與語意解析，以及考量成員間與群眾智慧間觀點之「虛擬社群成員參考度解析模組」。

### 2.3 小結

本研究之研究主題涉及「虛擬社群成員行為之探討」與「虛擬社群知識內容之探勘」等兩大研究方向。於「虛擬社群成員行為之探討」議題中 (2.1 小節) 得知，於個體成員知識分享意圖之層面中，主要針對成員之知識分享意圖進行探討，並以質化及量化之方式探討得成員自發性分享知識，以及影響其分享意圖之關鍵因素；於群體成員群眾智慧層面中，主要針對群眾智慧之影響力，進行相關之探討，並獲致群眾智慧將為成員及虛擬社群帶來正向影響之結論。此外，於各領域應用層面中，主要面



向巨量知識資料之探勘，以及群眾智慧之應用層面，進行相關演算法之設計與系統之建置，期望透過對虛擬社群中龐大之資料，以及雲端化平台之應用，以提供知識與服務需求者更具價值之實際應用。

於「虛擬社群知識內容之探勘」議題中 (2.2小節)，於問答核心資訊萃取之層面，大多以詞頻為基進行核心資訊之萃取，而少數具備語意概念之研究，需先以人工預處理資料，或藉語料庫進行訓練，方以解析；於語句及詞彙相關性分析之層面，大多朝向具備語意概念之解析進行發展，且分為「以語料庫為基」與「以上下文為基」。另一方面，於成員專業領域偏向分析層面，大多以成員之個人檔案及歷史使用行為作為分析基礎，有少數研究以成員間互動關係之觀點為分析資訊；於成員知識可信度分析之層面，大部分朝向成員之歷史發文進行解析，而解析方式多以詞頻為基進行相似度之衡量。

### 3. 提升成員知識分享意願之虛擬社群知識審核模式

本研究所提之「提升成員知識分享意願之虛擬社群知識審核模式」主要針對「虛擬社群問答契合度」，以及「虛擬社群成員參考度」等兩面向之議題進行解析，首先，於問答契合度解析之部分，乃以發問者之「發問標題」、「發問內容」與「討論串所屬領域」，以及回答者之「回答內容」等資料做為分析之基礎，並結合「領域性關鍵詞彙集合」之概念、「文章內容核心資訊萃取」、「文章相似度判定」，以及「詞彙語意分析」等技術，以判定發問者及回答者之「提問品質」、「回答品質」、「問答內文契合度」與「問答語意契合度」，並於最終藉上述數據之整合，分析並計算得「目標虛擬社群討論串提問與回答之整體問答契合度得分」。另一方面，於虛擬社群成員參考度解析之部分，本研究期望以虛擬社群成員之「歷史發文」為基礎，透過對目標虛擬社群成員歷史發文所屬之領域，及其與該領域中其他討論串所對應之分群結果，以於最終獲致實具參考價值之目標虛擬成員之領域知識可參考程度 (如圖3所示)。

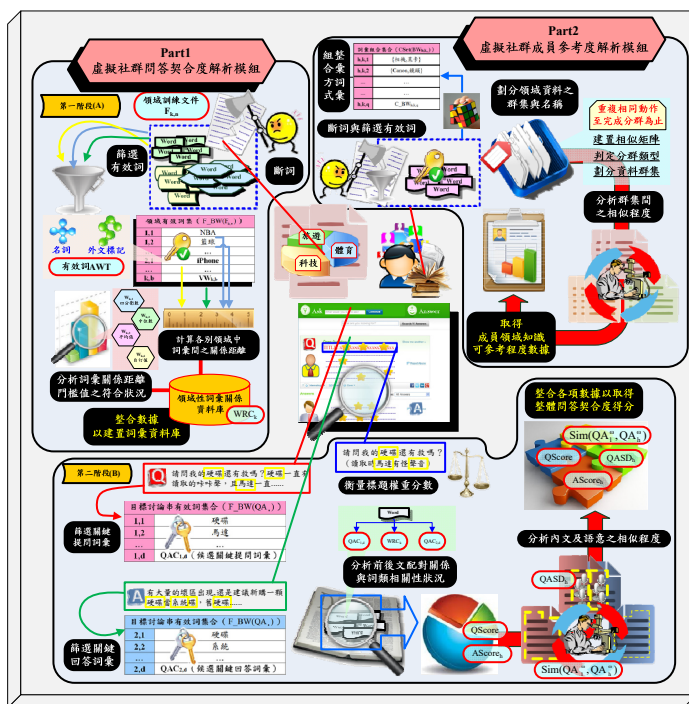


圖3、提升成員知識分享意願之虛擬社群知識審核模式架構圖

#### 3.1 虛擬社群問答契合度解析模組

##### 第一階段(A)—領域性詞彙關鍵詞彙集合之建立

本階段係運用中研院開發之 CKIP 中文斷詞系統，先行對訓練文件進行斷詞，分析詞彙間之關係，並依「中研院平衡語料庫詞類標記集」中對詞類定義之方式，於彙整各詞彙之詞類發生狀況後，將詞類屬於有效詞 (AWT) 之詞彙進行保留，除有效詞中定義之詞類外，其他詞類之詞彙將不予保留，以此過濾不具意義之詞彙，進而獲得相關之有效詞集。待獲得各領域中之有效詞後，即改良並運用 Cilibrasi 與 Vitanyi (2007) 於分析類別訓練文件中，以兩個有效詞為一單位 (組) 之各有效詞關聯程度，最後，對所有有效詞間之關聯程度進行組織及彙整，同時藉詞彙組合之分割處理，以實具合理之方式由詞彙組中取得獨立之詞彙，並令詞彙為節點且藉保留之詞彙關係，彙整並建構相關領域之「領域性關鍵詞彙集合」，以視後續第二階段方法論中核心 (概念) 萃取步驟之前置動作。此部分運作之方式及流程如圖4所示。

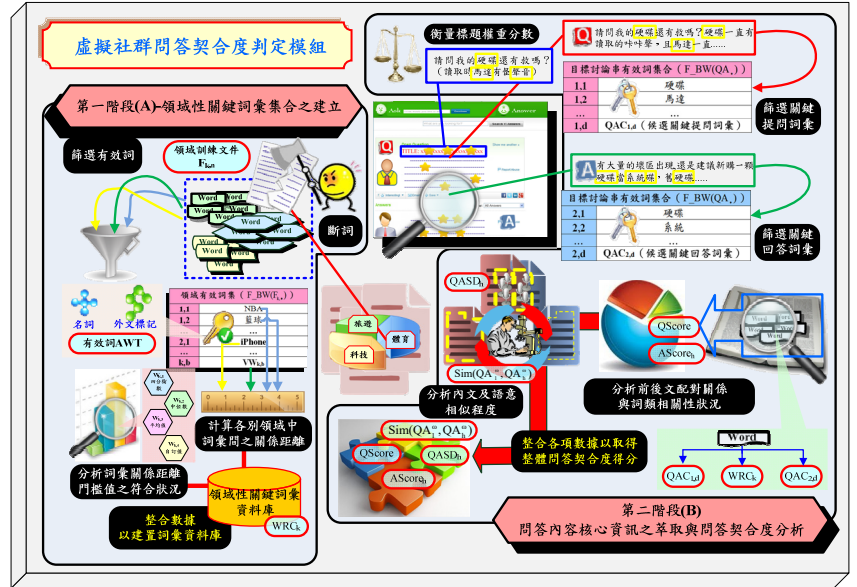


圖 4、領域性關鍵詞彙集之建立運作示意圖 圖 5、問答內容核心資訊之萃取與問答契合度分析之運作示意圖

### 步驟(A1)—對訓練文件斷詞並篩選有效詞

完成此任務之首要步驟，即選定欲建構領域性關鍵詞彙集之類型，爾後，使用中研院開發之 CKIP 中文斷詞系統，對已分類完成之所有訓練文件  $F_{k,}$  進行斷詞，以取得類別訓練文件中詞彙之分佈狀況  $BW(F_{k,})$  (如公式(1)所示)。其次，即根據「中研院平衡語料庫詞類標記集」中對詞類定義之方式，於彙整各詞彙之詞類發生狀況後，將詞類屬於有效詞 (AWT) 之詞彙進行保留，除有效詞中定義之詞類外，其他詞類之詞彙將不予保留，以此過濾不具意義之詞彙，並於最終獲得實具意義之領域有效詞彙集  $F\_BW(F_{k,})$  (如公式(2)所示)。

$$BW(F_{k,}) = \{BW_{k,1}, BW_{k,2}, \dots, BW_{k,a}\} \quad (1)$$

$$F\_BW(F_{k,}) = \{VW_{k,1}, VW_{k,2}, \dots, VW_{k,b} \mid VW_{k,b} \in AWT\} \quad (2)$$

### 步驟(A2)—計算領域有效詞間之關係距離

待獲得領域有效詞集  $F\_BW(F_{k,})$  後，即藉具專業領域知識訓練文件  $F_{k,n}$  (如虛擬社群領域性文章) 之挹注，同時結合知識分類概念與計算詞彙關聯程度之方法，除了運用 Cilibrasi 與 Vitanyi (2007) 中分析詞彙關係距離之方法外，本研究更同時將原為無知識領域概念之詞彙關聯程度計算方式，昇華為實具知識領域概念之方法論，於此透過以一個主要有效詞 ( $AVW_{k,i}$ ) 及次要有效詞 ( $BVW_{k,j}$ ) 為一單位之領域有效詞間搜尋結果數，亦即經 Google 搜尋引擎「站內搜尋」功能搜尋後之「主要有效詞搜尋結果數 ( $GSH(AVW_{k,i})$ )」、「次要有效詞搜尋結果數 ( $GSH(BVW_{k,j})$ )」與「同時包含主要與次要有效詞之搜尋結果數 ( $GSH(AVW_{k,i}, BVW_{k,j})$ )」，以及領域訓練文件之總數  $TF_k$ ，計算得領域有效詞間之關係距離  $RD(AVW_{k,i}, BVW_{k,j})$  (如公式(3)所示)，其結果整理如表 1 所示。當中，Cilibrasi 與 Vitanyi (2007) 所提方法論中之「Google 搜尋結果數」乃直接透過 Google 搜尋引擎搜尋所獲得，本研究於此更改為以「站內搜尋」進行搜尋，主要目的在於藉由站內搜尋之應用，將可針對已分類完成之領域文件進行搜尋，以於最終獲致具備「領域屬性」之搜尋結果。

$$RD(AVW_{k,i}, BVW_{k,j}) = \frac{\text{Max}\{\log GSH(AVW_{k,i}), \log GSH(BVW_{k,j})\} - \log GSH(AVW_{k,i}, BVW_{k,j})}{\log TF_k - \text{Min}\{\log GSH(AVW_{k,i}), \log GSH(BVW_{k,j})\}} \quad (3)$$

### 步驟(A3)—建立領域性關鍵詞彙集

本步驟係以「領域詞彙組關係距離」為基礎，即一個主要有效詞  $AVW_{k,i}$ ，以及次要有效詞  $BVW_{k,j}$  為一組合之領域詞彙組關係距離  $RD(AVW_{k,i}, BVW_{k,j})$  (相關方法詳述於「步驟 A2」中)，並令詞彙為節點且藉保留之詞彙關係，彙整與建構領域性關鍵詞彙集 (如公式(4)所示)。



$$ARD_{k,p} = \frac{RD(AVW_{k,i}, BVW_{k,j})}{\sum_{\text{all } i,j} RD(AVW_{k,i}, BVW_{k,j})} \quad (4)$$

最終，為使詞彙組合分離為獨立詞彙，首先，必須先行將符合門檻之領域詞彙組合予以分割，如公式(5)所示，其次，依據此些詞彙組內，獨立詞彙  $S\_HRD_{k,w}$  達成門檻值次數  $N(HRD_k, S\_HRD_{k,w})$  進行統計，於最終，若該詞彙達成門檻值之次數大於所有詞彙之平均值，則該詞彙即屬於領域性關鍵詞彙集合  $WRC_k$  中之詞彙，如公式(6)所示。

$$\text{If } (AVW_{k,i}, BVW_{k,j}) \in HRD_k \text{ Then } S\_HRD_{k,w} = AVW_{k,i} \text{ and } S\_HRD_{k,w+1} = BVW_{k,j} \quad (5)$$

Where  $S\_HRD_{k,w} \notin S\_HRD'_{k,\bullet}$ .

$$\text{If } N(HRD_k, S\_HRD_{k,w}) \geq \frac{\sum N(HRD_k, S\_HRD_{k,w})}{N(S\_HRD_{k,\bullet})} \text{ Then } S\_HRD_{k,w} \in WRC_k \quad (6)$$

表 1、主要有效詞與次要有效詞關係距離彙整表

主要有效詞 次要有效詞	AVW <sub>1,1</sub>	AVW <sub>2,1</sub>	...	AVW <sub>k,1</sub>	...	AVW <sub>k,i</sub>
BVW <sub>1,1</sub>	RD(AVW <sub>1,1</sub> , BVW <sub>1,1</sub> )	RD(AVW <sub>2,1</sub> , BVW <sub>1,1</sub> )	...	RD(AVW <sub>k,1</sub> , BVW <sub>1,1</sub> )	...	RD(AVW <sub>k,i</sub> , BVW <sub>1,1</sub> )
BVW <sub>2,1</sub>	RD(AVW <sub>1,1</sub> , BVW <sub>2,1</sub> )	RD(AVW <sub>2,1</sub> , BVW <sub>2,1</sub> )	...	RD(AVW <sub>k,1</sub> , BVW <sub>2,1</sub> )	...	RD(AVW <sub>k,i</sub> , BVW <sub>2,1</sub> )
...	...	...	...	...	...	...
BVW <sub>k,1</sub>	RD(AVW <sub>1,1</sub> , BVW <sub>k,1</sub> )	RD(AVW <sub>2,1</sub> , BVW <sub>k,1</sub> )	...	RD(AVW <sub>k,1</sub> , BVW <sub>k,1</sub> )	...	RD(AVW <sub>k,i</sub> , BVW <sub>k,1</sub> )
...	...	...	...	...	...	...
BVW <sub>k,i</sub>	RD(AVW <sub>1,1</sub> , BVW <sub>k,i</sub> )	RD(AVW <sub>2,1</sub> , BVW <sub>k,i</sub> )	...	RD(AVW <sub>k,1</sub> , BVW <sub>k,i</sub> )	...	RD(AVW <sub>k,i</sub> , BVW <sub>k,i</sub> )

