

南 華 大 學

資訊管理學系

碩士論文

虛擬社群知識群聚度判定模式

**A Virtual Community Knowledge Clustering Degree**

**Determination Model**

研 究 生：陳俞佑

指 導 教 授：楊士霆

中華民國 103 年 6 月 9 日

南 華 大 學

資訊管理學系暨研究所

碩 士 學 位 論 文

虛擬社群知識群聚度判定模式

研究生： 陳俞佑

經考試合格特此證明

口試委員：夏子明  
洪銘建  
楊心建

指導教授：楊心建

系主任(所長)：洪銘建

口試日期：中華民國

103 年 6 月 9 日

## 南華大學資訊管理學系碩士論文著作財產權同意書

立書人：陳俞佑之碩士畢業論文

中文題目：虛擬社群知識群聚度判定模式

英文題目：A Virtual Community Knowledge Clustering Degree Determination Model

指導教授：楊士霆 博士

學生與指導老師就本篇論文內容及資料其著作財產權歸屬如下：

- 共同享有著作權
- 共同享有著作權，學生願「拋棄」著作財產權
- 學生獨自享有著作財產權

學生：陳俞佑 (請親自簽名)

指導老師：楊士霆 (請親自簽名)

中華民國 103 年 6 月 9 日

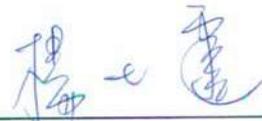
南華大學碩士班研究生  
論文指導教授推薦函

資訊管理系碩士班 陳俞佑 君所提之論文

虛擬社群知識群聚度判定模式

係由本人指導撰述，同意提付審查。

指導教授



103年6月9日

## 誌謝

為期兩年研究所的求學生涯，終究告一段落，這兩年來所看到的、聽到的、經歷的種種階段，無論是喜、怒、哀或樂，無疑皆令我的人生烙下了各種銘心的回憶，這兩年內接觸到的許多老師與同學，感謝你們能出現在我生命中，因你們的教導、關心、陪伴、協助與支持，除了使我得以順利完成碩士論文外，更令我的人生有更深層的學習與成長。

感謝指導教授 楊士霆 老師，付出相當多時間與心力，以嚴格的方式指導我撰寫論文與處理事務，使我潛移默化的學會細心、嚴謹、替他人著想，以及反思的態度，老師未曾對學生有過不合理的要求，或是不明事理的指責，老師皆能以信任學生的角度為出發點，傾聽學生解釋，並適當的指正學生的過錯，所以我很感謝老師能罵我，也很感激老師能兇我，我信服老師的教導，而每一次正確的指責也都是令我成長的動力，令我學會以多方之觀點考量他人的想法與需求，更令我因此養成反思的能力，我真的相當感謝老師，能以如此嚴格的方式帶領著我學習，也相當感激老師在非常時期時，能每天批改我進度，每天聽我練習預報，感謝老師三年來（包含大學時期）對我的諄諄教誨。

在課業的學習過程中，我相當感謝 楊美蓮 老師，令學生扎實且徹底了解資訊管理理論精隨，且當其他同學認為沒必要如此認真撰寫報告時，老師能發現學生撰寫報告的用心，並給予有別於其他學生高分的鼓勵，激勵並令我認知到擇善固執與努力後所得之成果；相當感謝 尤國任 老師耐心傾聽學生的想法，引導學生正確思考與學習，除了課程知識得以扎實習得外，更提升了學生多元思考能力；也相當感謝 陳仁義 老師，近乎每次課堂皆須撰寫心得感想的作法，這相當扎實地提升學生表達能力與培養反思態度。兩年研究所生涯中，沒人 Cover 的情況下，自行撰寫論文、扮演學長角色帶領學弟妹，以及處理老師交代事項，坦言之，起初我相當害怕，在任何能力皆尚未養成之前提下，因須兼顧課業、論文、學弟妹與老師事務等層面，令我倍感壓力，但因指導老師的耐心指導，以及好友 韓元翰 的支持與鼓勵，使我撐了下來並開始習慣這些壓力，最終，也因而了培養專業能力，以及諸多待人與處理事務的隱性能力。我很慶幸研究所生涯是在如此環境下展開，因為，在這看似相對較多荊棘的路上，我比其他人經歷更多，也學習更多，配合指導老師耐心且細心的教誨，真的令我這兩年來，過得相當相當相當的充實，謝謝 楊美蓮、尤國任 與 陳仁義 老師給予鼓勵與教導，謝謝老天所給的逆境，謝謝 韓元翰 支持與鼓勵，謝謝 阿昌 回國後幫忙處理學校事務。最重要的，感謝 楊士霆 老師這兩年來，對我的指導、照顧、支持、鼓勵與肯定，謝謝老師，真的，相當感謝！

## 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，本研究乃發展虛擬社群知識群聚度判定模式，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群聚度判定系統。此外，為驗證本系統績效，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「虛擬社群知識群聚度判定系統」，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。

**關鍵字：**知識管理、虛擬社群、知識群聚度判定、問答契合度解析、成員參考度解析。

## ABSTRACT

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's) 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 Clustering Degree Determination (KCD) 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. 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. The verification results show that the developed system is a high-performance knowledge clustering degree determination system. As a whole, this paper provides an approach for virtual community to efficiently evaluate the clustering degree of knowledge.

**Keywords:** *Knowledge management, Virtual community, Data mining, Q&A suitability, Reference degree determination*

# 目錄

碩士學位論文考試合格證明.....	I
碩士論文著作財產權同意書.....	II
論文指導教授推薦函.....	III
誌謝.....	IV
摘要.....	V
ABSTRACT.....	VI
目錄.....	VII
圖目錄.....	IX
表目錄.....	XVI
第一章、研究背景.....	1
1.1 研究動機與目的.....	1
1.2 研究步驟.....	5
第二章、文獻回顧.....	9
2.1 研究定位.....	9
2.2 虛擬社群成員行為之探討.....	11
2.2.1 虛擬社群知識凝聚行為探討.....	11
2.2.2 虛擬社群知識應用領域探討.....	18
2.3 虛擬社群知識內容之探勘.....	23
2.3.1 虛擬社群問答內文契合狀況解析.....	23
2.3.2 虛擬社群成員個人資訊解析.....	31
2.4 小結.....	39
第三章、虛擬社群知識群聚度判定模式.....	45
3.1 虛擬社群問答契合度解析模組.....	46

3.2 虛擬社群成員參考度解析模組.....	63
第四章、系統架構.....	78
4.1 虛擬社群知識群聚度判定系統之核心架構.....	78
4.2 系統功能架構.....	80
4.3 資料模式定義.....	85
4.4 系統流程.....	90
4.4.1 系統功能流程.....	90
4.4.2 系統資料流程.....	95
4.5 系統開發工具.....	96
第五章、系統實作與案例分析.....	98
5.1 系統案例之應用流程.....	98
5.2 系統案例驗證與評估.....	128
5.2.1—(A) 虛擬社群問答契合度解析之驗證.....	132
5.2.2—(B) 虛擬社群成員參考度解析之驗證.....	142
5.2.3—(C) 虛擬社群知識群聚度判定管理意涵之探討.....	157
第六章、結論與未來發展.....	164
6.1 論文總結.....	164
6.2 未來展望.....	167
參考文獻.....	168
附錄、系統功能操作說明.....	176

## 圖目錄

圖 1.1、虛擬社群群眾智慧運作之既有模式 AS-IS MODEL.....	2
圖 1.2、虛擬社群群眾智慧運作之期望模式 TO-BE MODEL .....	4
圖 1.3、研究架構.....	8
圖 2.1、研究定位圖.....	10
圖 3.1、虛擬社群知識群聚度判定模式之流程架構圖.....	46
圖 3.2、領域性關鍵詞彙集合之建立運作示意圖.....	47
圖 3.3、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.....	54
圖 3.4、虛擬社群成員參考度解析模組之運作示意圖.....	65
圖 4.1、虛擬社群知識群聚度判定系統之流程架構.....	79
圖 4.2、虛擬社群知識群聚度判定系統之功能架構.....	82
圖 4.3、虛擬社群知識群聚度判定系統運作架構.....	84
圖 4.4、虛擬社群知識群聚度判定系統之資料關聯.....	89
圖 4.5、虛擬社群資料維護模組—領域文件資料維護之功能流程.....	91
圖 4.6、虛擬社群資料維護模組—討論串維護之功能流程.....	91
圖 4.7、虛擬社群資料維護模組—成員歷史發文維護之功能流程.....	92
圖 4.8、虛擬社群問答契合度解析模組之功能流程.....	93
圖 4.9、虛擬社群成員參考度解析模組之功能流程.....	94
圖 4.10、系統參數設定模組之功能流程.....	95
圖 4.11、系統資料流程.....	96
圖 5.1、虛擬社群知識群聚度判定系統之應用流程.....	99
圖 5.2、「搜搜問問」之精華知識文章.....	99
圖 5.3、「搜搜問問」之討論串資料.....	100
圖 5.4、領域訓練文件基本資料之輸入.....	101
圖 5.5、領域訓練文件資料之預處理與新增.....	102
圖 5.6、詞彙關係網路建構之選擇畫面.....	103
圖 5.7、領域文件斷詞與有效詞篩選畫面.....	103
圖 5.8、領域文件有效詞集彙整畫面.....	103
圖 5.9、有效詞劃分與搜尋結果輸入畫面.....	104