## 第二階段(B)—問答內容核心資訊之萃取與問答契合度分析

本階段係以「第一階段(A)—領域性關鍵詞彙集合之建立」所得之結果(領域性關鍵詞彙集合)為基礎，並依序透過「Part1—取得提問品質與回答品質得分(步驟(B1)至步驟(B4))」、「Part2—取得問答內文契合及語意關聯程度(步驟(B5)及步驟(B6))」與「Part3—取得整體問答契合度得分(步驟(B7))」等三大面向流程之運行，以 Tung 與 Lu (2010) 為鑒借，同時結合「領域性關鍵詞彙集合」之概念，建構自動化之「問答內容核心資訊萃取方法」，以萃取虛擬社群討論串(討論過程)中發問者相關之「關鍵提問詞彙」，以及回答者相關之「關鍵回答詞彙」，並藉此等關鍵詞彙於目標討論串中所得代表性(重要性)得分，分析得目標虛擬社群之提問與回答品質得分(Part1)，不僅止於此，本階段亦針對問答之內文及語意(Cilibrasi 與 Vitanyi (2007))進行雙向之分析，以取得問答內文契合及語意關聯程度(Part2)，最終，本階段藉「提問及回答品質得分」與「問答內文及語意關聯程度」之兩相結合，分析並計算得目標虛擬社群提問與回答之整體問答契合度得分(Part3)，以視本研究第二階段方法論「問答內容核心資訊之萃取與問答契合度分析」之最終目的，而本階段相關運作之方式及流程如圖 5 所示。

### 步驟(B1)—篩選候選關鍵提問與回答詞彙並計算前後文配對及標題加權整體分數

完成此任務之首要步驟，即藉中研院開發之 CKIP 中文斷詞系統，對目標虛擬社群討論串中之資料(QA<sub>h</sub>)進行斷詞(h=1 為提問內容；h>1 為其他回答)，以取得所有資料之斷詞詞彙集合 BW(QA<sub>h</sub>) (如公式(7)所示)，爾後，承接於上述斷詞之結果，透過有效詞(AWT)篩選之方式，進行有效詞之篩選，以過濾不具意義之詞彙，並於最終獲致目標虛擬社群討論串相關資料之有效詞彙集合 F\_BW(QA<sub>h</sub>) (如公式(8)所示)，以視資料預處理動作之完成。當中，本模組亦定義有效詞彙為候選關鍵提問或回答詞彙(依資料所屬之種類而定)，並套用於後續之方法論中，若該筆資料係屬「提問」，其所包含之有效詞彙即為「候選關鍵提問詞彙」，反之，若該資料為「回答」，其包含之有效詞則為「候選關鍵回答詞彙」。

$$BW(QA_h) = \{QAW_{h,1}, QAW_{h,2}, \dots, QAW_{h,c}\} \quad (7)$$

$$F\_BW(QA_h) = \{QAC_{h,1}, QAC_{h,2}, \dots, QAC_{h,d} \mid QAC_{h,d} \in AWT\} \quad (8)$$

待完成相關資料之預處理動作後，本步驟即計算各觸發詞於候選關鍵提問與回答詞彙中，前段語句 QAC<sub>h,d</sub><sup>P</sup> 與後段語句 QAC<sub>h,d</sub><sup>N</sup> 於提問及回答內分佈之比例，於此相加並取平均值，以分析得候選關鍵

提問與回答詞彙之前後文配對平均分數，當中，分析得之結果值越大即表示該候選關鍵提問或回答詞彙越趨近於內文之核心，此外，由於提問標題相對於內文而言，更能貫穿整體提問之核心與主軸，故本步驟亦分析候選關鍵提問詞彙與提問標題之對應關係，賦予提問標題更高之權重，於最終，對整體得分進行相對應之加權，以合理之提問標題與內文權重概念，分析得更具代表性與精確度候選關鍵提問詞彙之前後文配對及標題加權整體分數  $OScore(QAC_{h,d})$  (如公式(9)所示)，當中，由於大部分虛擬社群中之「回答」，皆無法針對回答訂定相關之標題，故於標題加權之分數中即不予以計算。

If  $h=1$

$$\text{Then } OScore(QAC_{h,d}) = \left[ \sum_{\text{alla}} \left( \frac{FQ(QAC_{h,d,a}^P, QAC_{h,d}^P)}{N(QAC_{h,\bullet}^P)} + \frac{FQ(QAC_{h,d,a}^P, QAC_h^P)}{N(QAC_{h,\bullet}^P)} \right) \right] \cdot \left[ \frac{1}{N(QAC_{h,\bullet}^P)} \frac{1}{N(QAC_{h,\bullet}^N)} \frac{1}{FQ(QCW_{h,d}, QCW_h)} \right] \quad (9)$$

$$\text{Else } OScore(QAC_{h,d}) = \left[ \sum_{\text{alll}} \left( \frac{FQ(QAC_{h,d,a}^P, QAC_{h,d}^P)}{N(QAC_{h,\bullet}^P)} + \frac{FQ(QAC_{h,d,a}^P, QAC_h^P)}{N(QAC_{h,\bullet}^P)} \right) \right] \cdot \left[ \frac{1}{N(QAC_{h,\bullet}^P)} \frac{1}{N(QAC_{h,\bullet}^N)} \right]$$

$$\left[ \sum_{\text{allm}} \left( \frac{FQ(QAC_{h,d,b}^N, QAC_{h,d}^N)}{N(QAC_{h,\bullet}^N)} + \frac{FQ(QAC_{h,d,b}^N, QAC_h^N)}{N(QAC_{h,\bullet}^N)} \right) \right]$$

### 步驟(B2)—計算候選關鍵提問與回答詞彙之詞類相關性平均分數

於此步驟中，本研究運用 **Tung 與 Lu (2010)** 分析文章詞性關聯規則之方法，於計算候選關鍵提問與回答詞彙之詞類相關性平均分數中，首先，承接於步驟(B1)中公式(7)經斷詞後所得之詞彙集合  $BW(QA_{\bullet})$ ，並依「中研院平衡語料庫詞類標記集」中對詞類定義之方式，統計全部訓練文件中各詞彙所屬之詞類所佔比率，並藉統計之數據建置相關之詞類距離採納表，以獲致整體之詞類比重分布狀況，其次，於選定詞類距離採納值  $ARV(POS_{\bullet})$  後，本步驟即藉此採納值建置各個候選關鍵提問與回答詞彙之「前推算集合  $Set(BQA_{h,\bullet,ARV(POS_{\bullet})})$ 」(如公式(10)所示)。

$$Set(BQA_{h,\bullet,ARV(POS_{\bullet})}) = \begin{bmatrix} BQA_{h,1,1} & BQA_{h,2,1} & \cdots & BQA_{h,d,1} \\ BQA_{h,1,2} & BQA_{h,2,2} & \cdots & BQA_{h,d,2} \\ \vdots & \vdots & & \vdots \\ BQA_{h,1,n} & BQA_{h,2,n} & \cdots & BQA_{h,d,n} \end{bmatrix} \quad (10)$$

其次，即針對集合內詞類分佈之情形進行分析，以統計出相關之詞類發生狀況  $FQ(Set(BQA_{h,d,n}))$  與  $RFQ(Set(BQA_{h,d}))$ ，於最終藉由候選關鍵提問與回答詞彙於整體訓練文件之分布狀況，以及前推算集合內詞彙之發生狀況，分析並計算得候選關鍵提問與回答詞彙之詞類相關性平均分數  $RScore(QAC_{h,d})$  (如公式(11)所示)，當中，若發生「候選關鍵提問或回答詞彙可向前推算之詞彙總數  $PWN(QAC_{h,d})$ 」小於「選定之詞類距離採納值  $ARV(POS_{\bullet})$ 」之情況時，將以  $PWN(QAC_{h,d})$  取代原本之詞類採納距離值。

$$\text{IF } PWN(QAC_{h,d}) < ARV(POS_{\bullet}) \text{ Then } RScore(QAC_{h,d}) = \frac{\sum_{\text{alln}} \frac{FQ(Set(BQA_{h,d,n}))}{FQ(BQA_{h,d,n}, BQA_h)}}{PWN(QAC_{h,d}) \cdot N(QAC_{h,\bullet}) \cdot RFQ(Set(BQA_{h,d})) + 1} \quad (11)$$

$$\text{Else } RScore(QAC_{h,d}) = \frac{\sum_{\text{alln}} \frac{FQ(Set(BQA_{h,d,n}))}{FQ(BQA_{h,d,n}, BQA_h)}}{ARV(POS_{\bullet}) \cdot N(QAC_{h,\bullet}) \cdot RFQ(Set(BQA_{h,d})) + 1}$$

### 步驟(B3)—取得候選關鍵提問與回答詞彙之代表性得分

為了於目標虛擬社群討論串中，篩選得實具代表性之內文核心詞彙，本步驟整合於步驟(B1)及步驟(B2)中所得候選關鍵提問與回答詞彙之「前後文配對及標題加權整體分數  $OScore(QAC_{h,d})$ 」，以及「詞類相關性平均分數  $RScore(QAC_{h,d})$ 」，同時藉由此兩面向所得分數之結合，計算得各候選關鍵提問與回答詞彙之代表性(重要性)得分  $IScore(QAC_{h,d})$  (如公式(12)所示)。

$$IScore(QAC_{h,d}) = OScore(QAC_{h,d}) \cdot OScoreP(QAC_{h,d}) + RScore(QAC_{h,d}) \cdot RScoreP(QAC_{h,d}) \quad (12)$$

$$\text{Where } OScoreP(QAC_{h,d}) + RScoreP(QAC_{h,d}) = 1$$

### 步驟(B4)—計算目標虛擬社群討論串中提問與回答品質之得分

一個實具品質的提問，將能催化使用者間互動之情況，進而提升使用者於虛擬社群中群聚有效知識之可能性，反之，若發問者發佈一則品質低下之提問，其能群聚無效知識之可能性亦大幅增加，是故，提問品質係影響虛擬社群知識分享意願關鍵因素之一，為了分析得實具準確性之發問者提問品質得分，本步驟即整合步驟(B3)中所得「候選關鍵提問詞彙代表性得分  $IScore(QAC_{1,d})$ 」，以及第一階段步驟(A3)所得「領域性關鍵詞彙集合  $WRC_k$ 」，透過「詞彙重要性分佈」與「領域性關鍵詞彙集合」兩相概念之結合，計算得實具準確性目標虛擬社群討論串中，發問者提問之領域性提問品質得分  $QScore$  (如公式(13)及公式(14)所示)。

$$IF D_T \in WRC_k \text{ and } QAC_{1,d} \in WRC_k \text{ Then } QV\_QWM_y = IScore(QAC_{1,d}) \quad (13)$$

$$QScore = \sum_{all y} QV\_QWM_y + \sum_{all y} \frac{QV\_QWM_y \cdot (FQ(QWM_y, QA_1) - 1)}{FQ(QWM_y, QA_1)} \quad (14)$$

另一方面，影響虛擬社群知識群聚之因子，不僅止於發問者提問之品質，回答者之「回答品質」亦實具相當之重要性，是故，本模組亦針對回答品質進行相關之解析，而於本步驟回答品質之分析中，主要針對回答者與發問者間，回答與提問之契合關係進行計算，本步驟於此即整合步驟(B1)中所得「目標虛擬社群候選關鍵提問詞彙集合  $F\_BW(QA_1)$ 」，以及步驟(B3)中所得「候選關鍵回答詞彙代表性得分  $IScore(QAC_{h,d})$ 」，透過關鍵提問與回答詞彙間符合性之兩相比對，以及詞彙重要性分佈概念之整合，以計算得回答內容之回答品質得分  $AScore_h$  (如公式(15)及公式(16)所示)。

$$IF QAC_{h,d} \in F\_BW(QA_1) \text{ Then } QV\_AWM_z = IScore(QAC_{h,d}), \text{ Where } h > 1 \quad (15)$$

$$AScore_h = \sum_{all z} QV\_AWM_z + \sum_{all z} \frac{QV\_AWM_z \cdot (FQ(AWM_z, QA_h) - 1)}{FQ(AWM_z, QA_h)}, \text{ Where } h > 1 \quad (16)$$

### 步驟(B5)—計算目標虛擬社群討論串中之問答內文契合度得分

不僅止於關鍵提問與回答詞彙間字面上之比對，本研究亦採納文章相似度判定概念，以建構兼具全面性與精準度之問答契合判定方法，因此，本步驟即藉由向量空間模型方法論之應用，以分析目標虛擬社群討論串中「提問 ( $QA_1$ )」與「回答 ( $QA_h$ )」間內文契合之狀況，以獲致兩資料之相似程度  $Sim(QA_1^o, QA_h^o)$ ，並視為目標虛擬社群討論串提問與回答之問答內文契合度得分 (如公式(17)所示)。

$$QA_1^o = [QAW_{1,1}^Q, QAW_{1,2}^Q, \dots, QAW_{1,p}^Q]^T \text{ and } QA_h^o = [QAW_{h,1}^A, QAW_{h,2}^A, \dots, QAW_{h,q}^A]^T$$

$$Sim(QA_1^o, QA_h^o) = \frac{QA_1^o \cdot QA_h^o}{\|QA_1^o\| \times \|QA_h^o\|}, \text{ Where } h > 1 \quad (17)$$

### 步驟(B6)—計算目標虛擬社群討論串中之問答語意關係距離

綜觀自然語言領域相關之處理技術，不僅止於文章相似度之解析，「詞彙語意解析」之相關技術亦實存重要性，是故，借鑑於 **Cilibrasi 與 Vitanyi (2007)** 中詞彙關係距離計算之方式，本步驟將原為計算兩詞彙間關係距離之方法，改良為計算目標虛擬社群討論串中提問與回答間整體關係距離之模式，藉由對整體候選關鍵提問及回答詞彙之語意解析，分析並計算得提問與回答之問答語意關係距離  $QASD_h$  (如公式(18)所示)。當中，問答語意關係距離之數值越低，即代表提問與回答間之語意關聯程度越大。

$$QASD_h = \frac{\sum_{all i} \frac{\text{Max}\{\log GSH(QAC_{1,i}^Q), \log GSH(QAC_{h,j}^A)\} - \log GSH(QAC_{1,i}^Q, QAC_{h,j}^A)}{\log GN - \text{Min}\{\log GSH(QAC_{1,i}^Q), \log GSH(QAC_{h,j}^A)\}}}{N(QAC_{1,\bullet}^Q) \cdot N(QAC_{h,\bullet}^A)}, \text{ Where } h > 1 \quad (18)$$

### 步驟(B7)—計算目標虛擬社群討論串中之整體問答契合度得分

為了於目標虛擬社群中提問與回答間分析得全面性、準確性與鑑別度兼具之整體問答契合度得分，於分析問答品質之範疇中，本研究結合步驟(B4)中所得「領域性提問品質得分  $QScore$ 」與「回答品質得分  $AScore_h$ 」，其次，於問答契合度解析之層面中，本研究整合步驟(B5)，以及步驟(B6)中所得「問答內文契合度得分  $Sim(QA_1^o, QA_h^o)$ 」及「問答語意關係距離  $QASD_h$ 」，於最終，本步驟藉問答品質與契合度兩大面向之結合，經正規化後分析並計算得目標虛擬社群討論串中提問與回答之整體問答契合度得分  $OMScore_h$  (如公式(19)所示)。



$$OMScore_h = \frac{QScore \cdot AScore_h + \frac{Sim(QA_l^{\omega}, QA_h^{\omega})}{QASD_h}}{\sum_{all h} QScore \cdot AScore_h + \frac{Sim(QA_l^{\omega}, QA_h^{\omega})}{QASD_h}}, \text{Where } h > 1 \quad (19)$$

### 3.2 虛擬社群成員參考度解析模組

本研究建構之「虛擬社群問答契合度解析模組」係分析虛擬社群討論串中提問與回答之契合狀況（詳細內容可參照章節 3.1），換言之，該模組即針對虛擬社群中所蘊含「知識內容」進行相關之解析，然而，虛擬社群不僅止存在「知識內容」，其亦包含虛擬社群成員間互動之過程，有別於一般知識平台，虛擬社群中之知識乃因包含了成員間互動之因子而實存不同之價值，每位成員皆有各自不全然相同之背景及專業知識，因而致使其所分享知識於各領域中可供參考之程度不同，亦即相同成員於不同領域中所分享知識之可參考性將有所不同，有鑑於此，為了避免知識擷取者因背景知識不足，而導致錯誤知識之擷取，以及提升知識擷取者擷取知識之效能與效率，本研究即建構「虛擬社群成員參考度解析模組」，針對虛擬社群成員之領域知識進行可參考度之解析，以提供知識擷取者參考。

本研究所提「虛擬社群成員參考度解析模組」係以 Zhang 等人 (2010) 所提之分群演算法作為資料分群之基礎，並依序透過步驟(C1)至步驟(C5)中「斷詞與篩選有效詞」、「篩選頻繁項目集」、「建置相似矩陣」、「判定分群類型」與「劃分領域資料之群集」流程之運行，藉此獲得「虛擬社群領域性討論串分佈群集」與「目標虛擬成員領域性歷史發文分佈群集」。其次，為了使知識擷取者以更為直觀之方式瀏覽，以及獲取虛擬社群成員知識專業性偏向及可參考程度之具體指標，本研究亦根據 Zhang 等人 (2010) 所提之群集命名法則，透過步驟(C6)進行相對應群集之命名，藉此分別取得實具代表性虛擬社群，以及目標虛擬成員領域性討論串及歷史發文分佈群集之群集名稱。最終，本研究即整合與改良「向量空間模型」及 Cilibrasi 與 Vitanyi (2007) 所提之「NGD 演算法」，同時輔以 Zhang 等人 (2010) 所提群集相似度計算之概念，透過步驟(C7)以分析虛擬社群領域性討論串，以及目標虛擬成員領域性歷史發文中各群集間之相似關係，以此獲致「目標虛擬成員之領域知識可參考程度」以視本模組之最終目的。當中，「目標虛擬成員領域性歷史發文」等相關資料係目標虛擬社群成員發文（包含分享、提問與回應），並經儲存後所保留之公開資訊，而「虛擬社群領域性討論串」係指虛擬社群中所包含各項類別討論區內之討論串，此外，一則完整之虛擬社群領域性討論串係由該討論串中之標題，以及所有提問及回答所組成。綜上所述，本研究所提之「虛擬社群成員參考度解析模組」其整體運作情況，如圖 6 所示。本研究將此模組分成七個步驟，以進行具階層性之敘述與方法設計。



圖 6、虛擬社群成員參考度解析模組之運作示意圖



### 步驟(C1)—斷詞與篩選有效詞

本模組資料前置處理作業，主要分為兩大面向，其一為「資料之斷詞」，其二為「資料中有效詞彙之篩選」，於本步驟中，即針對此兩面向進行相關之資料預處理動作。於「資料之斷詞」中，本步驟運用中研院開發之 CKIP 中文斷詞系統，對已彙整完成資料集合  $Set_{h,k}$  內所有資料進行斷詞，以取得各資料中詞彙之分佈狀況  $BW(AH_{h,k,\bullet})$  (如公式(20)所示)，而此部份所得之資訊將運用於步驟(C2)，以及步驟(C7)之公式(36)。

$$BW(AH_{h,k,\bullet}) = \{BW_{h,k,1}, BW_{h,k,2}, \dots, BW_{h,k,p}\} \quad (20)$$

於「資料中有效詞彙之篩選」中，承接於詞彙斷詞後之結果，本步驟根據「中研院平衡語料庫詞類標記集」中對詞類定義之方式，於彙整各詞彙之詞類發生狀況後，將詞類屬於有效詞 (AWT) 之詞彙進行保留，除有效詞中定義之詞類外，其他詞類之詞彙將不予保留，以此過濾不具意義之詞彙，進而獲致實具意義之領域資料有效詞集合  $F\_BW(AH_{h,k,\bullet})$  (如公式(21)所示)，而此部份所得之資訊將運用於步驟(C7)中之公式(37)。

$$F\_BW(AH_{h,k,\bullet}) = \{VW_{h,k,1}, VW_{h,k,2}, \dots, VW_{h,k,q} \mid VW_{h,k,q} \in AWT\} \quad (21)$$

### 步驟(C2)—篩選頻繁項目集

本步驟乃針對類別中之所有資料，進行「頻繁項目集」之篩選，而於此之頻繁項目集係指「出現頻率」大於或等於最小支持度 (MinSup) 之詞彙組合。於開始進行頻繁項目集之篩選前，必須先行根據領域中之所有資料，建構不考量順序詞彙組合之集合  $CSet(BW_{h,k,\bullet})$  (如公式(22)所示)。

$$CSet(BW_{h,k,\bullet}) = \{C\_BW_{h,k,1}, C\_BW_{h,k,2}, \dots, C\_BW_{h,k,e}\} \quad (22)$$

其次，則針對詞彙組合  $C\_BW_{h,k,e}$ ，出現於領域資料  $AH_{h,k,i}$  中之指標函數  $I(C\_BW_{h,k,e}, AH_{h,k,i})$  進行界定，若詞彙組合出現於其中即定義指標函數為 1；反之，若詞彙組合無出現則定義指標函數為 0 (如公式(23)所示)。

$$I(C\_BW_{h,k,e}, AH_{h,k,i}) = \begin{cases} 1, & \text{If } C\_BW_{h,k,e} \text{ exist in } AH_{h,k,i} \\ 0, & \text{Otherwise} \end{cases} \quad (23)$$

於最終即藉由「領域中所有資料總數 ( $N(AH_{h,k,\bullet})$ )」與「最小支持度 (MinSup)」之乘積，建置頻繁項目集之篩選門檻，若詞彙組合之指標函數總和，大於或等於篩選門檻，則該詞彙組合即隸屬於頻繁項目集  $FI(AH_{h,k,\bullet})$  (如公式(24)所示)。

$$FI(AH_{h,k,\bullet}) = \{C\_BW_{h,k,1}, C\_BW_{h,k,2}, \dots, C\_BW_{h,k,e} \mid \sum_{\text{all } i} I(C\_BW_{h,k,e}, AH_{h,k,i}) \geq N(AH_{h,k,\bullet}) \cdot \text{MinSup}\} \quad (24)$$

### 步驟(C3)—建置相似矩陣

本步驟乃運用領域中所有資料間之相似度建置「相似矩陣」，以視進行分群前相似度分析之前置動作。於建置相似矩陣之前，必須先行計算資料間之相似程度，首先，由於相似度之判定乃針對「兩資料間」之相似度進行解析，故於此必須先行將類別中之所有資料 ( $AH_{h,k,\bullet}$ )，分別定義為  $AH_{h,k,a}^A$  與  $AH_{h,k,b}^B$  等兩型態之獨立資料表示方式，以進行後續相似度之解析。其次，待完成資料之定義後，即運用 Zhang 等人 (2010) 所提結合「頻繁項目集」於相似矩陣中相似度判定法則之方法，以計算資料與資料間之相似程度  $\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)$  (如公式(25)所示)，當中，為使後續步驟能順利且有效率的運行，於此即限定兩資料間相似度之運算必須符合「 $a < b$  for all  $a, b$ 」之條件，以避免發生計算兩個完全相同資料之相似度 (如  $\text{Sim}(AH_{h,k,1}^A, AH_{h,k,1}^B)$ )，以及重複計算相似度 (如  $\text{Sim}(AH_{h,k,1}^A, AH_{h,k,2}^B)$  與  $\text{Sim}(AH_{h,k,2}^A, AH_{h,k,1}^B)$ ) 等情況。

$$\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) = N(\text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,a}^A) \cap \text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,b}^B)) \quad (25)$$

where  $a < b$  for all  $a, b$  and  $C\_BW_{h,k,\bullet} \in FI(AH_{h,k,\bullet})$

最終，即藉公式(24)中所計算得各資料間相似度數據之彙整，以建置相似矩陣  $SM[AH_{h,k,\bullet}]$  (如公式(25)所示)，當中，由於相似度之計算無須考量兩資料間前後順序之關係，故兩資料間相似度之計算  $\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)$  與  $\text{Sim}(AH_{h,k,b}^B, AH_{h,k,a}^A)$  將有相同之結果，此外，兩個完全相同之資料其相似程度亦

全然相同，雖仍可表示其相似度（相似度最大），但因後續所採用之分群法，必須運用相似矩陣中「最大相似度」之數據進行分群，以及針對完全相同資料進行分群無實質意義等因素，於公式(24)中針對兩個完全相同之資料即不予運算，而於相似矩陣中（公式(26)）則將此些未計算資料之相似度表示為「0」，以使後續之分群能順利且合理運行，另一方面，為提升分群之效率，針對重複相似度之計算（如  $\text{Sim}(\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B)$  與  $\text{Sim}(\text{AH}_{h,k,2}^A, \text{AH}_{h,k,1}^B)$ ）之處理方式，其作法亦同於上。

$$\text{SM}[\text{AH}_{h,k,\bullet}] = \begin{bmatrix} 0 & \text{Sim}(\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B) & \text{Sim}(\text{AH}_{h,k,1}^A, \text{AH}_{h,k,3}^B) & \cdots & \text{Sim}(\text{AH}_{h,k,1}^A, \text{AH}_{h,k,b}^B) \\ 0 & 0 & \text{Sim}(\text{AH}_{h,k,2}^A, \text{AH}_{h,k,3}^B) & \cdots & \text{Sim}(\text{AH}_{h,k,2}^A, \text{AH}_{h,k,b}^B) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (26)$$

#### 步驟(C4)—彙整數據並判定分群類型

於開始實行分群之前，必須先行完成相關資料對及數據彙整之前置動作，以判定分群之類型，因此，本步驟即藉由相似矩陣中所蘊含之相似度進行相關必要資訊之彙整。首先，針對相似矩陣中具備最大相似度之各個資料對，進行集合之彙整，以獲致相似矩陣中具備最大相似度資料對之集合  $\text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}])$ （如公式(27)所示），當中，資料對中必須存在至少一筆未經分群之資料，而於此之「資料對」係指相似度判定中之兩資料，如  $\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)$  則資料對即為  $(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)$ 。

$$\text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}]) = \left\{ \begin{array}{l} (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B), (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,3}^B), (\text{AH}_{h,k,2}^A, \text{AH}_{h,k,3}^B), \dots, (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \\ | \text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \in \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) \\ \text{and } \text{AH}_{h,k,a}^A \text{ or } \text{AH}_{h,k,b}^B \text{ has not been assigned to any cluster} \end{array} \right\} \quad (27)$$

其次，即針對相似矩陣中具備最小相似度，且該最小相似度符合數值不為 0 條件之數據進行蒐集，以取得相似矩陣中符合條件之最小相似度數值  $\text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}])$ （如公式(28)所示）。

$$\text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) = \text{Min}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) \quad (28)$$

$$\text{Where } \text{Min}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) \neq 0$$

最終，藉由相似矩陣中相似度之最大值與符合條件最小值之兩相比對，以取得後續實行分群時所採取之分群類型  $\text{CType}$ （如公式(29)所示），當中， $\text{CType}$  分為 1、2 與 3，若最大相似度同時不等於及大於符合條件之最小相似度，則  $\text{CType}=1$ ；最大相似度等於符合條件之最小相似度，則  $\text{CType}=2$ ；若最大相似度為 0，則  $\text{CType}=3$ 。

$$\text{CType} = \begin{cases} 1, \text{ If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) \neq \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ \quad \text{and } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) > \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ 2, \text{ If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) = \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ 3, \text{ If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) = 0 \end{cases} \quad (29)$$

#### 步驟(C5)—劃分領域資料之群集

本步驟係根據 Zhang 等人 (2010) 所提之分群方式，以建置相關分群法則，而選定分群法則前置條件係由步驟(C4)中所取得分群類型 ( $\text{CType}$ ) 而定，此外，當該批次 ( $\text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}])$ ) 之所有資料皆完成分群動作時（包含已判定但未被歸納於群集之資料），必須將此批次資料中所有資料對相似度設置為 0，並返回至步驟(C4)，再次取得新分群資料，持續相同動作直至所有資料皆完成分群為止。分群法則詳細說明如下：