圖 5.10、語意關係距離計算畫面.....	104
圖 5.11、整體關係距離計算與門檻選擇畫面.....	105
圖 5.12、門檻值計算與詞彙組門檻符合狀況篩選畫面.....	105
圖 5.13、獨立詞彙劃分與門檻符合狀況篩選畫面.....	106
圖 5.14、領域性詞彙關係網路建立完成畫面.....	106
圖 5.15、系統參數設定.....	107
圖 5.16、系統參數設定成功畫面.....	107
圖 5.17、討論串基本資料之輸入.....	108
圖 5.18、討論串各筆回答資料之輸入.....	108
圖 5.19、討論串提問資料之預處理與新增.....	109
圖 5.20、討論串各筆回答資料之預處理與新增.....	109
圖 5.21、討論串查詢與分析選擇介面.....	110
圖 5.22、目標提問與各筆回答資料之預處理畫面.....	110
圖 5.23、前後文配對及標題加權整體分數計算畫面.....	111
圖 5.24、詞類相關性平均分數計算畫面.....	111
圖 5.25、關鍵詞彙代表性得分解析畫面.....	112
圖 5.26、提問與回答品質得分計算畫面.....	112
圖 5.27、斷詞詞彙詞頻資訊彙整畫面.....	113
圖 5.28、問答內文契合度得分計算畫面.....	113
圖 5.29、提問與回答之有效詞組合劃分畫面.....	114
圖 5.30、各詞彙組合語意關係距離之計算畫面.....	114
圖 5.31、提問與回答有效詞組合之劃分畫面.....	114
圖 5.32、整體問答契合度得分計算畫面.....	115
圖 5.33、討論串資料之查詢.....	116
圖 5.34、檢視詳細討論串資訊.....	116
圖 5.35、成員歷史發文基本資料之輸入.....	117
圖 5.36、目標成員各筆歷史發文資料之輸入.....	117
圖 5.37、目標成員各筆歷史發文資料之預處理與新增.....	118
圖 5.38、查詢成員歷史發文資料.....	119
圖 5.39、檢視目標成員詳細歷史發文.....	119

圖 5.40、統計成員歷史發文領域分佈.....	119
圖 5.41、成員領域歷史發文之資料預處理.....	120
圖 5.42、彙整領域有效詞資訊.....	120
圖 5.43、篩選頻繁項目集.....	121
圖 5.44、建置相似矩陣並判定分群類型.....	121
圖 5.45、劃分領域資料之群集(1).....	122
圖 5.46、劃分領域資料之群集(2).....	123
圖 5.47、彙整最終分群結果.....	123
圖 5.48、查詢成員歷史發文資料.....	124
圖 5.49、檢視目標成員詳細歷史發文.....	124
圖 5.50、統計成員歷史發文領域分佈.....	124
圖 5.51、彙整成員領域歷史發文之分群資訊.....	125
圖 5.52、定義群集名稱.....	125
圖 5.53、彙整並預處理領域討論串之資訊.....	126
圖 5.54、計算領域討論串與群集間之內文相似度.....	126
圖 5.55、劃分群集與討論串之有效詞組合(1).....	126
圖 5.56、劃分群集與討論串之有效詞組合(2).....	127
圖 5.57、計算領域討論串與群集間之語意關係距離.....	127
圖 5.58、計算目標社群成員領域知識可參考程度.....	127
圖 5.59、成員歷史發文資料之查詢.....	128
圖 5.60、檢視詳細成員歷史發文資訊.....	128
圖 5.61、系統驗證與評估之架構.....	129
圖 5.62、「搜搜問問」中之知識頻道.....	132
圖 5.63、「搜搜問問」知識頻道內之精華知識.....	132
圖 5.64、第一階段問答契合度解析召回率之分佈趨勢.....	139
圖 5.65、第一階段問答契合度解析正確率之分佈趨勢.....	140
圖 5.66、各驗證週期間問答契合度解析之績效分佈趨勢.....	141
圖 5.67、第一階段成員參考度解析召回率之分佈趨勢.....	148
圖 5.68、第一階段成員參考度解析正確率之分佈趨勢.....	149
圖 5.69、各驗證週期成員參考度解析之績效分佈趨勢.....	150

圖 5.70、受測者於本研究與搜搜問問領域可參考程度之分佈趨勢.....	154
圖 5.71、本研究整體系統管理意涵探討之示意圖.....	158
圖 A.1、領域訓練文件基本資料之輸入.....	177
圖 A.2、領域訓練文件資料之預處理與新增.....	177
圖 A.3、領域文件資料之查詢.....	178
圖 A.4、欲修改領域文件資料之查詢.....	179
圖 A.5、領域文件資料之修改(1).....	179
圖 A.6、完成領域文件資料之修改(1).....	179
圖 A.7、領域文件資料之修改(2).....	180
圖 A.8、領域文件資料預處理確認介面.....	180
圖 A.9、完成領域文件資料之修改(2).....	180
圖 A.10、欲刪除領域文件資料之查詢.....	181
圖 A.11、確認欲刪除領域文件之資料.....	182
圖 A.12、完成領域文件資料之刪除.....	182
圖 A.13、討論串基本資料之輸入.....	183
圖 A.14、討論串各筆回答資料之輸入.....	184
圖 A.15、討論串提問資料之預處理與新增.....	184
圖 A.16、討論串各筆回答資料之預處理與新增.....	184
圖 A.17、討論串資料之查詢.....	185
圖 A.18、檢視詳細討論串資訊.....	186
圖 A.19、欲修改討論串資料之查詢.....	188
圖 A.20、討論串詳細資料之檢視.....	188
圖 A.21、討論串資料之修改(1).....	188
圖 A.22、完成討論串資料之修改(1).....	189
圖 A.23、討論串資料之修改(2).....	189
圖 A.24、提問內容資料預處理確認介面.....	189
圖 A.25、完成討論串資料之修改(2).....	190
圖 A.26、討論串資料之修改(3).....	190
圖 A.27、回答內容資料預處理確認介面.....	190
圖 A.28、完成討論串資料之修改(3).....	191

圖 A.29、討論串資料之修改(4) .....	191
圖 A.30、提問與回答內容資料預處理確認介面 .....	191
圖 A.31、完成討論串資料之修改(4) .....	192
圖 A.32、欲刪除討論串資料之查詢 .....	193
圖 A.33、檢視詳細討論串資訊 .....	193
圖 A.34、確認欲刪除之回答資料 .....	193
圖 A.35、完成討論串回答資料之刪除 .....	194
圖 A.36、確認欲刪除之討論串資料 .....	194
圖 A.37、完成討論串資料（包含該筆討論串所有回答）之刪除 .....	194
圖 A.38、成員歷史發文基本資料之輸入 .....	196
圖 A.39、成員基本資料之比對 .....	196
圖 A.40、新成員資料之建置與各筆歷史發文資料之輸入 .....	196
圖 A.41、目標成員各筆歷史發文資料之預處理與新增 .....	197
圖 A.42、成員歷史發文資料之查詢 .....	198
圖 A.43、檢視詳細成員歷史發文資訊 .....	198
圖 A.44、欲修改成員歷史發文資料之查詢 .....	199
圖 A.45、成員歷史發文詳細資料之檢視 .....	200
圖 A.46、成員歷史發文資料之修改(1) .....	200
圖 A.47、完成成員歷史發文資料之修改(1) .....	200
圖 A.48、成員歷史發文資料之修改(2) .....	201
圖 A.49、發文內容資料預處理確認介面 .....	201
圖 A.50、完成成員歷史發文資料之修改(2) .....	201
圖 A.51、欲刪除成員資料之查詢 .....	202
圖 A.52、檢視詳細成員資料資訊 .....	203
圖 A.53、確認欲刪除之歷史發文資料 .....	203
圖 A.54、完成發文資料之刪除 .....	203
圖 A.55、確認欲刪除之成員資料 .....	204
圖 A.56、完成成員資料（包含目標成員所有歷史發文）之刪除 .....	204
圖 B.1、詞彙關係網路建構之選擇畫面 .....	207
圖 B.2、領域文件斷詞與有效詞篩選畫面 .....	207

圖 B.3、領域文件有效詞彙彙整畫面 .....	207
圖 B.4、有效詞劃分與搜尋結果輸入畫面 .....	208
圖 B.5、語意關係距離計算畫面(1).....	208
圖 B.6、語意關係距離計算畫面(2).....	208
圖 B.7、整體關係距離計算與門檻選擇畫面 .....	209
圖 B.8、門檻值計算與詞彙組門檻符合狀況篩選畫面 .....	209
圖 B.10、獨立詞彙劃分與門檻符合狀況篩選畫面 .....	209
圖 B.11、領域性詞彙關係網路建立完成畫面 .....	210
圖 B.12、討論串查詢與分析選擇介面 .....	213
圖 B.13、目標提問與各筆回答資料之預處理畫面 .....	213
圖 B.14、關鍵詞彙前後文語句分割畫面 .....	213
圖 B.15、關鍵詞彙前後文語句斷詞畫面 .....	214
圖 B.16、前後文配對及標題加權整體分數計算畫面 .....	214
圖 B.17、詞彙距離採納集合建置畫面 .....	214
圖 B.18、詞類相關性平均分數計算畫面 .....	215
圖 B.19、前後文與詞類數據彙整畫面 .....	215
圖 B.20、關鍵詞彙代表性得分解析畫面 .....	215
圖 B.21、提問與回答品質分析資訊彙整畫面 .....	216
圖 B.22、提問與回答品質得分計算畫面 .....	216
圖 B.23、問答內文契合度分析資訊彙整畫面 .....	216
圖 B.24、斷詞詞彙詞頻資訊彙整畫面(1).....	217
圖 B.25、斷詞詞彙詞頻資訊彙整畫面(2).....	217
圖 B.26、問答內文契合度得分計算畫面 .....	217
圖 B.27、問答語意關係距離分析資訊彙整畫面 .....	218
圖 B.28、提問與回答之有效詞組合劃分畫面 .....	218
圖 B.29、各詞彙組合語意關係距離之計算畫面(1).....	218
圖 B.30、各詞彙組合語意關係距離之計算畫面(2).....	219
圖 B.31、提問與回答有效詞組合之劃分畫面 .....	219
圖 B.32、整體問答契合度得分計算畫面 .....	219
圖 C.1、查詢成員歷史發文資料 .....	222

圖 C.2、統計成員歷史發文領域分佈 .....	222
圖 C.3、成員領域歷史發文之資料預處理(1).....	222
圖 C.4、成員領域歷史發文之資料預處理(2).....	223
圖 C.5、彙整領域有效詞與輸入最小支持度 .....	223
圖 C.6、篩選頻繁項目集 .....	223
圖 C.7、建置相似矩陣與判定分群類型 .....	224
圖 C.8、劃分領域資料之群集(1).....	224
圖 C.9、劃分領域資料之群集(2).....	224
圖 C.10、彙整最終分群結果 .....	225
圖 C.11、查詢成員歷史發文資料 .....	227
圖 C.12、統計成員歷史發文領域分佈 .....	227
圖 C.13、彙整成員領域歷史發文之分群資訊 .....	227
圖 C.14、定義群集名稱 .....	228
圖 C.15、彙整並預處理領域討論串之資訊 .....	228
圖 C.16、計算領域討論串與群集間之內文相似度 .....	228
圖 C.17、劃分群集與討論串之有效詞組合 .....	229
圖 C.18、計算領域討論串與群集間之語意關係距離 .....	229
圖 C.19、計算目標社群成員領域知識可參考程度 .....	229

## 表目錄

表 2.1、個體成員知識分享意圖之文獻彙整表.....	14
表 2.2、群體成員群眾智慧之文獻彙整表.....	18
表 2.3、虛擬社群知識應用領域探討之文獻彙整表.....	22
表 2.4、問答核心資訊萃取之文獻彙整表.....	26
表 2.5、語句及詞彙相關性分析之文獻彙整表.....	30
表 2.6、成員專業領域偏向分析之文獻彙整表.....	34
表 2.7、成員知識可信度分析之文獻彙整表.....	38
表 2.8、本研究與過去文獻差異彙整表.....	41
表 5.1、系統管理者所蒐集之部份精華知識文章.....	100
表 5.2、一般使用者所蒐集之部份討論串資料（應用於「問答契合度解析」）.....	100
表 5.3、一般使用者所蒐集之部份討論串資料（應用於「成員參考度解析」）.....	101
表 5.3、中文知識型虛擬社群質化之比較.....	130
表 5.4、虛擬社群問答契合度解析訓練資料表（部分資料）.....	134
表 5.5、虛擬社群問答契合度績效判定測試資料彙整表（已轉換為繁體中文）.....	134
表 5.6、第一階段驗證資料—評斷一致性程度分析彙整表.....	137
表 5.7、第二階段驗證資料—評斷一致性程度分析彙整表.....	137
表 5.8、虛擬社群問答契合度解析—第一階段驗證結果（共 200 筆訓練資料）.....	139
表 5.9、虛擬社群問答契合度解析績效彙整.....	141
表 5.10、各項驗證指標成長率之彙整表.....	142
表 5.11、虛擬社群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） .....	144
表 5.12、虛擬社群成員參考度解析訓練資料表（部分資料）.....	145
表 5.13、虛擬社群成員參考度解析—第一階段驗證結果（共 200 筆訓練資料）.....	148
表 5.14、虛擬社群成員參考度解析績效彙整.....	150
表 5.15、各項驗證指標成長率之彙整表.....	151
表 5.16、本系統判定 10 位受測者所得成員參考度之彙整表.....	152
表 5.17、搜搜問問衡量 10 位受測者所得成員參考度之彙整表.....	153
表 5.18、本系統與搜搜問問成員參考性判定結果之比較.....	154

表 5.19、虛擬社群知識群聚度判定之測試資料(1) (部分內容) ..... 159

表 5.20、虛擬社群知識群聚度判定之測試資料(2) (部分內容) ..... 160

表 5.21、虛擬社群知識群聚度系統判定結果..... 161



# 第一章、研究背景

本章將針對本論文之研究動機、研究目的與研究步驟依序進行說明。首先乃透過研究動機與目的之描述，點出本論文研究之緣由。其次，於研究步驟中，說明本論文規劃進行之流程。各小節詳細內容說明如下。

## 1.1 研究動機與目的

過去虛擬社群尚未風行之時，網際網路上之知識創造行為，主要形式以網頁知識為主，僅能以知識貢獻者己身之專業知識能力，進行封閉且回饋性質低之知識創建；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「Mobile01」、「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，進而提升虛擬社群之整體發展與造就群眾智慧之累積（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 1.1 所示。



圖 1.1、虛擬社群群眾智慧運作之既有模式 As-Is Model

如圖 1.1 所示，於既有之虛擬社群群眾智慧運作機制中，知識貢獻者將自發性地對資訊進行評論及評分，並根據個人認知分享己身見解以累積群眾智慧，而知識需求者將自行評估資訊可採納之程度，進行相關知識之學習與累積，並轉換為知識貢獻者，散播並貢獻所學之知識，藉以令知識持續流通並累積更多群眾智慧，雖既有之運作機制能為虛擬社群持續累積群眾智慧，但如此作法長遠觀之將衍伸些許問題，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))。

如以「Mobile01」為例，知識需求者於此發佈提問型之討論串，以尋求社群成員對該事件之見解，爾後，許多成員相繼以己身之知識專業能力回覆相關之建議，當中之回覆不乏出現差異兩極之內容(如相同之產品，成員 A 極力推薦，而成員 B 相當不推薦)，此些內容亦由社群成員對其做出評論，於此藉成員之回覆群聚相關之知識，而最終最佳建議之採納，將由知識需求者綜合其他成員之想法與己身經驗後以進行評估，但此作法將因需求者對該事件認知之不足，致使低品質資訊之採納與學習，當社群成員學習並散播相關低品質之資訊，將因社群管理者難以對低品質資訊之傳播進行實質之管控，導致整體虛擬社群中群眾智慧品質之下降，同時降低社群成員分享知識之意願，進而影響虛擬社群之發展。

有鑑於上述問題，本研究乃建構「虛擬社群知識群聚度判定」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。

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「虛擬社群知識群聚度判定」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，於此將相關判定得之數據提供予社群成員與社群管理者參考，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；針對社群管理者而言，其將可參考相對應數據，進行相關管理措施之制定，藉以使虛擬社群獎勵機制之互惠規範得以順利運作，以於最終改善既有虛擬社群群眾智慧運作機制中所面臨之問題。

綜上所述，改善後虛擬社群群眾智慧運作機制(即虛擬社群群眾智慧運作之 TO-BE 模式)，如圖 1.2 所示。



圖 1.2、虛擬社群群眾智慧運作之期望模式 To-Be Model

整體而言，本研究係針對虛擬社群之「知識內文」與「使用者行為」等兩面向之資料進行探勘，判定虛擬社群之知識群聚度，並將相關判定得之數據提供予社群成員與社群管理者參考，於社群管理者之層面，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；於知識需求者之層面，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，並減少低品質資訊再

傳遞之行為；由於知識貢獻者持續分享之關鍵因素，可溯根於分享時他人對其之評價，以及己身對收穫之認知（Jin 等人（2013）），因此，於知識貢獻者層面中，除虛擬社群成員對知識貢獻者所提供之評價外，本研究亦將藉由對問答契合與成員參考度之解析，對知識貢獻者之歷史發文內容，提供具公正性與正確性之評量數據，如此將因知識貢獻者對己身知識可參考性之認知與確認，提升持續分享知識之意願，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。