- 當  $\text{CType}=1$  時，若資料對  $((\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B))$  中，兩筆資料皆不屬於已存在之群集 ( $\text{CR}'_{h,k,\bullet}$ )，則針對此些符合條件之資料對，各別增加新的群集  $\text{CR}_{h,k,c+n}$ （如公式(30)所示），當中， $n$  值為 1 代表第一個新群集， $n$  值為 2 代表第二個新群集...以此類推。

$$\text{CR}_{h,k,c+n} = \left\{ \begin{array}{l} (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B), (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,3}^B), (\text{AH}_{h,k,2}^A, \text{AH}_{h,k,3}^B), \dots, (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \\ | \text{CType} = 1 \text{ and } \text{AH}_{h,k,a}^A \notin \text{CR}'_{h,k,\bullet} \text{ and } \text{AH}_{h,k,b}^B \notin \text{CR}'_{h,k,\bullet} \end{array} \right\} \quad (30)$$

and  $\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) = 0 \forall a, b$

Where  $(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \in \text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}])$

- 當 CType=1 時，若於資料對  $((AH_{h,k,a}^A, AH_{h,k,b}^B))$  中，有其中一筆資料  $(AH_{h,k,a}^A$  或  $AH_{h,k,b}^B)$  屬於已存在之群集  $(CR'_{h,k,\bullet})$ ，則將此資料對與已存在之群集進行聯集，以取得合併後之群集  $CR_{h,k,c}$  (如公式(31)所示)，當中，若有任一筆資料屬於新群集  $(CR_{h,k,c+n})$ ，則此條件不成立 (亦即不將此資料對進行合併)。

$$CR_{h,k,c} = \left\{ \begin{array}{l} (AH_{h,k,1}^A, AH_{h,k,2}^B), (AH_{h,k,1}^A, AH_{h,k,3}^B), (AH_{h,k,2}^A, AH_{h,k,3}^B), \dots, (AH_{h,k,a}^A, AH_{h,k,b}^B) \\ | CType = 1 \text{ and } AH_{h,k,a}^A \in CR'_{h,k,\bullet} \text{ or } AH_{h,k,b}^B \in CR'_{h,k,\bullet} \\ \text{and } AH_{h,k,a}^A \notin CR_{h,k,c+n} \text{ and } AH_{h,k,b}^B \notin CR_{h,k,c+n} \end{array} \right\} \cup CR'_{h,k,c}$$

(31)

$$\text{and Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) = 0 \forall a, b$$

$$\text{Where } (AH_{h,k,a}^A, AH_{h,k,b}^B) \in \text{MaxSet}(\text{SM}[AH_{h,k,\bullet}])$$

- 當 CType=2 時，則針對此批資料中  $(\text{MaxSet}(\text{SM}[AH_{h,k,\bullet}]))$ ，未歸屬於任一群集之資料，全部進行合併，以形成一個新的群集  $CR_{h,k,c+n+1}$  (如公式(32)所示)。

$$CR_{h,k,c+n+1} = \left\{ \begin{array}{l} AH_{h,k,1}^A, AH_{h,k,2}^B, \dots, AH_{h,k,a}^A, AH_{h,k,b}^B \\ | CType = 2 \text{ and } AH_{h,k,a}^A \notin CR'_{h,k,\bullet} \text{ or } AH_{h,k,b}^B \notin CR'_{h,k,\bullet} \end{array} \right\}$$

(32)

$$\text{and Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) = 0 \forall a, b$$

$$\text{Where } (AH_{h,k,a}^A, AH_{h,k,b}^B) \in \text{MaxSet}(\text{SM}[AH_{h,k,\bullet}])$$

- 當 CType=3 時，則將尚未歸屬於任一群集之全部資料進行合併，並建構為一個新的群集  $CR_{h,k,c+n+2}$  如公式(33)所示)。

$$CR_{h,k,c+n+2} = \{AH_{h,k,1}, AH_{h,k,2}, \dots, AH_{h,k,i} \mid CType = 3 \text{ and } AH_{h,k,i} \notin CR'_{h,k,\bullet}\} \quad (33)$$

### 步驟(C6)—定義群集名稱

本步驟係針對於步驟(C5)中所得之各群集，進行相對應群集之命名，以使後續分析得之「目標成員領域知識可參考程度」能以更具體之方式呈現。對於群集之命名，於本步驟中，主要以更具體之方式表達群集，以及名稱不重複之原則進行命名，因此，在命名過程的首要條件中，即根據該群集中具備最長長度之頻繁項目，以及同時出現於群集內其他資料中次數最高頻繁項目之名稱，直接對群集進行命名，以此獲致群集之名稱  $\text{Topic}_{h,k,c}$  (如公式(34)所示)，當中，若群集中沒有符合條件之頻繁項目，則將該群集中之頻繁項目歸納於候選群集名稱集合  $(\text{CTSet\_Topic}_{h,k,c})$  中，並於後續公式(35)中進行更進一步之處理。

$$\text{If } N(CR_{h,k,c}) \neq 1 \text{ and } \text{FI}(AH_{h,k,\bullet}) \in \text{ML}(\text{FI}(AH_{h,k,\bullet}), CR_{h,k,c})$$

$$\text{and } \text{FI}(AH_{h,k,\bullet}) \in \text{MN}(\text{FI}(AH_{h,k,\bullet}), CR_{h,k,c})$$

$$\text{Then } \text{FI}(AH_{h,k,\bullet}) = \text{Topic}_{h,k,c} \quad (34)$$

$$\text{Else } \text{FI}(AH_{h,k,\bullet}) \in \text{CTSet\_Topic}_{h,k,c}$$

$$\text{Where } \text{FI}(AH_{h,k,\bullet}) \in CR_{h,k,c} \text{ and } \text{CTSet\_Topic}_{h,k,c} \notin \text{Topic}_{h,k,\bullet}$$

若群集中無符合首要條件之頻繁項目，以致無法直接進行群集之命名，則根據該群集中同時出現於群集內其他資料中次數最高頻繁項目之名稱，針對群集進行命名，而當中所訂定之名稱不得為已存在之群集名稱  $(\text{Topic}_{h,k,\bullet})$ ，以符合名稱不重複之原則 (如公式(35)所示)。

$$\text{If } \text{FI}(AH_{h,k,\bullet}) \in \text{MN}(\text{FI}(AH_{h,k,\bullet}), CR_{h,k,c}) \text{ and } \text{FI}(AH_{h,k,\bullet}) \notin \text{Topic}_{h,k,\bullet}$$

$$\text{Then } \text{FI}(AH_{h,k,\bullet}) = \text{Topic}_{h,k,c} \quad (35)$$

$$\text{Where } \text{FI}(AH_{h,k,\bullet}) \in \text{CTSet\_Topic}_{h,k,c}$$

### 步驟(C7)—取得目標虛擬成員領域知識可參考程度數據

本研究鑒借於「向量空間模型」，以及 Cilibrasi 與 Vitanyi (2007) 所提之「NGD 演算法」，以「內文」及「語意」兩面向相似度解析之方式為基礎，並同時輔以 Zhang 等人 (2010) 所提之群集相似度計算方法，於整合及改良後發展出群集間整合內文及語意相似度解析之方法，以用於分析虛擬社群成員領域知識可參考之程度。於開始進行相關之解析前，必須先行根據於步驟(C5)中所得之各群集，依對象「虛擬社群」與「目標虛擬社群成員」及其所包含之各群集  $(CR_{h,k,c})$ ，分別劃分並定義為「 $CR_{1,k,x}^X$ 」



與「 $CR_{2,k,y}^Y$ 」等兩型態之獨立群集表示方式，以進行後續群集相似度之運算。其次，於群集間內文相似度解析之部份，本步驟以「向量空間模型」計算群集間各資料之相似程度，並根據各資料所屬之群集進行相對應相似程度之加總，於完成加總後則依兩群集中所包含資料之總數進行數值平均之動作，以避免因群集中資料總數龐大之落差，而產生之不對稱相似度判定結果（如公式(36)所示），當中，於此之兩群集係指以「虛擬社群（ $h=1$ ）」及「目標虛擬社群成員（ $h=2$ ）」為對象所劃分領域資料之群集，此外，於此部份中各資料之集合向量係根據步驟(C1)中公式(20)所得之斷詞結果所建置。

$$\begin{aligned} AH_{1,k,a}^{A\omega} &= [BW_{h,k,a,1}^A, BW_{h,k,a,2}^A, \dots, BW_{h,k,a,v}^A]^T \\ AH_{2,k,b}^{B\omega} &= [BW_{h,k,b,1}^B, BW_{h,k,b,2}^B, \dots, BW_{h,k,b,w}^B]^T \\ \text{SimVS}(CR_{1,k,x}^X, CR_{2,k,y}^Y) &= \frac{\sum_{AH_{1,k,a}^A \in CR_{1,k,x}^X} \sum_{AH_{2,k,b}^B \in CR_{2,k,y}^Y} \frac{AH_{1,k,a}^{A\omega} \cdot AH_{2,k,b}^{B\omega}}{\|AH_{1,k,a}^{A\omega}\| \cdot \|AH_{2,k,b}^{B\omega}\|}}{N(AH_{1,k,\bullet}^A) + N(AH_{2,k,\bullet}^B)} \end{aligned} \quad (36)$$

Where  $AH_{1,k,\bullet}^A \in CR_{1,k,x}^X$  and  $AH_{2,k,\bullet}^B \in CR_{2,k,y}^Y$

再者，本步驟鑒借「NGD演算法」中計算兩詞彙語意關係距離之方法，於改良後同時輔以群集相似度計算之概念，將其擴展為得以分析兩群集間語意關係距離之方法，以用於解析兩群集間除「內文相似程度」外之「語意關係相似程度」（如公式(37)所示），當中，計算所得之數值越低，即代表兩群集間之語意關係距離越大，而於此部份中所指稱之「兩群集」，其定義亦同於公式(36)中所定義之對象，此外，由於此部份係解析兩群集間「語意」之關係，故對於資料中所存在無具代表性之詞彙而言（如虛詞），本研究即予以屏除，以避免過多雜訊影響分析之結果，並以此獲致實具代表性兩群集間之語意關係距離，而此部份中各資料實具代表性之詞录，即藉步驟(C1)中公式(21)有效詞篩選之結果所取得。

$$\begin{aligned} &\sum_{AH_{1,k,a}^A \in CR_{1,k,x}^X} \sum_{AH_{2,k,b}^B \in CR_{2,k,y}^Y} \\ &\quad \text{Max} \{ \log \text{GSH}(VW_{1,k,m}^A, AH_{1,k,a}^A), \log \text{GSH}(VW_{2,k,n}^B, AH_{2,k,b}^B) \} \\ &\quad - \log \text{GSH}((VW_{1,k,m}^A, AH_{1,k,a}^A), (VW_{2,k,n}^B, AH_{2,k,b}^B)) \\ \text{SimNGD}(CR_{1,k,x}^X, CR_{2,k,y}^Y) &= \frac{\sum_{\text{all } m,n} \log \text{GN} - \text{Min} \{ \log \text{GSH}(VW_{1,k,m}^A, AH_{1,k,a}^A), \log \text{GSH}(VW_{2,k,n}^B, AH_{2,k,b}^B) \}}{N(VW_{1,k,\bullet}^A) \cdot N(VW_{2,k,\bullet}^B)} \\ &\quad \frac{N(AH_{1,k,\bullet}^A) + N(AH_{2,k,\bullet}^B)}{N(AH_{1,k,\bullet}^A) + N(AH_{2,k,\bullet}^B)} \end{aligned} \quad (37)$$

Where  $VW_{1,k,m}^A \in AH_{1,k,a}^A$  and  $VW_{2,k,n}^B \in AH_{2,k,b}^B$  and  $AH_{1,k,\bullet}^A \in CR_{1,k,x}^X$  and  $AH_{2,k,\bullet}^B \in CR_{2,k,y}^Y$

最終，本步驟將公式(36)與公式(37)所得之結果進行兩相之整合，於正規化後分析並計算得目標虛擬成員之領域知識可參考程度  $MDR(D_{k,x}, (M_T, DK_{k,y}))$ （如公式(38)所示）。

$$MDR(D_{k,x}, (M_T, DK_{k,y})) = \frac{\frac{\text{SimVS}(CR_{1,k,x}^X, CR_{2,k,y}^Y)}{\text{SimNGD}(CR_{1,k,x}^X, CR_{2,k,y}^Y)}}{\sum_{\text{all } h,y} \frac{\text{SimVS}(CR_{1,k,x}^X, CR_{h,k,y}^Y)}{\text{SimNGD}(CR_{1,k,x}^X, CR_{h,k,y}^Y)}} \quad (38)$$

Where  $h > 1$

綜上所述，本研究提之「虛擬社群成員參考度解析模組」，主要分成三大面項之主軸進行虛擬社群成員參考度之解析，第一部份由步驟(C1)至步驟(C5)構成，其目的為劃分領域資料之群集，第二部份為步驟(C6)定義群集之名稱，而第三部份為步驟(C7)之取得目標虛擬成員領域知識可參考程度數據，本研究藉由對 Zhang 等人 (2010)、向量空間模型，以及 Cilibrasi 與 Vitanyi (2007) 等三項方法論之鑒借，經本研究之改良後，整合並發展出一套適用於虛擬社群之成員參考度解析模組，知識擷取者將可參考經本模組運作後所得「目標虛擬成員之領域知識可參考程度」，以避免因背景知識不足，而導致錯誤知識之擷取，並於最終提升其擷取知識之效能與效率。

#### 4. 系統應用流程

根據第三章發展之方法論，本研究乃開發一套提升成員知識分享意願之虛擬社群知識審核系統以確認模式可行性，當中，本研究將系統使用者分為一般使用者與系統管理者，並依權限而有不同執行權力，如圖7所示，此外，系統管理者乃針對「搜搜問問」進行訓練文件蒐集（如圖8所示），以作為判定後續解析之基礎。另一方面，一般使用者須蒐集未判定虛擬社群討論串資料，以透過本系統幫助其判定問答契合度及成員參考程度，於此乃以「搜搜問問」之討論串（如圖9所示），作為驗證資料樣本，以測試系統之可行性與績效。