## 1.2 研究步驟

本研究之目的係判定虛擬社群之知識群聚度，包含問答契合程度，以及成員領域知識可參考程度，並將相關判定得之數據提供予知識擷取者參考，以減少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，本研究之研究架構可分為五大步驟。藉由研究動機與目的以進行文獻回顧與探討，並於文獻回顧後確立本研究定位方向，以發展本研究之方法論，並依照方法論開發雛形系統，最後以一實例進行系統案例驗證，以確認系統之運作與成效，並評估本研究方法論之實用性，其各步驟詳細說明分述如下：

### 步驟一、背景資料蒐集與探討

根據本論文之研究背景、動機與目的進行相關資料蒐集，本研究主要係針對虛擬社群之「知識內文」與「使用者行為」解析等兩層面進行探討，並彙整所涉及主題包含「虛擬社群知識凝聚行為」與「虛擬社群知識內容之探勘」等議題，透過文獻蒐集與研讀，以了解問答契合程度及成員領域知識可參考程度解析之現行作法與發展方向，進而建構本研究之方法論與系統模型。

### 步驟二、研究方向定位

透過相關文獻之蒐集與探討即可得知，於虛擬社群問答內文契合狀況解析之層面中，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，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。

於虛擬社群成員個人資訊解析之層面中，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，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，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，然群眾智慧亦實存價值。

有鑑於上述，本研究於「虛擬社群問答契合度解析」層面之目標乃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「虛擬社群問答契合度解析模組」，當中，包含「領域性關鍵詞彙集合建立」與「虛擬社群問答契合度解析」兩層面之核心推論，另一方面，本研究於「虛擬社群成員個人資訊解析」層面之目標乃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」。當中，包含「虛擬社群成員資料分群」與「虛擬社群成員參考度解析」兩層面之核心推論。

整體而言，本研究發展之「虛擬社群知識群聚度判定模式」分為「虛擬社群問答契合度解析模組」與「虛擬社群問答契合度解析模組」，本研究乃整合上述兩模組，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，提供相關數據予知識擷取者，以減少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，活絡虛擬社群之利用率，進而催化虛擬社群持續性之發展。

### 步驟三、研究模式之建立與系統開發

本研究共有四大主題需完成，分別為「虛擬社群知識群聚度判定模式規劃」、「虛擬社群問答契合度解析方法論」、「虛擬社群成員參考度解析方法論」與「系統功能開發」，以下為各項主題之細述說明：

#### 主題一、虛擬社群知識群聚度判定模式規劃

蒐集並回顧虛擬社群問答契合度與成員參考度解析等議題之相關文獻

瞭解虛擬社群問答契合度與成員參考度解析相關技術，以設計相關方法論

瞭解虛擬社群問答契合度與成員參考度解析技術之差異，以建構整合性之模式

## 主題二、虛擬社群問答契合度解析方法論

解析虛擬社群問答資料與領域訓練文件

設計虛擬社群問答契合度解析之方法論

## 主題三、虛擬社群成員參考度解析方法論

解析虛擬社群成員領域歷史發文內容

設計虛擬社群成員參考度解析之方法論

## 主題四、系統功能開發

開發領域性關鍵詞彙集合建立功能

開發虛擬社群問答契合度解析功能

開發虛擬社群成員資料分群功能

開發虛擬社群成員參考度解析功能

## **步驟四、案例驗證**

於此步驟乃將本研究所建構之「虛擬社群知識群聚度判定系統」為基，並尋求相關之實務應用案例進行驗證，以確認本研究所發展系統模組之正確性與實用性。

## **步驟五、成果分析與結論**

透過案例驗證之執行成效與分析，了解本研究預期成果與實際成效間之符合程度，藉，並評估本論文所發展之方法論與系統模組之實用性與準確性。最後，藉由分析評估結果規劃本研究之未來發展與應用方向。

綜合上述之研究步驟之說明，本研究首先乃根據研究動機與目的相關文獻，以釐清研究定位並確立研究方向，其次，依據研究方向建立本研究之研究模式，並以此為基礎開發一套虛擬社群知識群聚度判定系統，最後，再以案例驗證本研究所發展之方法論與系統模組，以確認本研究之實用價值與發展性。本論文之研究架構如圖 1.3 所示。

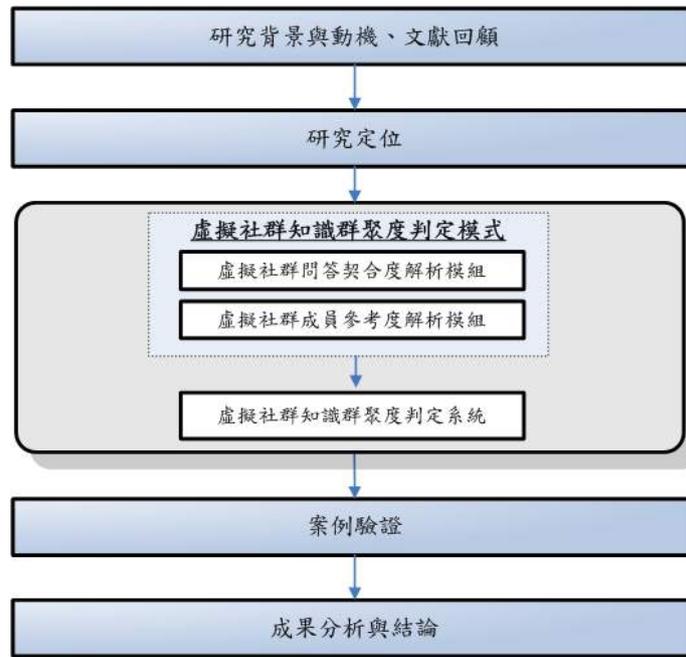


圖 1.3、研究架構



## 第二章、文獻回顧

本研究主要目的係以自動化方式同時衡量與評估問答內文契合程度，以及知識貢獻者可參考程度，於此將相關判定得之數據提供予知識需求者參考，以降低其篩選知識所需花費之時間成本，且同時提升其獲取所需且正確知識之機率，間接減少錯誤資訊再傳遞行為，藉以增加實具品質群眾智慧之累積。因此，針對上述研究目的，於探討相關文獻前乃先行釐清本研究之研究定位，以瞭解本研究與現今相關研究之差異性及本研究之研究價值。

### 2.1 研究定位

針對本研究所發展之「虛擬社群知識群聚度判定模式」而言，主要目的係面向虛擬社群問答之契合程度，以及成員可參考程度進行相關之解析，因此，針對模式相關演算法開發之主軸而言，本研究以「知識內容」與「使用者行為」等兩大層面進行文獻之回顧，而為確認本研究知識群聚度凝聚觀點於應用層面中之適切性，本研究亦針對相關管理面向之文獻進行探討，以期由技術層面，以及管理面向雙向議題之整合，發展一套實具績效及應用性之模式與系統。

有鑑於上述，為符合於本研究「虛擬社群知識群聚度判定模式」之主軸，於此，本研究將研究主題劃分為「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。以下即針對此兩大主題之相關研究進行文獻回顧及探討。

於「虛擬社群成員行為之探討」議題中，針對過去研究進行探討，從中可歸納得既有研究之主軸，主要可劃分為「虛擬社群知識凝聚行為」及「虛擬社群知識應用領域」兩子議題。於「虛擬社群知識凝聚行為」中，過去研究主要針對個體成員之知識分享意圖，以及群眾智慧之影響力，以質化及量化之方式探討得個體成員自發性分享知識，以及群眾智慧對虛擬社群整體發展之影響，因此，於此子議題中亦可進一步劃分為「個體成員知識分享意圖」與「群體成員群眾智慧」兩方面進行探討；於「虛擬社群知識應用領域」中，進行相關文獻之回顧，本研究發現就虛擬社群知識之應用面而言，大多數之研究主要面向「教育領域」及「醫療領域」，執行相關知識管理之策略，以藉由虛擬社群之網路化平台，進行跨地區知識與資訊之傳播，除上述兩領域外，部份研究則針對資訊超載，以及社交網路進行相關深入之探討，以期透過知識之管理與公眾議題之探測，進行相關「其他領域」之應用。

於「虛擬社群知識內容之探勘」議題中，本研究於研讀大量文獻後，歸納得過去研究主要面向虛擬社群文章相似度分析、關鍵知識內容萃取，以及社群成員專業性等課題進行探討，亦即既有研究大多以「問答核心資訊萃取」、「語句及詞彙相關性分析」、「成員專業領域偏向」與「成員知識可信度」等課題為主軸，進行相關虛擬社群技術層面知識內容之探勘。於此，為契合於本研究「虛擬社群知識群聚度判定模式」之主軸，本研究將「問答核心資訊萃取」及「語句及詞彙相關性分析」課題整合為「虛擬社群問答內文契合狀況解析」子議題，進行相關整合性深入之探討，藉以用於發展一套整合語料庫與上下文之自動化解析之「虛擬社群問答契合度解析模組」，另一方面，本研究則將「成員專業領域偏向」與「成員知識可信度」課題整合為「虛擬社群成員個人資訊解析」子議題，進行整合性文獻之回顧，以此為同時具備詞頻與語意解析之「虛擬社群成員參考度解析模組」提供相關演算法設計之方向。