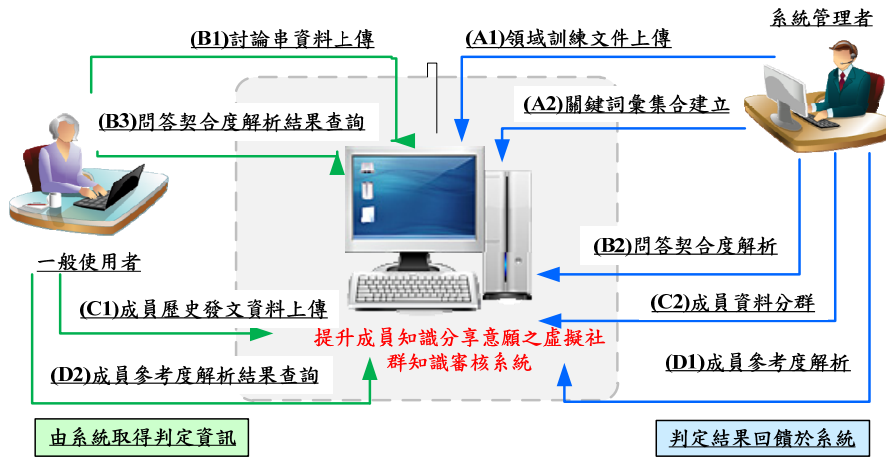


圖 7、提升成員知識分享意願之虛擬社群知識審核系統運作架構



圖 8、「搜搜問問」之精華知識文章



圖 9、「搜搜問問」之討論串資料

#### ➤ (A) 領域性關鍵詞彙集建立功能

##### (A1) 系統管理者上傳領域訓練文件

當管理者進入此功能輸入如標題「電腦...」及內容「有沒有...」，並選擇如類別「電腦/數碼」，上傳自定義路徑如「D:\data」領域文件原始檔案後（如圖 10 所示），系統將進行斷詞與有效詞篩選之資料預處理，以分別取得如「有沒有(ADV) 什麼(DET)...」及「方法；家；...」等結果，並將結果儲存與維護於資料庫中（如圖 11 所示）。



圖 10、領域訓練文件基本資料之輸入



圖 11、領域訓練文件資料之預處理與新增





平均分數與代表性得分，並解析得「0.73966」及「1.49634」等問答品質得分（如圖 18 所示），爾後，於統計詞彙發生頻率後，將計算得問答內文契合度得分「0.52」，接著，系統以排列組合方式劃分與計算各語意關係距離（如圖 19 所示），並根據各語意關係及有效詞彙組合總數「81」，計算得問答語意關係距離「0.20681」（如圖 20 所示），最終，整合問答品質、內文契合度及語意關係距離，將於正規化後計算得整體問答契合程度「0.42771」，亦同時儲存分析結果於資料庫中（如圖 21 所示）。

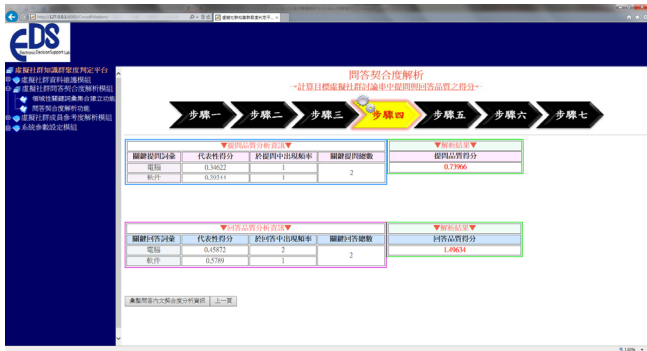


圖 18、提問與回答品質得分計算



圖 19、各詞彙組合語意關係距離計算



圖 20、問答語意關係距離計算

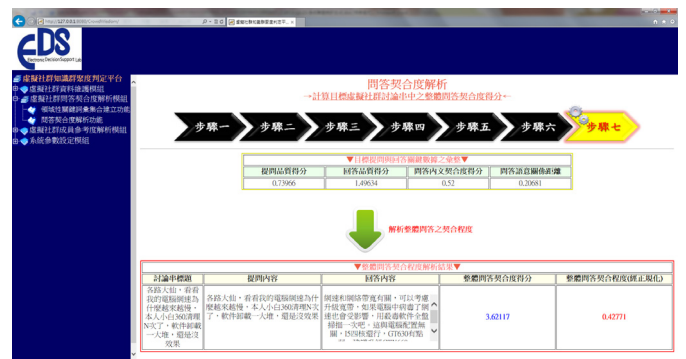


圖 21、整體問答契合度得分計算

### (B3)一般使用者查詢目標討論串問答契合度解析結果

系統管理者完成討論串問答契合度之解析後，一般使用者可透過「討論串資料查詢」功能，以查詢得欲了解目標討論串問答契合度解析之結果，如圖 22 與圖 23 所示。



圖 22、討論串資料之查詢

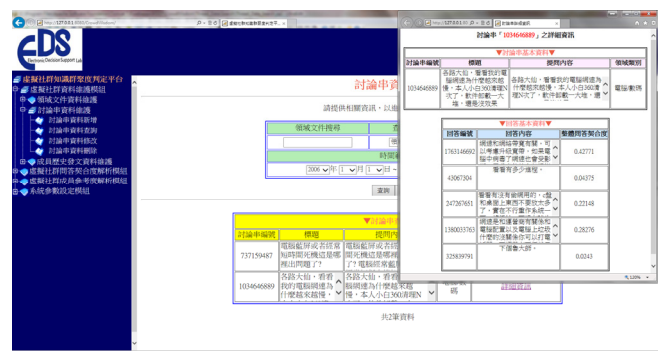


圖 23、檢視詳細討論串資訊

### ➤ (C)成員資料分群功能

#### (C1)一般使用者上傳未判定成員歷史發文資料

當使用者輸入成員「galant7072」及發文筆數「8」等資訊後，系統將創建相應欄位，使用者依序輸入標題「東京...」...等內容（如圖 24 所示），系統即執行資料預處理分別取得如「這(DET)...」及「航空；...」等結果，並儲存於資料庫中（如圖 25 所示）。



圖 24、目標成員各筆歷史發文輸入



圖 25、目標成員各筆歷史發文預處理與新增

(C2) 管理者執行成員資料分群核心步驟

管理者針對「galant7072」領域歷史發文執行資料預處理(如圖 26 所示),並根據最小支持度篩選得如「{淺草寺}」頻繁項目集(如圖 27 所示),接著,系統以此交集結果為基分析發文間相似程度,並於彙整相似度數據後,判定得目前資料所屬分群類型為「2」,接著,系統乃劃分領域資料群集,並將最大相似度數據取代為 0 以持續判定分群類型(如圖 28 所示),最後,待目領域資料皆完成分群時,系統即彙整並呈現分群結果,如「Cluster1」中乃包含編號「4」及「10」等歷史發文(如圖 29 所示)。



圖 26、成員領域歷史發文之資料預處理



圖 27、篩選頻繁項目集



圖 28、劃分領域資料之群集



圖 29、彙整最終分群結果

➤ (D) 成員參考度解析功能

(D1) 管理者執行成員參考度解析核心步驟

管理者透過系統針對「galant7072」統整領域分佈數據(如圖 30),接著,當選取「電腦/數碼」領域後,系統將彙整相應分群資訊,並根據頻繁項目集與群集間之關係,定義如群集編號「1」群集名稱為「淺草寺、餐廳」(如圖 31 所示),爾後,系統彙整領域訓練討論串並執行資料預處理,藉由向量空間模型計算得整體內文相似度「0.35933」,此外,系統以群集為單位劃分有效詞彙,並計算得整體語意關係距離「0.32373」(如圖 32 所示),最終,系統整合所有資訊後,將解析得目標社群成員「galant7072」領域知識可參考程度為「0.21268」(如圖 33 所示)。

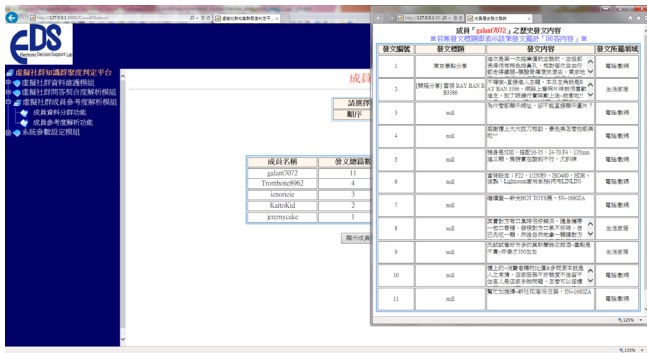


圖 30、查詢成員歷史發文資料



圖 31、彙整分群資訊並定義群集名稱



圖 32、計算討論串與群集語意關係距離



圖 33、計算成員領域知識可參考程度

(D2)一般使用者查詢目標成員領域參考度解析結果

系統管理者完成成員領域參考度解析後，一般使用者可查詢得目標成員領域知識可參考度解析之結果，如圖 34 與圖 35 所示。



圖 34、成員歷史發文資料之查詢



圖 35、檢視詳細成員歷史發文資訊

5. 系統驗證與評估

為驗證本系統於實務應用之績效，本研究以「搜搜問問」虛擬社群真實問答資料作為驗證資料樣本（如圖 36 所示），並以該社群之精華知識文章，作為訓練資料之樣本（如圖 37 所示），借鑑於同為分析中文虛擬社群資料 Liu 等人 (2013)，以及 Toba 等人 (2014) 所採之驗證方式，經由相關之改良後，以驗證本研究發展方法論與系統之績效。



圖 36、「搜搜問問」之問答資料



圖 37、「搜搜問問」之精華知識文章



## 5.1 驗證資料之蒐集與建置

於問答契合度之驗證中，本研究乃於「搜搜問問」中蒐集 680 筆具備相當品質，以及領域概念之問答內容，作為領域訓練資料（如表 2 所示），而評斷標準將以兩面向，包含人工審核（由兩位自願者審核），以及參考贊同與不贊同數比值（由其他社群成員點擊）等面向，進行 Cohen's Kappa 統計，將兩位審核者、以及其他社群成員視為兩個觀測者，以評估不同觀測者對同一事件（衡量回答品質）評斷之一致性，並將 Kappa 統計值維持於 0.97 間（參考 Toba 等人 (2014)），而本研究亦隨機挑選 20 筆領域問答內容（各類別 2 筆）作為問答契合度判定之測試資料（如表 3 所示）。

表 2、虛擬社群知識群聚度判定系統訓練資料表（部分資料）

領域類別	提問內容	最佳解答	贊同/不贊同數量	審核結果 A	審核結果 B
奧運體育	WWE 的 8B 去哪了	Batista (巴蒂斯塔) 已經在去年 5 月份就離開 WWE 了, 前...	13/0	高品質知識	非高品質知識
社會/人文	喝酒為什麼要碰杯	喝酒為什麼要碰杯? 目前有兩種說法。一種說法是職人...	1491/4	高品質知識	高品質知識
電腦/數碼	如何讓電話一直有電	智能手機確實好用, 但這電量問題實在是很麻煩。平時班...	82/51	非高品質知識	高品質知識

表 3、問答契合度績效判定測試資料彙整表（已轉換為繁體中文之部份資料）

領域類別	提問內容	最佳解答
電腦/數碼	怎麼才能學會查看電腦配置?	電腦的配置一般是指電腦的硬件配件的高檔程度...
生活家居	怎樣選擇實木家具	實木多用於用料較少的品種和局部, 而且貴重...
奧運體育	游泳有什麼好處, 有什麼壞處?	游泳的好處非常多, 游泳時身體直接浸泡在...

於成員參考度驗證中，本研究沿用蒐集得之 680 筆高品質領域訓練資料（如表 2 所示），另一方面，本研究乃於「搜搜問問」中，隨機針對各領域隨機挑選 5 位回答超過 200 則問題之社群成員（參考 Liu 等人 (2013)），共 10 個領域與 50 位社群成員，以視為系統績效驗證之受測者，並根據社群成員於領域類別中之「綜合聲望（由搜搜問問提供）」，由高至低，以分別排序各領域類別中社群成員可參考度之排名（即第一至第五名），亦即「綜合聲望之排名」為「實際成員領域可參考價值之排名」，並針對 50 位社群成員之領域歷史發文內容，依其所屬之領域各隨機蒐集並匯入 10 筆資料於系統中，共計 500 筆資料，以作為受測者之基礎測試資料（如表 4 所示）。

表 4、成員參考度績效判定受測者資料彙整表（電腦/數碼領域部分資料）

領域類別	回答者	綜合聲望	領域排名	回答編號	回答內容（隨機挑選）
電腦/數碼	王戰勝	378	1	1	路由器是組建局域網實現...
				2	有兩種辦法可以實現: 1. 改變...
				3	寬帶提示錯誤 769: 網卡禁用...
				4	呵呵, WIN7 的 EXCEL 是什...
				5	斷開你的路由器與上級路由器...

## 5.2 提升成員知識分享意願之虛擬社群知識審核模式之驗證

### ➤ 系統驗證方式說明

本研究由 680 份訓練資料中隨機挑選 200 份作為訓練資料，視為第一階段驗證，將其逐一匯入系統中，以建置「領域性關鍵詞彙集合」，並將之視為兩核心模組判定之基礎，並透過對「召回率」與「正確率」指標之應用，藉此階段性地評估本系統對於「問答契合度」與「成員參考度」解析之績效。

首先，於問答契合度解析績效之評估中，「問答契合度解析召回率」與「問答契合度解析正確率」乃為一相對比例值，分別為「實際高契合問答與推論高契合問答相符之個數」與「實際高契合問答個數」之比例，以及「實際高契合問答與推論高契合問答相符之個數」與「推論契合問答個數」之比例，期望藉由此項指標評估系統推論之問答契合度與實際問答契合狀況間之差異程度。

次之，於成員參考度解析績效之評估中，「成員領域可參考性排名召回率」與「成員參考度解析正確率」乃為一相對比例值，分別為「實際成員領域可參考性排名與推論成員領域可參考性排名相符之個數」與「實際成員領域可參考性總名次個數」之比例，以及「實際成員領域可參考性排名與推論成員領域可參考性排名相符之個數」與「推論成員領域可參考性總名次個數」之比例，期望藉由此項指標評估系統推論之成員領域可參考性排名，以及實際成員領域可參考性排名狀況間之差異程度。

最後，待完成上述第一階段驗證後，於第二階段中，本研究將剩餘之 480 份領域訓練文件分為 6

週期持續匯入系統中，每週分別匯入 80 份討論串資料，最後，於各週期中利用前述所選定之「20 筆領域問答內容」與「50 位社群成員」(測試資料)，分別重新進行「問答契合度解析」及「成員參考度解析」之推論，以分析系統(績效指標)於不同訓練問答討論串數量下之長期學習趨勢，當中，兩階段共 680 份具備相當品質之領域訓練資料，其 Kappa 統計值彙整如表 5。

表 5、領域訓練資料—評斷一致性程度分析彙整表

驗證資料評斷一致性		審核結果 A		總數
		高品質	非高品質	
審核結果 B	高品質	663	8	671
	非高品質	9	0	9
總數		672	8	680
<b>Kappa 值 ≈ 0.9744</b>				

## ➤ 系統驗證實作與績效結果分析

### (I) 第一階段驗證結果分析 (20 筆問答內容；50 位受測者)

在 200 份高品質問答內容為訓練資料基礎下，系統針對 20 筆測試資料進行判定，獲得「問答契合度解析」平均召回率為 30%、平均正確率為 30%；「成員參考度解析」平均召回率為 26%、平均正確率為 26%。當中，於「問答契合度」層面，實際高契合度問答數量共為 20 個，系統推論契合問答數量為 15 個，符合高契合度問答數量為 6 個；於「成員參考度」層面，實際排名與系統推論排名符合個數為 6 個。此階段詳細之推論結果如表 6 與表 7 所示。

表 6、問答契合度解析—第一階段驗證結果 (共 200 筆訓練資料)

領域類別	實際高契合度問答數量	契合度門檻值 0.6	契合度門檻值 0.8	召回率	正確率
		推論契合問答數量	符合高契合度問答數量		
電腦/數碼	2	2	2	100%	100%
生活家居	2	1	0	0%	0%
奧運體育	2	2	0	0%	0%
休閒/愛好	2	1	0	0%	0%
藝術/文學	2	1	0	0%	0%
社會/人文	2	1	0	0%	0%
教育/科學	2	2	2	100%	100%
健康/醫療	2	2	1	50%	50%
商業/理財	2	2	1	50%	50%
娛樂/明星	2	1	0	0%	0%
平均值				30%	30%

表 7、成員參考度解析—第一階段驗證結果 (共 200 筆訓練資料)

領域類別	實際總名次個數	推論總名次個數	實際與推論排名符合個數	召回率	正確率
電腦/數碼	5	5	2	40%	40%
生活家居	5	5	1	20%	20%
奧運體育	5	5	1	20%	20%
休閒/愛好	5	5	1	20%	20%
藝術/文學	5	5	1	20%	20%
社會/人文	5	5	1	20%	20%
教育/科學	5	5	1	20%	20%
健康/醫療	5	5	2	40%	40%
商業/理財	5	5	2	40%	40%
娛樂/明星	5	5	1	20%	20%
平均值				26%	26%

透過上述圖表統計之結果得知，系統於第一階段之虛擬社群知識群聚度判定之驗證結果中，「問答契合度」之召回率與正確率其約落於 0%、50%及 100%間，且整體平均值約為 30%；「成員參考度」之召回率與正確率其約落於 20%及 40%間，且整體平均值約為 26%。是故，根據本階段驗證結果中所獲致之召回率及正確率而言，虛擬社群知識群聚度判定之準確率與績效欠佳，尚無法準確判斷回答契合度與成員參考度。

## (II) 第二階段驗證結果分析 (20 筆問答內容；50 位受測者)

本研究將第二階段驗證分為六個週期，並於各週期中分別匯入 80 份領域訓練文件，以觀察隨訓練文件增加，各週期驗證指標變化之趨勢，於此，各週期驗證相關結果整理如表 8 與表 9。

表 8、問答契合度解析績效彙整

虛擬社群 問答契合度解析		各週期間答契合度解析驗證—訓練問答討論串匯入數量							平均
		第一階段	第二階段						
		第一週期 200 份	第二週期 280 份	第三週期 360 份	第四週期 440 份	第五週期 520 份	第六週期 600 份	第七週期 680 份	
召回率	平均值	30%	50%	60%	70%	75%	80%	80%	64%
	成長率	-	20%	10%	10%	5%	5%	0%	8%
正確率	平均值	30%	55%	65%	75%	80%	85%	85%	68%
	成長率	-	25%	10%	10%	5%	5%	0%	9%

表 9、成員參考度判定績效彙整

虛擬社群 成員參考度判定		各週期成員參考度判定驗證—訓練問答討論串匯入數量							平均
		第一階段	第二階段						
		第一週期 200 份	第二週期 280 份	第三週期 360 份	第四週期 440 份	第五週期 520 份	第六週期 600 份	第七週期 680 份	
召回率	平均值	26%	44%	58%	68%	76%	82%	82%	59%
	成長率	-	18%	14%	10%	8%	6%	0%	9%
正確率	平均值	26%	44%	58%	68%	76%	82%	82%	59%
	成長率	-	18%	14%	10%	8%	6%	0%	9%

綜合而言，以每週期增加 80 份訓練文件為單位，平均每週期「問答契合度解析」，以及「成員參考度解析」績效指標，其整體成長率分別為「召回率 8%；正確率 9%」與「召回率 9%；正確率 9%」，而針對最終第七週期（共匯入 680 份訓練資料）之驗證結果而言，績效將分別由第一週期之「召回率 30%；正確率 30%」與「召回率 26%；正確率 26%」提升至「召回率 80%；正確率 85%」及「召回率 82%；正確率 82%」，藉由對上述驗證結果之觀察，本研究所開發之兩核心模組，具備學習能力與相當程度之正確性。

## (III) 驗證結果整體量化分析

整體驗證結果整理如表 10 所示。由表 10 可得知，各項驗證指標之「收斂前每週期平均成長率」及「整體每週期平均成長率」皆為正數成長，且各指標皆於第六週期呈現收斂狀態。整體而言，由第一與第二階段中，可得知當系統進行「問答契合度解析」與「成員參考度解析」時，訓練資料與週期乃影響系統績效之主要原因，而系統績效將隨著不同週期及訓練量之增加，呈現持續成長狀況，並於最終達到穩定且良好之績效水平。故本系統皆確實能準確判定相關數據，並確實為使用者提供具精確性之參考依據。

表 10、各項驗證指標成長率之彙整表

驗證指標	整體平均值	收斂週期	收斂前每週期平均成長率	整體每週期平均成長率
問答契合度解析召回率	64%	第六週期	11.25%	8.7%
問答契合度解析正確率	68%	第六週期	12.50%	11.2%
成員參考度判定召回率	59%	第六週期	11.2%	8.7%
成員參考度判定正確率	59%	第六週期	11.2%	8.7%

## (IV) 驗證結果整體質化分析

本系統與搜搜問問計算成員參考度之方法與結果，存在相當程度之差異性，而為能針對兩者進行整體化之分析與比較，以評估本系統所分析得「成員領域可參考程度」，對使用者所提供之實際參考性與貢獻性。於此，本研究乃先行針對搜搜問問與本系統之判定結果進行彙整，如表 11 所示。針對表 11 進行質化層面之比較，首先，於判定時採用發文相差總量層面中，搜搜問問於計算過程中所須採用之巨量資料，將為系統帶來龐大負擔，而本研究僅須隨機蒐集少許目標回答者於虛擬社群之歷史發文，即可執行相關判定，而就判定得結果正確性而言，本系統亦存在相當程度之量化績效。其次，於判定後呈現之領域可參考程度結果層面中，搜搜問問將因其輸出領域參考度結果均偏低下之故，致使知識需求者難以直接透過對百分比數據之參考，以評估社群成員實際領域專業性程度，而本研究參考過去研究，並經相關改良後所設計之成員參考度判定方法論，具備更為嚴謹與合理理論架構，故透



過本系統分析得之領域可參考性程度，對知識需求者而言更具參考之價值。最後，於成員參考度判定結果應用性之層面中，本系統與搜搜問問所判定得之「領域可參考程度」，皆可廣泛適用於虛擬社群中之各領域類別，而本系統與搜搜問問相較之，本系統僅須透過少量樣本數之隨機蒐集，即可完成實具績效（標準差趨近於 0）之判定結果，而搜搜問問則須藉由目標成員大量樣本之全數蒐集，方可取得具備績效性之判定結果。

表 11、本系統與搜搜問問成員參考性判定結果之比較

領域類別	受測者	分析時採用歷史發文數量		採用發文 相差總量	領域可參考程度		領域參考度 相差機率
		本研究	搜搜問問		本研究	搜搜問問	
電腦/數碼	王戰勝	10	59665	59655	76%	31%	45%
生活家居	古順均	10	36781	36771	70%	24%	46%
奧運體育	Stinger	10	15951	15941	67%	22%	45%
休閒/愛好	陳清華	10	19943	19933	65%	23%	42%
藝術/文學	魏婕馨	10	30242	30232	86%	37%	49%
社會/人文	蘆天鵬	10	37106	37096	78%	33%	45%
教育/科學	章炳華	10	14671	14661	75%	30%	45%
健康/醫療	文馳流星	10	33232	33222	71%	26%	45%
商業/理財	杜江	10	27246	27236	61%	18%	43%
娛樂/明星	喵喵同學	10	16944	26934	61%	28%	33%
平均		10	29178	30168	71%	27%	43%
樣本標準差					7.94%	5.71%	

## 6. 結論

虛擬社群成員透過開放式之網路社群以發佈與回覆等行為，進行主題式之討論藉以群聚相關之知識，進而提升社群之整體發展與造就群眾智慧之累積。然而，群聚大量知識固然容易，但因資訊共享之便利性，促使大量知識相繼湧入虛擬社群，僅透過既有社群成員間自發性評論及評分機制之運用，對低品質群眾智慧之散播難以進行實質之管控，如此之問題就知識需求者而言，將因蒐集到低品質資訊而產生額外評估知識之成本，就知識貢獻者言之，將降低社群成員持續分享知識之意願，而綜觀於對整體虛擬社群之影響，將導致社群中群眾智慧整體品質之下降，進而影響虛擬社群之發展。為改善既有虛擬社群運作機制面臨之問題，藉由對近期相關文獻之借鑑與改良，本研究發展一套「提升成員知識分享意願之虛擬社群知識審核模式」，當中，核心模組主要分為「虛擬社群問答契合度解析」與「虛擬社群問答契合度解析」模組，於前者中，本研究以群眾智慧影響性為切入點，發展一套整合語料庫與上下文之自動化解析，且通用於虛擬社群之方法論；於後者中，本研究考量個體成員知識分享意圖之觀點，發展一套同時具備詞頻與語意解析，以及考量成員間與群眾智慧間觀點之方法論，最終，整合兩核心模組，針對虛擬社群之「知識內文」與「使用者行為」等兩面向，進行知識群聚度之判定，提供相關數據予使用者，減少其篩選知識之成本，且同時提升獲取所需且正確知識之機率。另一方面，本研究除發展方法論外，亦根據此方法論建構一套以網際網路為基之「虛擬社群知識群聚度判定系統」以進行案例驗證，從而確認方法論與技術之可行性。

## 參考文獻

1. Agarwal, N., Galan, M., Liu, H. and Subramanya, S., 2010, "WisColl: Collective wisdom based blog clustering," *Information Sciences*, Vol. 180, pp. 39-61.
2. Alali, H. and Salim, J., 2013, "Virtual Communities of Practice Success Model to Support Knowledge Sharing behaviour in Healthcare Sector," *Procedia Technology*, Vol. 11, pp. 176-183.
3. Ba, S. and Wang, L., 2013, "Digital health communities: The effect of their motivation mechanisms," *Decision Support Systems*, Vol. 55, No. 4, pp. 941-947.
4. Blooma, M. J., Kurian, J. C., Chua, A. Y. K., Goh, D. H. L. and Lien, N. H., 2013, "Social question answering: Analyzing knowledge, cognitive processes and social dimensions of micro-collaborations," *Computers & Education*, Vol. 69, pp. 109-120.
5. Cantador, I. and Castells, P., 2011, "Extracting multilayered communities of interest from semantic user profiles: Application to group modeling and hybrid recommendations," *Computers in Human Behavior*, Vol. 27, No. 4, pp. 1321-1336.
6. Cao, Y., Yang, W.-Y., Lin, C.-Y. and Yu, Y., 2011, "A structural support vector method for extracting contexts and answers of questions from online forums," *Information Processing & Management*, Vol. 47, No. 6, pp. 886-898.