綜合以上所述，本論文所涉及之各項主題領域可以圖 2.1 呈現架構關係；圖中灰色部分乃代表本研究所強調之研究主題。如圖 2.1 所示，以本研究所發展之方法論為主軸，本研究根據過去大量研究成果依不同研究議題蒐集相關文獻資料，並針對不同主題進行相關細節之說明。本章文獻回顧即針對「虛擬社群成員行為之探討」與「虛擬社群知識內容之探勘」兩大議題進行說明。

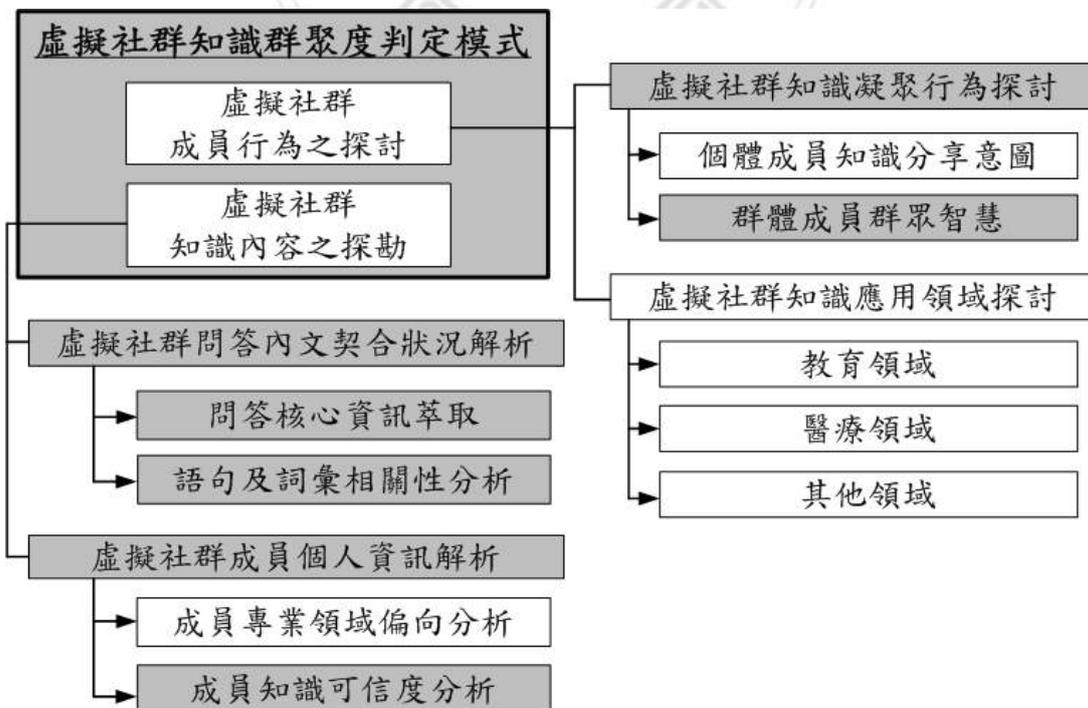


圖 2.1、研究定位圖

## 2.2 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。

### 2.2.1 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，首先，該研究結合正義與信任理論，於一般成員層面，探討得分配及人際正義，對成員間信任關係具正向影響；於管理者層面，探討得程序與資訊正義，對成員與管理者間信任程度具正向影響；最後，該研究將心理學相關之利他主義與嚴謹自律性整合於模型中，並歸納得成員給予之鼓勵（讚許）及人際關係對待等公平性，以及管理者管理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。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，於前者中，該研究認為成員分享知識行為，取決於其本身對知識貢獻之自發性、主動性與熱絡性，於後者中，該研究認為成員分享知識之動機，歸因於其評估貢獻知識後所獲得之利益，綜合而言，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，首先，於資訊系統成功模式

層面中，分別以成員「知覺資訊品質」、「知覺系統品質」與「滿意度」作為探討目標，其次，於從後採用研究之概念中，則以成員「知覺效益」探討其使用社群之感受，最後，藉由成員使用經驗有效樣本之蒐集、分析與討論，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，為此，該研究蒐集 218 份有效問卷，針對科技類虛擬社群成員分享知識之意圖進行深入探討，並從中發現樂觀度對知覺易用及有用性，以及創新度對知覺易用與相容性具備顯著之正向影響，此外，於知識分享意圖之層面，知覺易用及有用性對其亦存顯著之正向影響，綜上結果，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。而 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，分析結果顯示，自我效能與認知相對優勢，皆為知識貢獻與蒐集行為之影響因子，而認知相容性僅為貢獻行為之影響因子，其次，信任乃影響知識貢獻與蒐集行為之因子，而互惠規範之因子僅影響蒐集行為，再者，知識蒐集行為將雙方面提升知識利用率及社群運作效能，而知識貢獻行為僅提升前者，綜上而論，該研究認為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，而貢獻與蒐集行為係主要提升虛擬社群運作效能之前置因素。

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，並藉由對知識平台中 301 筆問卷資料之蒐集與分析，發現「美譽度之提升」、「互惠」與「助人樂趣」皆可提升知識貢獻者「知識自我效能」與「確認感」，確認感亦將單方面影響其「滿意度」，而知識自我效能則雙方面影響「滿意度」及「持續分享意圖」，此外，滿意度係持續分享意圖之最大影響因子，綜上而論，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) 以部落格社群 (Blogging Communities; BCs) 為基礎，融合社會認同理論發展一套整合性模型，以探討社群成員

分享知識之動機與因素，於此，該研究蒐集問卷並以量化方式，分析得成員於社群中之「參與度」、「善意性」、「社交技巧性」及「創造力」對「知識貢獻性」具顯著影響，尤以「參與度」為重，亦對其他因子具間接之影響，藉以形塑社群成員「知識貢獻性」，綜合言之，該研究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，分別以成員「認知動機」、「認知目標」與「知識擷取策略」等三層面，針對成員擷取知識採用之策略進行探討，經由相關問卷蒐集與分析後，該研究認為當成員分享知識具備足夠之信心（自我效能），並於分享之同時感受相當成就（愉悅性），且擁有足夠時間支持其分享行為（時間資源），成員將因個人所擁有之「時間資源」而設置相應目標，總體而言，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) 整合自我知覺理論與流動性因子，將虛擬社群成員知識創建行為，劃分為「知識分享行為」、「流動性經驗」與「知識創造」等三層面進行探討，於蒐集社群成員資料以量化方式進行分析後，該研究發現成員之求知與貢獻知識行為，將因個體認知之不同，產生相應程度之愉悅性及焦點關注度，而其產生之愉悅及對社群發展之關切（焦點關注度），將提升創造知識之可能性，是故，該研究認為管理者應透過對外環境因子之調整，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。

綜合於上述之探討，本研究亦根據「個體成員知識分享意圖」議題彙整相關結論，如表 2.1 所示。

表 2.1、個體成員知識分享意圖之文獻彙整表

過去研究	採用理論或方法	探討對象與目的	探討得之結論
Fang 與 Chiu (2010)	正義、信任理論；心理學	成員與管理者對知識分享意圖之影響	成員待人、管理者管理之原則性及公平性，將改變成員分享知識之意圖。
Tsai 與 Pai (2013)	遠端、近端影響因子；調節變項；定性、定量研究。	成員自發性之參與行為	整體滿意度與可識別性，將溯根於成員態度、背景、熱絡度與資訊之可用性，並影響成員自發性之參與。
Zhao 等人 (2013)	非自覺性意象；自覺性意象	成員持續使用社群之意圖	以新成員為主要目標，訂定相關策略，令其感受貢獻知識獲致之利益，提升其參與社群時間及貢獻知識頻率，進而為社群提供持續性發展契機。
Zheng 等人 (2013)	資訊系統成功模式；資訊系統從後採用研究概念	成員持續分享知識之動機	管理者應將社群視為一資訊系統，以提升成員需求及滿意度作為最終管理策略施行之方針，藉以維繫成員持續使用虛擬社群之動機。
Hung 與 Cheng (2013)	科技準備度；科技接受模型	成員分享知識之意圖	成員之樂觀度、創新度、知識內容之易用及有用性，將影響成員分享知識之意圖。
Chen 與 Hung (2010)	社會認知理論；社會交換理論	成員知識分享與蒐集知識之因素	知識貢獻與蒐集行為之產生，可溯根於自我效能、認知相對優勢、相容性與互惠規範，而貢此些行為亦提升虛擬社群運作效能之前置因素。
Jin 等人 (2013)	期望確認理論；知識分享因子；知識自我效能	知識貢獻者持續提供回應之因素	評價、對收穫之認知，將提升貢獻者自我效能及確認感，以獲致滿足感並提升持續分享意願。
Kim、Zheng 與 Gupta (2011)	社會認同理論	成員知識分享之動機與因素	社群成員決定性之參與，以及成員本身於網路社群中之善意性（樂於解答）、社交技巧性及創造力，將直接影響其分享知識之可能性。
Kim、Song 與 Jones (2011)	社會認知理論；目標設定理論	成員對社群服務之需求	服務提供者應以直覺且簡便化方式呈現原為龐大且繁雜之資訊，藉以使成員願意設置及達成認知目標，進而提升社群知識間之運用及流動狀況。
Yan 等人 (2013)	自我知覺理論；流動性因子	成員求知與貢獻知識行為	成員之求知與貢獻知識行為，將因認知之不同，產生相應程度之自覺愉悅性及焦點關注度，促使並提升其創造知識之可能性。

## (B) 群體成員群眾智慧

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，該研究分別於知識內容面探討「事實型」、「概念型」、「程序型」與「後設型」知識，於認知過程面探討使用者「記憶」、「理解」、「應用」、「分析」、「評估」與「創造」能力，於社交面探討使用者間「情感」、「互動」、「凝聚」與「議論」行為，於使用者群聚知識過程中，對其他使用者學習與社群發展效益之影響，討論結果發現「程序型知識」、「使用者理解能力」、「使用者間互動行為」為提升效益最關鍵因素，綜上而論，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，首先，該研究將環境中之相似知識，分別預先設定於低、中與高等層次，並於使用者創建知識之同時，隨機調整知識相似之程度，於此蒐集並分析相似知識對群眾智慧形成時之影響，其次，則分析相對知識對於上述環境中之影響，最終，結合上述分析結果，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，首先，該研究針對蒐集得之社交網路資料，進行結構化動作以標準化方式儲存之，接著，該研究使用子空間分群演算法，將社群中之個體（文章）群聚為多個標題，爾後，則針對成員發文狀況，將成員劃分至最近似相應群集中，接著，該研究即以鏈結分析演算法，解析各群集間節點之關係鏈結程度，最後，即進行社交網路鏈結圖之自動生成，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) 認為解析知識關聯性間之變更，即可掌握新趨勢走向，而趨勢將取決於知識貢獻者間對知識概念循序漸進與動態之集體轉移，是故，該研究發展以動態群眾知識網路為基之新興趨勢分析模式，首先，該研究蒐集大量歷史文章，並以結構化方式表達文章中之關鍵字，次之，將各關鍵字視為節點分析關聯性，並根據關聯程度生成知識群集，爾後，即根據群集間之關係距離建置知識群集網路，最後，該研究以文章發布時間為基礎，根據時序狀況賦予各節點相應權重係數，透過關鍵字向量權重之解析，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，首先，該研究蒐集各文章中之標籤與標題資訊，並根據標籤間與標題間之相似度進行整合 k-means 與階層式之分群，並以圖形化方式進行呈現；次之，執行預處理後即以向量空間模型進行文章集合向量之表示，視為文章編碼，並結合潛在語意分析與奇異值分解法 (Singular Value Decomposition; SVD) 以建置文章相似矩陣；最後，將標籤分群資訊，整合於文章相似矩陣中進行分群，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之結果，藉以同時解

法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) 以擴散激發理論、資訊覓食理論及協同進化模型為基礎，發展整合型擴展資訊分析模型，於實驗一中，該研究蒐集受測者之瀏覽及標記行為，並從中發現受測者自身興趣偏好將影響其對標記行為；實驗二中，於受測者可觀看他人對內容標記之開放環境下，進行相同測試，並從中發現當受測者內心偏好之標記與群眾標籤不符合時，其將檢索相關資訊頁面，對資訊內容進行了解與學習，以修訂原先之標記；綜上而論，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之議題，所致巨量搜尋結果之問題，於系統核心模組中，將藉由群眾標籤與知識本體之匹配，令原以關鍵字形態儲存之標籤轉換為具語意架構之樹狀結構，此外，該系統亦針對使用者輸入之搜尋關鍵字、知識涵義及其指定之搜尋條件，進行語意概念分析，並以語意概念圖方式進行呈現，除提供更契合搜尋結果外，更因成員擁有編修知識本體之權限，而為群眾協同作業之應用提供更進一步且嶄新之擴展。

另一方面，Huang 等人 (2014) 透過對社交網路概念之借鑑，將群眾標籤之資源應用於使用者關係之鏈結與推薦，以拓展群眾標籤之應用層面，首先，該研究面向使用者所註記標籤頻率，相關之歷史行為進行探勘，以從中洞悉使用者興趣之偏向性，次之，即由興趣偏向性中解析使用者間偏向之相似狀況，以計算得各使用者間之關係鏈結程度，另一方面，該研究亦根據使用者歷史註記之時序（註記標籤之順序），以及各時期之註記偏向，分析得使用者興趣偏好之成長曲線，於最終，整合所有分析結果，該研究即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。而 Yang 與 Chen (2008) 融合社交網路與語意分析之概念及技術，發展互動性協同知識分享平台，而該平台之核心模組分別為知識資源管理模組，以及搜尋行為轉譯模組，透過對核心模組之應用，該研究針對知識內容進行知識本體結構化之處理，並藉語意技術分析知識需求者搜尋行為，且透過點對點社交網路三層面概念包含群眾協作、人際關係網路，以及知識網路之應用，解析社群成員間之關係鏈結，以取得成員間之關係程度，並建置社交鏈結網路，為此，該研究除提供知識需求者適配之搜尋需求，更增進同質性成員互動之頻率，進而於最終提升群眾智慧凝聚之品質。

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，首先，以預先定義之單字辭典，進行單字拼字之比對，藉以過濾錯誤拼字之詞彙，接著，運用語意相似性功能，分析得各關鍵字相似度，最後，則透過高相似度關鍵字之彙整，由各知識文章中萃取得具鑑別性之知識關鍵項目(Key-Term)，並取得兩資料集合間知識關鍵項目之重疊狀況，於此，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。

綜上所述，就虛擬社群中知識凝聚行為之議題而言，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，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面向之管理層面中，主要針對群眾智慧之影響力，包含對成員個人能力之改變，以及對虛擬社群整體發展之影響，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，而於技術層面中，則主要著重於虛擬社群整體性群眾智慧(力量)之應用，包含社群探測、趨勢探索及協同作業等。整體而論，經此兩層面之文獻回顧後，本研究從中歸納得，於虛擬社群中凝聚而得之群眾智慧，具備不可取代與相當之價值，而群眾智慧之形成亦可溯根於個體成員之分享行為，亦即群眾智慧與個體成員之關係乃相輔相成，因此，就本研究以「知識群聚度解析」之主軸而言，若欲解析得實具參考價值之結果，面向於個體成員與群眾智慧進行相關之分析乃勢在必行。

綜合於上述之探討，本研究亦根據「群體成員群眾智慧」議題，彙整相關探討之結論，如表 2.2 所示。

表 2.2、群體成員群眾智慧之文獻彙整表

過去研究	研究方法	探討對象與目的	結論或研究貢獻
Bloomer 等人 (2013)	以使用者群眾知識觀點，進行整體化定性層面之探討。	群眾智慧對虛擬社群成員帶來之效益	程序型知識、使用者理解能力、互動行為係提升虛擬社群成員效益之因素。
Moskaliuk 等人 (2012)	以維基百科知識分享環境為基礎，探討群眾智慧之形成。	相似與相對知識對群眾智慧形成時造成之影響	若存在中度相似與高度相對知識，群眾智慧形成之效率及品質將達最高點，且群眾智慧係由個體與社會網路間相互動態外化與內化過程形塑而得。
Zhao 等人 (2012)	發展以主題為導向社群探測模式	虛擬社群探測	整合社群個體分群及鏈結分析技術，提供協同學學習、專家尋找、知識分享等研究參考，以運用於群眾智慧探勘領域之發展。
Liu 等人 (2013)	發展動態群眾知識網路為基之新興趨勢分析模式	群眾智慧於趨勢探索之應用	建置知識群集網路並視為趨勢演進圖，提供決策者參考，進而針對趨勢之演進與方向訂定決策目標。
Agarwal 等人 (2010)	發展以群眾智慧為基整合型之分群演算法	群眾智慧於分群演算法之應用	以群眾標籤聚類得具使用者觀點之結果，並提供不具結構及資訊較少文章語意相似度之解析。
Cress 等人 (2013)	以擴展資訊分析模型，探討標籤對使用者行為之影響。	社交網路中群眾標籤之影響力	群眾智慧將影響受測者標記方式，亦促使偶發性之自主學習，而此行為將與群眾智慧進行雙向回饋，以修正成更為正確之標籤資訊。
Huang 等人 (2012)	發展語意標籤為基群眾註記系統	群眾智慧於協同作業之應用	整合群眾標記資源、知識本體、語意搜尋及視覺概念圖，提供契合使用者所需搜尋結果，並為群眾協同作業之應用提供更進一步且嶄新之擴展。
Huang 等人 (2014)	發展群眾標籤為基成員關係鏈結與推薦方法	群眾智慧於成員關係鏈結與推薦之應用	分析成員興趣偏向性、關係鏈結程度，建置社交網路拓樸，應用於推薦機制中，以拓展並延伸群眾標籤資源之應用。
Yang 與 Chen (2008)	發展整合社交網路與語意分析協同知識分享平台	群眾智慧於協同作業之應用	分析成員搜尋行為，並解析成員間之關係鏈結，以提供成員適切之搜尋結果，並增進同質性成員互動之頻率，進而於提升群眾智慧凝聚之品質。
Hong 與 Scardamalia (2014)	發展詞彙分析工具於萃取關鍵知識之方法	群眾智慧重點知識之萃取	有效過濾大量知識文章中所蘊含之錯誤資訊，並提供具鑑別性之知識關鍵項目予知識需求者參考。