7. Caverlee, J., Liu, L. and Webb, S., 2010, "The SocialTrust framework for trusted social information management: Architecture and algorithms," *Information Sciences*, Vol. 180, No. 1, pp. 95-112.
8. Chen, C.-J. and Hung, S.-W., 2010, "To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities," *Information & Management*, Vol. 47, No. 4, pp. 226-236.
9. Chen, P.-I. and Lin, S.-J., 2011, "Word AdHoc network: Using Google core distance to extract the most relevant information," *Knowledge-Based Systems*, Vol. 24, No. 3, pp. 393-405.
10. Chen, Y.-J. and Chen, Y.-M., 2012, "Knowledge evolution course discovery in a professional virtual community," *Knowledge-Based Systems*, Vol. 33, pp. 1-28.
11. Chen, Y.-J., Chu, H.-C., Chen, Y.-M. and Chao, C.-Y., 2013, "Adapting domain ontology for personalized knowledge search and recommendation," *Information & Management*, Vol. 50, No. 6, pp. 285-303.
12. Cilibrasi, R. L. and Vitanyi, P. M. B., 2007, "The Google similarity distance," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 3, pp. 370-383.
13. Cress, U., Held, C. and Kimmerle, J., 2013, "The collective knowledge of social tags: Direct and indirect influences on navigation, learning, and information processing," *Computers & Education*, Vol. 60, pp. 59-73.
14. Dascalu, M. I., Bodea, C. N., Lytras, M., Pablos, P. O. D. and Burlacu, A., 2014, "Improving e-learning communities through optimal composition of multidisciplinary learning groups," *Computers in Human Behavior*, Vol. 30, pp. 362-371.
15. Daud, A., Li, J., Zhou, L. and Muhammad, F., 2010, "Temporal expert finding through generalized time topic modeling," *Knowledge-Based Systems*, Vol. 23, No. 6, pp. 615-625.
16. Fang, Y.-H. and Chiu, C.-M., 2010, "In justice we trust: Exploring knowledge-sharing continuance intentions in virtual communities of practice," *Computers in Human Behavior*, Vol. 26, No. 2, pp. 235-246.
17. Ferrández, A., 2011, "Lexical and syntactic knowledge for information retrieval," *Information Processing & Management*, Vol. 47, No. 5, pp. 692-705.
18. Galitsky, B. A., Rosa, J. L. and Dobrocsi, G., 2012, "Inferring the semantic properties of sentences by mining syntactic parse trees," *Data & Knowledge Engineering*, Vol. 81-82, pp. 21-45.
19. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H., 2009, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter*, Vol. 11, No. 1, pp. 10-18.
20. Han, L. and Chen, G., 2009, "A fuzzy clustering method of construction of ontology-based user profiles," *Advances in Engineering Software*, Vol. 40, No. 7, pp. 535-540.
21. Hong, H.-Y. and Scardamalia, M., 2014, "Community knowledge assessment in a knowledge building environment," *Computers & Education*, Vol. 71, pp. 279-288.
22. Huang, C.-L., Yeh, P.-H., Lin, C.-W. and Wu, D.-C., 2014, "Utilizing user tag-based interests in recommender systems for social resource sharing websites," *Knowledge-Based Systems*, Vol. 56, pp. 86-96.
23. Huang, S.-L., Lin, S.-C. and Chan, Y.-C., 2012, "Investigating effectiveness and user acceptance of semantic social tagging for knowledge sharing," *Information Processing & Management*, Vol. 48, No. 4, pp. 599-617.
24. Huh, J., Yetisgen-Yildiz, M. and Pratt, W., 2013, "Text classification for assisting moderators in online health communities," *Journal of Biomedical Informatics*, Vol. 46, No. 6, pp. 998-1005.
25. Hung, S.-W. and Cheng, M.-J., 2013, "Are you ready for knowledge sharing? An empirical study of virtual communities," *Computers & Education*, Vol. 62, pp. 8-17.
26. Iwashita, M., Shimogawa, S. and Nishimatsu, K., 2011, "Semantic analysis and classification method for customer enquiries in telecommunication services," *Engineering Applications of Artificial Intelligence*, Vol. 24, No. 8, pp. 1521-1531.
27. Jin, X.-L., Zhou, Zhongyun., Lee, M. K. O. and Cheung, C. M. K., 2013, "Why users keep answering questions in online question answering communities: A theoretical and empirical investigation," *International Journal of Information Management*, Vol. 33, No. 1, pp. 93-104.
28. Karamolegkos, P. N., Patrikakis, C. Z., Doulamis, N. D., Vlacheas, P. T. and Nikolakopoulos, I. G., 2009, "An evaluation study of clustering algorithms in the scope of user communities assessment," *Computers & Mathematics with Applications*, Vol. 58, No. 8, pp. 1498-1519.
29. Kim, H.-W, Zheng, J. R. and Gupta, S., 2011, "Examining knowledge contribution from the perspective of an online identity in blogging communities," *Computers in Human Behavior*, Vol. 27, No. 5, pp. 1760-1770.
30. Kim, J., Song, J. and Jones, D. R., 2011, "The cognitive selection framework for knowledge acquisition strategies in virtual communities," *International Journal of Information Management*, Vol. 31, No. 2, pp. 111-120.
31. Kim, Y. A. and Ahmad, M. A., 2013, "Trust, distrust and lack of confidence of users in online social media-sharing communities," *Knowledge-Based Systems*, Vol. 37, pp. 438-450.
32. Kim, Y. A. and Phalak, R., 2012, "A trust prediction framework in rating-based experience sharing social networks without a Web of Trust," *Information Sciences*, Vol. 191, pp. 128-145.

33. Ko, J., Si, L. and Nyberg, E., 2010, "Combining evidence with a probabilistic framework for answer ranking and answer merging in question answering," *Information Processing & Management*, Vol. 46, No. 5, pp. 541-554.
34. Korovaiko, N. and Thomo, A., 2012, "Predicting Trust from User Ratings," *Procedia Computer Science*, Vol. 10, pp. 263-271.
35. Lai, C.-H., Liu, D.-R. and Lin, C.-S., 2013, "Novel personal and group-based trust models in collaborative filtering for document recommendation," *Information Sciences*, Vol. 239, pp. 31-49.
36. Li, J., Li, Q., Liu, C., Khan, S. U. and Ghani, N., 2014, "Community-based collaborative information system for emergency management," *Computers & Operations Research*, Vol. 42, pp. 116-124.
37. Li, Y., Ma, S., Zhang, Y., Huang, R. and Kinshuk., 2013, "An improved mix framework for opinion leader identification in online learning communities," *Knowledge-Based Systems*, Vol. 43, pp. 43-51.
38. Li, Y.-M., Liao, T.-F. and Lai, C.-Y., 2012, "A social recommender mechanism for improving knowledge sharing in online forums," *Information Processing & Management*, Vol. 48, No. 5, pp. 978-994.
39. Li, Z., Xiong, Z., Zhang, Y., Liu, C. and Li, K., 2011, "Fast text categorization using concise semantic analysis," *Pattern Recognition Letters*, Vol. 32, No. 3, pp. 441-448.
40. Liu, D.-R., Chen, Y.-H. and Huang, C.-K., 2014, "QA document recommendations for communities of question-answering websites," *Knowledge-Based Systems*, Vol. 57, pp. 146-160.
41. Liu, D.-R., Chen, Y.-H., Kao, W.-C. and Wang, H.-W., 2013, "Integrating expert profile, reputation and link analysis for expert finding in question-answering websites," *Information Processing & Management*, Vol. 49, No. 1, pp. 312-329.
42. Liu, W., Yan, H. and Xiao, J., 2011, "Automatically extracting user reviews from forum sites," *Computers & Mathematics with Applications*, Vol. 62, No. 7, pp. 2779-2792.
43. Liu, X., Jiang, T. and Ma, F., 2013, "Collective dynamics in knowledge networks: Emerging trends analysis," *Journal of Informetrics*, Vol. 7, No. 2, pp. 425-438.
44. Lykourantzou, I., Papadaki, K., Vergados, D. J., Polemi, D. and Loumos, V., 2010, "CorpWiki: A self-regulating wiki to promote corporate collective intelligence through expert peer matching," *Information Sciences*, Vol. 180, No. 1, pp. 18-38.
45. Maleszka, M., Mianowska, B. and Nguyen, N. T., 2013, "A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles," *Knowledge-Based Systems*, Vol. 47, pp. 1-13.
46. Meo, P. D., Nocera, A., Terracina, G. and Ursino, D., 2011, "Recommendation of similar users, resources and social networks in a Social Internetworking Scenario," *Information Sciences*, Vol. 181, No. 7, pp. 1285-1305.
47. Moreo, A., Navarro, M., Castro, J. L. and Zurita, J. M., 2012, "A high-performance FAQ retrieval method using minimal differentiator expressions," *Knowledge-Based Systems*, Vol. 36, pp. 9-20.
48. Moskaliuk, J., Kimmerle, J. and Cress, U., 2012, "Collaborative knowledge building with wikis: The impact of redundancy and polarity," *Computers & Education*, Vol. 58, No. 4, pp. 1049-1057.
49. Ni, X., Lu, Y., Quan, X., Wenying, L. and Hua, B., 2012, "User interest modeling and its application for question recommendation in user-interactive question answering systems," *Information Processing & Management*, Vol. 48, No. 2, pp. 218-233.
50. Oliva, J., Serrano, J. I., Castillo, M. D. D. and Iglesias, A., 2011, "SyMSS: A syntax-based measure for short-text semantic similarity," *Data & Knowledge Engineering*, Vol. 70, No. 4, pp. 390-405.
51. Sato, A. and Costa-i-Font, J., 2013, "Social networking for medical information: A digital divide or a trust inquiry?" *Health Policy and Technology*, Vol. 2, No. 3, pp. 139-150.
52. Schall, D., 2012, "Expertise ranking using activity and contextual link measures," *Data & Knowledge Engineering*, Vol. 71, No. 1, pp. 92-113.
53. Sorg, P. and Cimiano, P., 2012, "Exploiting Wikipedia for cross-lingual and multilingual information retrieval," *Data & Knowledge Engineering*, Vol. 74, pp. 26-45.
54. Stieglitz, S. and Kaufhold, C., 2011, "Automatic full text analysis in public social media - Adoption of a software prototype to investigate political communication," *Procedia Computer Science*, Vol. 5, pp. 776-781.
55. Teplovs, C., 2008, "The Knowledge Space Visualizer: A tool for visualizing online discourse," *In Proceedings of the International Conference of the Learning Sciences*.
56. Tobarra, L., Gómez, A. R., Ros, S., Hernández, R. and Caminero, A. C., 2014, "Analyzing the students' behavior and relevant topics in virtual learning communities," *Computers in Human Behavior*, Vol. 31, pp. 659-669.
57. Tsai, H.-T. and Pai, P., 2013, "Explaining members' proactive participation in virtual communities," *International Journal of Human-Computer Studies*, Vol. 71, No. 4, pp. 475-491.
58. Tseng, F.-C. and Kuo, F.-Y., 2014, "A study of social participation and knowledge sharing in the teachers' online professional community of practice," *Computers & Education*, Vol. 72, pp. 37-47.



59. Tung, C.-M. and Lu, W.-H., 2012, "Predict depression tendency of web posts using negative emotion evaluation model," In *ACM SIGKDD Workshop on Health Informatics*.
60. Vertommen, J., Janssens, F., Moor, B. D. and Duflou, J. R., 2008, "Multiple-vector user profiles in support of knowledge sharing," *Information Sciences*, Vol. 178, No. 17, pp. 3333-3346.
61. Wang, B.-X., Liu, B.-Q., Sun, C.-J., Wang, X.-L and Sun, L., 2013, "Thread segmentation based answer detection in Chinese online forums," *Acta Automatica Sinica*, Vol. 39, No. 1, pp. 11-20.
62. Wang, G. A., Jiao, J., Abrahams A. S., Fan W. and Zhang, Z., 2013, "ExpertRank: A topic-aware expert finding algorithm for online knowledge communities," *Decision Support Systems*, Vol. 54, No. 3, pp. 1442-1451.
63. Wang, K.-Y., Ting, I.-H. and Wu, H.-J., 2013, "Discovering interest groups for marketing in virtual communities: An integrated approach," *Journal of Business Research*, Vol. 66, No. 9, pp. 1360-1366.
64. Wenyin, L., Quan, X., Feng, M. and Qiu, B., 2010, "A short text modeling method combining semantic and statistical information," *Information Sciences*, Vol. 180, No. 20, pp. 4031-4041.
65. Yan, J., Bracewell, D. B., Ren, F. and Kuroiwa, S., 2009, "Integration of multiple classifiers for Chinese semantic dependency analysis," *Electronic Notes in Theoretical Computer Science*, Vol. 225, No. 2, pp. 457-468.
66. Yan, Y., Davison, R. M. and Mo, C., 2013, "Employee creativity formation: The roles of knowledge seeking, knowledge contributing and flow experience in Web 2.0 virtual communities," *Computers in Human Behavior*, Vol. 29, No. 5, pp. 1923-1932.
67. Zhao, K., Stylianou, A. C. and Zheng, Y., 2013, "Predicting users' continuance intention in virtual communities: The dual intention-formation processes," *Decision Support Systems*, Vol. 55, No. 4, pp. 903-910.
68. Zhao, Z., Feng, S., Wang, Q., Huang, J. Z., Williams, G. J. and Fan, J., 2012, "Topic oriented community detection through social objects and link analysis in social networks," *Knowledge-Based Systems*, Vol. 26, pp. 164-173.
69. Zheng, H.-T., Kang, B.-Y. and Kim, H.-G., 2009, "Exploiting noun phrases and semantic relationships for text document clustering," *Information Sciences*, Vol. 179, No. 13, pp. 2249-2262.
70. Zheng, Y., Zhao, K. and Stylianou, A., 2013, "The impacts of information quality and system quality on users' continuance intention in information-exchange virtual communities: An empirical investigation," *Decision Support Systems*, Vol. 56, pp. 513-524.
71. Zolfaghar, K. and Aghaie, A., 2012, "A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data," *Knowledge-Based Systems*, Vol. 26, pp. 93-102.
72. 中研院 CKIP 中文斷詞系統，<http://ckipsvr.iis.sinica.edu.tw/>。

# 出席國際學術會議心得報告

105 年 07 月 25 日

計畫編號	MOST 104-2221-E-343-001
計畫名稱	提升成員知識分享意願之虛擬社群知識審核模式(I)
出國人員姓名 服務機關及職稱	楊士霆 南華大學資訊管理學系 助理教授
會議時間地點	July 19-21, 2016 Hokkaido, Japan
會議名稱	2016 Global Conference on Engineering and Applied Science (2016 GCEAS)
發表論文題目	A Member's Reference Degree Analysis Approach for Virtual Community

## 一、參加會議經過與心得

此次 2016 Global Conference on Engineering and Applied Science (2016 GCEAS)研討會乃安排於日本 (Japan) 之北海道 (Hokkaido) 中札幌 (Sapporo) 舉辦，配合研討會主辦單位之行程規劃與可行機位安排，個人於 07/17 上午八點，即出發前往桃園國際機場，進行登記作業且於上午 10:00 由桃園國際機場起飛，並於下午 15:00 抵達新千歲 New-Chitose-Airport Terminal 國際機場，辦理入境手續後即搭乘「JR (Japan Railway)」至札幌，並於當日下午 16:30 左右抵達住宿飯店「札幌三井花園飯店」(Mitsui Garden Hotel Sapporo) (此飯店離研討會舉辦飯店 Premier Hotel Tsubaki Sapporo 搭【地下鐵】東豐線，共需 30 分鐘左右交通時間)，台北與北海道時差為 1 小時；此程搭機、轉乘地鐵時間總計 8 小時左右，故辦理 Check in 手續並稍做休息後，於本日至 07/18、19 日上、下午即參觀札幌二條市場、白色戀人館、大通公園、狸小路、札幌巨蛋、小樽等周遭景點，以瞭解日本北海道之交通、飲食習慣、友善的風俗民情及北海道之建築、古蹟等風貌。