## 2.2.2 虛擬社群知識應用領域探討

### (A) 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，包含量化型與文字型資料，進行相關資料之探勘，首先，該研究藉由統計方法，針對量化型資料進行數據之統計，次之，針對文字型資料進行語意分析，以從中萃取得關鍵詞彙並以詞袋進行彙整與儲存，接著，該研究以 TF-IDF 演算法分析各討論串中詞袋之相似程度，並藉量化型資料統計得之數據，以解析並取得學生於社群中討論

主題之分布狀況，最後，藉由自動化主題偏向探測演算法之運用，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，**Dascalu 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，故該研究發展最佳化多學科學習小組探測之方法，並將其整合與應用於以本體為基之線上學習系統，於此，該研究將一個成員視為一單位，並針對成員歷史發文，以及成員自定義個人檔案資料運用粒子群聚演算法（Particle Swarm Optimization；PSO）進行階段性個體成員之分群，當中，該研究將分群目標分別訂定為「個體差異最大化」與「群集差異最小化」，藉此將因個體專業性差異最大化，以及小組專業性差異最小化之因素，形塑互助及互補之條件，以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)** 針對教師參與網路社群之行為進行深入探討，並以量化方式衡量教師對知識分享之實踐及社交關係之維持，對個人專業性與整體社群成長所帶來之影響，於此，該研究由 SCTNet 網路社群中，蒐集並分析使用者資料，並發現教師分享知識之意願，可溯根於回答問題之熱絡度、知識品質衡量能力，以及對於分享正確知識之信心，而上述因子係由教師對人際關係參與之活動力形塑之，因此，為了於網際網路中確切提升教師專業能力，該研究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。

## **(B) 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，首先，該研究由「WebMD」醫療型知識分享虛擬社群中以自動化方式蒐集大量醫療討論文章，次之，則根透過對詞袋、語言查詢詞數（Linguistic Inquiry Word Count；LIWC），以及討論串長度等概念與特徵之運用，整合  $X^2$  統計法則之應用，藉以降低數據不平衡所致分類效能之影響，以完成單純貝式分類器（Naïve Bayes Classifier）之訓練，最終，即應用已完善訓練之分類器，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，於此，該研究藉由 362 份有效問卷之蒐集、分析與討論，探討得虛擬社群若具備相當之知識、系統及服務品質，且社群界面與功能符合易用性及有用性之原則，則社群成員間自發性分享知識之熱絡度將得以維持，而如此探討得之結果，將可作為醫療組織建置知識分享虛擬社群，以及維持社群發展性之參考依據。

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，當中，該研究亦以年齡、教育程度與收入作為考量之因子，作為運算時之調節變數，以分析得更具代表性之結果，而藉由對最終分析結果之深入探討，該研究認為資訊取得之便利性，以及社群功能使用之容易性乃服務提供者須克服之關鍵，而若同時對數位落差隔所致溝通問題進行改善，將可有效提升使用者使用社交網路型雲端服務之意圖。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，於此，該研究蒐集虛擬社群及成員個人等量化數據，並透過對零膨脹卜瓦松迴歸 (Zero-Inflated Poisson Regression) 模式之應用，計算並探討得健康資訊虛擬社群得以持續成長，首要條件係先行鞏固社交網路之活絡性，而活絡性維繫之方式，將可面向成員自發性分享意願層面進行著手，針對付費與免費成員設計應類型，獎勵成員自發性分享之機制，促使成員自發性分享健康資訊，進而為健康資訊虛擬社群帶來整體性之持續成長。

### (C) 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，首先，該研究以自動化方式蒐集網際網路中各虛擬社群成員所發布之訊息，並以半結構化方式儲存資料，次之，則針對蒐集得之資料進行內文解析，以萃取得各資料中之關鍵主題，接著，即以關鍵主題相對於全部資料出現之頻率，計算得各主題之重要性比重，最後，以時間區段作為基準 (如一週) 持續進行資料自動化之蒐集與分析，藉以分析得公眾議題發展與變化之趨勢，而解析得主題趨勢分布之走向，將可為政府及企業提供未來策略規劃之相關參考依據。

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，首先，該研究根據成員對討論串之標籤，劃分得各討論串歸屬狀況，以彙整並建置為相關主題知識，其次，該研究透過對 Yahoo 斷詞程式之應用，並以自行建置之停用詞列表，進行候選討論串關鍵詞彙之篩選，接著，以 TF-IDF 演算法進行詞彙權重之計算，以萃取得各討論串中之關鍵詞彙，爾後，則針對各社群成員之歷史知識收藏與推薦行為，進行成員間交叉鏈結關係與可信度之解析，最後，整合於所有數據，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，首先，該研究針對社群中之文章發布者，以瀏覽文章人數及回應速度為基準，鑑別得具影響力之發布者，並根據其發布文章之屬性、特徵、響應頻率、引用資料、推薦狀況及其興趣特徵定義群集類別，爾後，運用 Hall 等人 (2009) 發展之 WEKA 資料探勘軟體，探索各成員間之關聯規則，並解析成員與群集類別間之關聯性，以建置消費者社交網路，最後，則根據消費者社交網路中之關聯性，於成員搜尋文章時，提供符合興趣之文章及產品予其參考，除縮短成員取得所需資訊時間外，更拓展企業推薦消費者所需產品之管道。

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，首先，該研究建構點對點虛擬社群環境，並將資料儲存於雲端資料庫中，次之，蒐集既有緊急事件領域之知識本體，並以 W3C 緊急資訊互動架構進行語意概念本體化之定義，接著，運用 SPARQL 作為標準化查詢語言，藉以匹配查詢語句至最接近語意概念之結果，最後，該研究整合所有功能並建構以 Web 為基之管理系統，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。

綜上所述，就虛擬社群知識應用領域之議題而言，於教育領域之層面中，主要以統計技術與量化概念整合之方式，針對提升教育績效方法進行全面性量化資料之蒐集與統計，藉以從中討論得相關教育績效提升之策略；於醫療領域之層面中，大多透過量化之方式針對醫療型虛擬社群成功之因素，以及成員分享知識之動機進行深入之探討，提供相關討論得之影響因子予線上醫療服務提供者參考，方以作為管理層面中管理措施之擬訂與執行；於其他領域之層面中，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。

綜合於上述之探討，本研究亦根據「虛擬社群知識應用領域探討」議題，彙整相關探討之結論，如表 2.3 所示。

表 2.3、虛擬社群知識應用領域探討之文獻彙整表

領域	過去研究	研究方法	探討對象與目的	結論或研究貢獻
教育	Tobarra 等人 (2014)	整合詞袋、TF-IDF 及標題網路方法，進行量化層面之探討。	線上學習學生學習狀況與問題	提供教育單位參考，使其瞭解學生討論議題之趨勢，以作為課程內容調整之依據。
	Tseng 與 Kuo (2014)	知識分享行為量化之分析	教師對知識分享之實踐	教師專業能力之養成雖須重視，但對於自發性參與社群活動力亦須維持。
	Dascalu 等人 (2014)	粒子群聚演算法	多學科學習小組之成立	聚類得多個學習小組，並具備跨領域討論與學習能力，且符合互助及互補條件。
醫療	Huh 等人 (2013)	運用單純貝式分類器發展醫療討論文章分類方法	醫療領域專家對線上社群之管理	呈現分類資訊予領域專家參考，以改善巨量討論資訊增加所致回答品質之影響。
	Alali 與 Salim (2013)	整合科技接受及資訊系統成功模型，進行量化層面之探討。	提升醫療領域知識分享行為因素	當社群具備相當知識、系統及服務品質，且功能符合易用及有用性原則，自發性分享知識熱絡度將得以維持。
	Ba 與 Wang (2013)	運用零膨脹卜瓦松回歸模式，進行量化層面之探討。	健康資訊虛擬社群中，成員自發性分享之動機。	藉由獎勵機制之應用，將提升成員自發性之分享，進而鞏固社交網路之活躍性。
	Sato 與 Costa-i-Font (2013)	以機率單位模型，進行量化層面之探討。	社交網路型雲端醫療服務之成功因素	資訊取得之便利性及社群功能使用之容易性乃須克服之關鍵，改善數位落差隔閡，將有效提升成員使用意圖。
其他	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	探測公眾議題成長趨勢	動態分析公眾議題發展與變化之趨勢，以提供政府及企業未來策略規劃參考依據。
	Liu 等人 (2014)	發展問答主題知識推薦模式	改善資訊超載問題	根據成員與群體偏向性關係，推薦符合偏向主題知識，以降低擷取知識時間。
	Wang 等人 (2013)	發展消費者興趣偏好探索與搜尋推薦模式與系統	社交網路消費者關係鏈結之應用	提供符合興趣偏向性之文章及產品予成員參考，縮短資訊取得時間，並拓展企業推薦消費者所需產品之管道。
	Li 等人 (2014)	建置以虛擬社群概念為基災害資訊協同共享與整合系統	緊急事件雲端即時化整合之管理	提供使用者共享、討論與檢索災害資訊，以雲端進行跨地區策略協同規劃，降低資訊傳遞與統合之時間。