由於個人住宿於「札幌三井花園飯店」(Mitsui Garden Hotel Sapporo)，離研討會會場 Premier Hotel Tsubaki Sapporo 需搭【地下鐵】東豐線，福住站，共需 30 分鐘交通時間，個人於 07/20 上午 09:30 隨即前往報到，完成報到手續 (如圖 1 及圖 2 所示)，並與國外與會學者、專家進行互動。

此次研討會之規模可算是中型規模，研討會總計發表篇數約為 400 篇左右，議程數為 55 個左右，當中 Oral 議程約 40 個場次、Poster 約 15 個場次左右。大會正式行程日期為 7/19 至 7/21 三日，7/19 乃為提前報到日期，正式發表日期亦為 7/20 至 7/21 兩日。本次研討會內容乃安排與此次會議主題相關之工程與應用科學 (Engineering and Applied Science) 專題演講與論文發表，再依不同論文主題每天分至 5 個時段 6 個左右平行 Session 進行發表。個人的論文被

安排於發表日 7/21 的上午 (11:00~12:00) 場次 (編號 Poster Sessions 6) 「Computer and Information Sciences (2)」發表，由於此研討會主題乃著重於電腦與資訊科技，個人研究 (資料探勘、知識管理、系統開發) 與其他學者甚為相似，故於發表後其他學者亦表示對此研究的高度興趣，詢問本研究之網頁 (知識文件) 探勘技術、問答解析 (詞彙解析) 分析與應用、系統智慧推論技術與其他研究之差異，個人並作完整回答，互動甚佳。此外個人亦參加多場與研究興趣較相關之發表場次，並對於其他學者發表內容提出詢問，對於知識管理、資訊科技管理等課題觸發新的研究靈感 (如圖 3 至圖 8 所示)。

除會議發表時間外，在其他交流活動時，個人與國際/國內學者亦有良好交流，於此次研討會認識多位先進，藉由討論瞭解許多國際/國內工業工程、資訊管理學者之研究方向，並規劃未來合作之可能作法，收穫極大，此對於個人學術經歷尚屬資淺而言，乃一大助益。



圖 1、抵達 2016 GCEAS 會場並註冊(1)



圖 2、抵達 2016 GCEAS 會場並註冊(2)

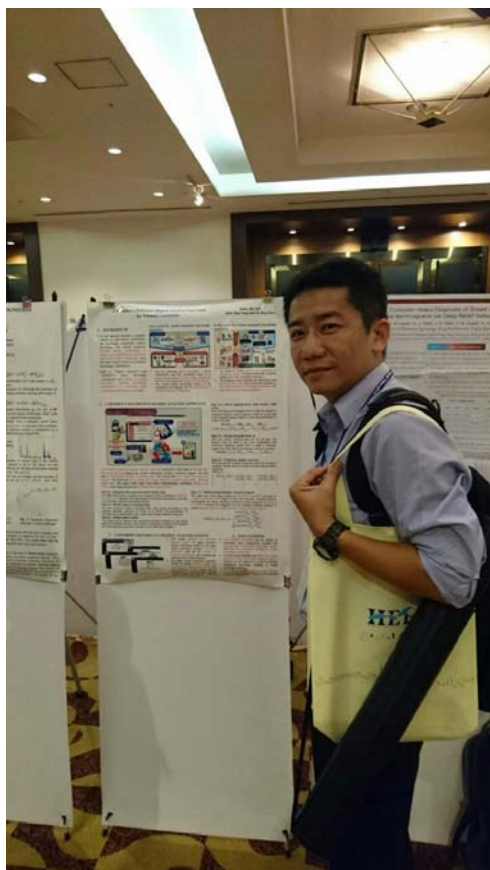


圖 3、論文發表與研討(1)



圖 4、論文發表與研討(2)



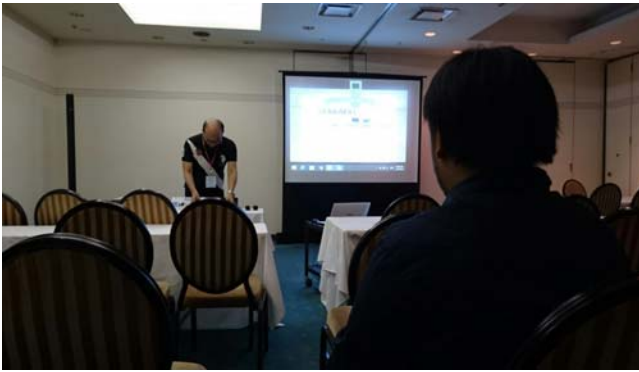


圖 5、論文發表與研討(3)



圖 6、論文發表與研討(4)



圖 7、論文發表與研討(5)



圖 8、論文發表與研討(6)

待研討會圓滿結束後，個人於隔日(7/22)即搭車前往新千歲 New-Chitose-Airport Terminal 國際機場，並搭機回桃園國際機場，結束此次 2016 GCEAS 學術研討活動。

### 三、建議

此次會議中的各項活動安排都可發現主辦單位頗為用心，對於遠道造訪之學者給予多項貼心之服務，為國內學校爭取主辦國際型研討會可加以參考之長處。然而，雖然主辦單位之用心可見，由於此次研討會乃屬中型之規模，雖然各與會學者之於會場中研討之熱絡，然相較於一般中大型國際研討會之會場可能規畫數個地點，或者數個樓層，此次研討會僅舉辦於 Premier Hotel Tsubaki Sapporo 之 B1 樓，故需受限於飯店之場地限制(如各議程場地較為狹小及休息區之規劃皆不甚完美)，此可提供國內學者於辦此類中、大型學術研討會之借鏡。

整體而言，本次大會舉辦頗為用心，個人於此行收穫豐富，且結識多位國際學者，希望能於未來建立更長遠的交流與合作。

### 四、攜回資料名稱及內容

1. 研討會論文集：含議程集 1 本、論文摘要集 1 本、論文全文電子檔 (USB 隨身碟)。
2. 國內外學者學術交流名片。

**A Member's Reference Degree Analysis Approach for Virtual Community**

**Shih-Ting Yang\* and Yu-You Chen**

<sup>a</sup>Department of Information Management, Nanhua University, Taiwan

stingyang@nhu.edu.tw

**1. Background**

In virtual community, a lot of knowledge contributors devote themselves to knowledge sharing, a great deal of information will surely be agglomerated, and the knowledge demanders are given the channels for more multivariate knowledge reference. However, for the clustering of referential and professional knowledge, there are considerable challenges. As the internet brings the convenience of information transfer, a lot of knowledge contributors are devoted to virtual communities successively, sharing knowledge spontaneously to obtain personal community status. For the management of numerous and plural knowledge contributors, most of virtual communities use incentive systems to encourage the members to share quality knowledge, and the violative words are removed manually, this practice is difficult to implement effective and substantial control on the knowledge contributors who accumulate community status rapidly by the defects in the incentive systems and keep spreading less referential knowledge. Therefore, the knowledge contributor's reference degree measurement depends on the spontaneous grading and comments of community members. The member measures the knowledge reference degree according to the understanding of domain expertise of knowledge contributors, discusses the comments of other members in the string and his professional ability, there will not be recognized standard due to the cognition difference between community members. Therefore, the knowledge demander must collect, probe into (poster and knowledge content) and learn domain expertise, so as to measure the knowledge contributor's reference value. However, as the existing virtual community has not provided the performance statistics of precise knowledge contributors, and the understanding of the contributor reference degree and domain expertise is insufficient, the evaluation is wrong, and a lot of information should be collected, the time cost of evaluating the knowledge contributor reference degree is increased. In addition, the less referential public knowledge is spread continuously, and the community status lacks justice and preciseness, the knowledge contributor's power to keep sharing knowledge will be reduced. Take a general view of the effect on the overall virtual community, the overall public knowledge quality in the community will degrade, then the development of virtual community is influenced.

In order to avoid the knowledge demander extracting wrong knowledge due to insufficient background knowledge, and to increase the knowledge demander's knowledge extraction effectiveness and efficiency, this paper builds the member's reference degree analysis approach,

to analyze the reference degree of domain knowledge of virtual community members for the knowledge demander's reference. Therefore, the proposed approach in this paper measures and evaluates the knowledge contributor reference degree automatically, and provides the determined data for the user, so as to solve the problems in the virtual community knowledge acquisition.

## **2. Methods**

The proposed approach in this paper uses the clustering algorithm (Zhang et al. (2010)) as the basis of data clustering, the "virtual community domainial discussion threads distribution cluster" and "target member domainial historical publication distribution cluster" can be obtained by "word segmentation and screen valid term", "screen frequent item set", "construct similar matrices", "determine cluster type" and "divide cluster of domain data". Secondly, in order to enable the knowledge extractor to browse more visually, and to obtain the specific indexes of virtual community member's knowledge specialty bias and reference degree, this paper nominates the corresponding clusters (Zhang et al. (2010)), so as to obtain the cluster names of representative virtual community, the target member domainial discussion threads and historical publication distribution clusters. Finally, this paper integrates the "vector space model" and the "NGD algorithm" proposed by Cilibrasi and Vitanyi (2007), and uses the concept of cluster similarity (Zhang et al. (2010)), to analyze the virtual community domainial discussion threads, and the similarity relationship between clusters in the target member domainial historical publications to obtain the domain knowledge reference degree of target member.

## **3. Conclusion**

The member's reference degree analysis approach proposed in this paper is divided into three major parts to analyze the virtual community member's reference degree. The first part is to divide the clusters of domain data. The second part is to define the names of clusters. The third part is to obtain the target member's domain knowledge reference degree data. This paper refers ideas from three methodologies, including Zhang et al. (2010), vector space model and Cilibrasi and Vitanyi (2007), which are improved and integrated to develop a member's reference degree analysis model applicable to virtual community. The knowledge demander can refer to the "target member's domain knowledge reference degree" obtained by this approach to avoid insufficient background knowledge resulting in wrong knowledge extraction, and the effectiveness and efficiency of knowledge extraction are enhanced eventually.

Keywords: Virtual community, Clustering, Data mining, Member's reference degree



# 科技部補助計畫衍生研發成果推廣資料表

日期:2016/10/25

科技部補助計畫	計畫名稱: 提升成員知識分享意願之虛擬社群知識審核模式(I)
	計畫主持人: 楊士霆
	計畫編號: 104-2221-E-343-001- 學門領域: 資訊系統
無研發成果推廣資料	

104年度專題研究計畫成果彙整表

計畫主持人：楊士霆			計畫編號：104-2221-E-343-001-			
計畫名稱：提升成員知識分享意願之虛擬社群知識審核模式(I)						
成果項目			量化	單位	質化 (說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等)	
國內	學術性論文	期刊論文		0	篇	
		研討會論文		3		
		專書		0	本	
		專書論文		0	章	
		技術報告		0	篇	
		其他		0	篇	
	智慧財產權及成果	專利權	發明專利	申請中	0	件
				已獲得	0	
			新型/設計專利		0	
		商標權		0		
		營業秘密		0		
		積體電路電路布局權		0		
		著作權		0		
		品種權		0		
		其他		0		
	技術移轉	件數		0	件	
		收入		0	千元	
	國外	學術性論文	期刊論文		1	篇
			研討會論文		1	
			專書		0	本
專書論文			0	章		
技術報告			0	篇		
其他			0	篇		
智慧財產權及成果		專利權	發明專利	申請中	0	件
				已獲得	0	
			新型/設計專利		0	
		商標權		0		
		營業秘密		0		
		積體電路電路布局權		0		
		著作權		0		
		品種權		0		
其他		0				

	技術移轉	件數	0	件	
		收入	0	千元	
參與計畫人力	本國籍	大專生	5	人次	
		碩士生	0		
		博士生	0		
		博士後研究員	0		
		專任助理	0		
	非本國籍	大專生	0		
		碩士生	0		
		博士生	0		
		博士後研究員	0		
		專任助理	0		
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)		此計畫分就不同主題發表於國內外各大研討會中，並獲「中國工業工程學會104年度年會暨學術研討會」-最佳論文獎。			



## 科技部補助專題研究計畫成果自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以100字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形（請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊）

論文： 已發表  未發表之文稿  撰寫中  無

專利： 已獲得  申請中  無

技轉： 已技轉  洽談中  無

其他：（以200字為限）

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性，以500字為限）

(1)理論方法層面：(A)以Tung與Lu（2010）、向量空間模型，以及Cilibrasi與Vitanyi（2007）為鑒借，藉由領域性關鍵詞彙集合之建置、問答內文核心之萃取、提問品質與問答品質之解析、問答內文契合度，以及問答語意關係距離之分析等技術，分析並計算目標虛擬社群討論串中，問答間內文及語意之整體契合程度，以供使用者作為知識品質參考依據。(B)藉由對Zhang等人（2010）、向量空間模型，以及Cilibrasi與Vitanyi（2007）等三項方法論之鑒借，經本研究之改良後，整合並發展出一套適用於虛擬社群之成員參考度解析模組，藉以分析得虛擬社群成員之領域知識可參考程度，進而避免因背景知識不足，而導致錯誤知識之擷取，並於最終提升其擷取知識之效能與效率。

(2)實務應用層面：(A)於社群管理者之層面，可為其提供標準化且具公正性群眾智慧品質之管理指標，以提升群眾智慧管理之實質效益。(B)於知識需求者之層面，可降低其篩選知識時所需花費之成本，同時提升其獲取實具契合性與品質知識之機率。

4. 主要發現

本研究具有政策應用參考價值： 否  是，建議提供機關

（勾選「是」者，請列舉建議可提供施政參考之業務主管機關）

本研究具影響公共利益之重大發現： 否  是

說明：（以150字為限）