## 2.3 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。

### 2.3.1 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) 提出一套以主題分割為基中文線上論壇之答案探測模式，於主題分割議題中，該研究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，該研究擴展回歸預測演算法，使系統能通用於多重策略問答模式中，以增添合併答案時之準確性，另一方面，於答案排名特徵函數中，該研究以「地名辭典」與「WordNet」、「維基百科」與「中央情報局世界概況」作為語料庫，以用於分析答案中詞彙間之相似度距離，最終，整合於所有功能，該研究即建構一套適用於各種問答環境與社群之答案排名及合併系統，以提升問答契合程度與問答匹配之準確性。

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) 提出一套運用結構支援向量機，萃取線上論壇中內文與解答資訊之模式，該研究藉其所開發之「通用型圖形表示法」，以該模組完整地表達包含「連續的句子」、「句子的上下文」，以及「多重標籤之獨立句子」等三種不同類型句子間之相互關係，以釐清多個提問間之交互作用，另一方面，該研究透過對「結構支援向量機」之應用，以模式化之方式萃取線上論壇中之資訊，不僅於此，該研究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，首先，該研究蒐集大量提問與回答內容，透過資料預處理以結構化方式建置訓練資料庫，接著，即以 MDE 演算法，針對問答內容進行最小微型化之分類與表達，藉此萃取得候選關鍵提問詞彙，爾後，針對各詞彙進行關聯程度之分析，並同時以關聯度為依據進行分類，最後，則根據候選關鍵詞彙與提問及類別契合程度，萃取得關鍵詞彙，藉以改善既有關鍵字萃取法可能導致資訊偏差之問題。

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、情緒與需求之模式，首先，該研究先行對文章之內容進行「事件」與「需求」之人工篩選，次之，即進行事件與需求之得分評量，以獲得發文者於文章中欲表達之「核心事件(主題)」與「重點需求(目的)」，接著，該研究即透過以情緒為基之事件需求表達模組，同時藉由七大情緒類別之判定，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，首先，該研究運用其所提出之「水平權重樹匹配演算法 (Level-weighted Tree Matching Algorithm)」，針對各別討論串中之所有評論，進行相似度之判定，以消除影響計算結果之雜訊，並同時建置該討論串中評論之樹狀關係；次之，為更進一步消除雜訊，該研究亦透過對「文檔導向模型樹 (Document Object Model Tree)」中節點一致性程度之比較，以萃取出討論串中符合最小子樹之評論內容；最後，針對於非雜訊之評論而言，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，首先，該研究運用詞類標記與語句分塊器，進行詞類標註及語句斷詞等資料預處理動作，次之，即以 TF-IDF

權重衡量演算法，衡量詞彙間及語句間之關係權重，以解析得問答間內文之契合程度，接著，則根據文章內之同義、反義及派生詞間之相互關係進行解析，以取得不同文章中詞彙間及語句間之語意關係距離，最終，經正規化後以取得整合型問答契合度數據，而此整合型分析結果，將有助提升既有資訊檢索系統之精確度。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，首先，該研究由維基百科蒐集文章，並以 TF-IDF 權重衡量法解析詞袋權重，以計算詞彙及概念於文章中之關鍵程度，次之，即運用「跨語言鏈結程度解析模組」，針對維基百科中不同種語言之外部連結，分別進行以文章與類別鏈結程度之推導，接著，經正規化後將取得跨語言詞彙及概念於文章中之關鍵程度，最後，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，以提升專業型虛擬社群中知識之利用率，進而促進虛擬社群成員間相互交流知識之行為。

而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，進行相關具語意特性之解析，並根據解析結果進行相對應問題屬性之定義與分類，首先，在語句本體層面，透過「型態分析」及「依存分析」等方法建構語句本體，次之，於語意本體層面，則藉由「詞頻」、「共現」、「變遷速率」等方式，計算語句之語意關聯程度，最終，該研究即於推論本體之層面中，進行語句項目關係之建構（定義），以獲致客戶提問內容所屬之類別，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。

綜合於上述之探討，本研究亦根據「問答核心資訊萃取」議題，彙整相關使用技術與分析方式，如表 2.4 所示。

表 2.4、問答核心資訊萃取之文獻彙整表

過去研究	使用技術或方法	分析方式與過程
Wang 等人 (2013)	主題分割；情緒指標界定；答案探測；支持向量機	經主題分割後進行正、負面情緒指標之界定，並以正面情緒之回答為主，藉支持向量機訓練答案語料庫，以判定最佳解答。
Ko 等人 (2010)	擴展型回歸預測法；答案排名；詞彙相似度計算	於回歸預測後以地名辭典與 WordNet 為答案排名，並分析答案之相似度，以作為問答匹配之指標。
Moreo 等人 (2012)	最小化微型表達法	以最小化微型表達法，對問答內容進行表達，萃取得候選關鍵提問詞彙，於分析詞彙關聯程度後，對提問進行分類，再分別計算得候選關鍵詞彙與提問及類別之契合程度。
Cao 等人 (2011)	通用型圖形表示法；結構支援向量機	以通用型圖形表示法表達句子間之相互關係，並以改善後之結構支援向量機萃取各種不同形式線上論壇中之內文及解答。
Tung 與 Lu (2010)	內文、詞性與標題核心得分之計算；情緒類別判定	於萃取核心事件與重點需求並進行評分後，藉情緒對事件與需求相關性之分析，以取得發文者於文章中隱含之需求及情緒。
Liu 等人 (2011)	水平權重樹匹配演算法；文檔導向模型樹；直接萃取法；封裝為基萃取法	以水平權重樹匹配法判定評論相似度、消除雜訊與建置樹狀關係，再藉文檔導向模型樹、直接與封裝為基萃取法萃取得最具代表性之評論。
Ferrández (2011)	詞類標記與語句分塊器；TF-IDF；語意解析	以詞類標記與語句分塊器預處理資料，並以 TF-IDF 解析問答間內文契度，於根據詞性取得語意關係後，經正規化以獲致內文及語意之契合程度。
Sorg 與 Cimiano (2012)	TF-IDF；跨語言鏈結程度解析；維度推導；概念關係解析	以 TF-IDF 計算概念之關鍵程度，並藉跨語言鏈結程度、維度推導、概念關係、內文相關性與概念空間之解析，於整合後以獲致跨語言文章間之整體相關程度。
Chen 與 Chen (2012)	萃取經驗知識；驗證知識品質；推論領域類別	透過萃取經驗知識、驗證知識品質、建構知識資料與推論並彙整知識類別，進行經驗知識之統整，並提供知識之彙整索引。
Iwashita 等人 (2011)	建構語句本體；語意解析；語句本體之擴展與定義	透過型態及依存分析以建構語句本體，並藉詞頻、共現、變遷速率等方式，計算語意關聯程度，最後根據上述數據，進行語句本體之擴展與類別之定義。

## (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，於檢驗結果之解讀中，上述分類法之正確性皆達 80% 以上，亦表示此些分類法適用於中文語句之解析，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，當中，由於 Google 搜尋平台已具備龐大且多樣之資料量，故該研究所提之方法，可直接運用 Google 搜尋引擎中自帶之索引資料庫，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合理性與精確度，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，因此，該研究即結合語法與語意之概念，建構一套整合語法及語意資訊之語句相似度分析模式，藉由兩種不同面向角度之考量，以分析得更具全面性之語句相似程度，當中，語意之資訊係藉詞彙語料庫所取得，而語法之資訊則透過深層分析多個語句後歸納並彙整所得，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，首先，該研究運用語料庫之詞彙語意資訊，根據短文本之內容建置相關詞彙相似矩陣，接著，即藉詞彙相似矩陣計算詞彙與文本之相似程度，次之，將上述所得運算結果彙整並建置為關係矩陣，最後，則根據關係矩陣中之數據建構文本向量，並藉建置所得之文本向量與其他文本進行比對，以分析得文本間之相似程度，當中，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，當中，該研究運用不同文章表示法，以解析各表示法產生之效能，於選定效能最佳之文章表示方法後，則將語句關係套用於以 WordNet 為基之分群演算法進行比較，最終之實驗結果顯示，於分群演算法中所選定之語意關係類型，其所獲致之效能由高至低依序為「上位關係」、「下位關係」、「部分—整體關係」與「整體—部分關係」，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。再者，Zhang 等人 (2010) 針對既有較具代表性之相關研究，包含「以頻繁詞彙序列為基之分群法」、「以最大頻繁序列為基之文檔分群法」、「基於頻繁項目之分群法」，以及「基於頻繁項目集之階層式分群法」進行深入之探討與解析，整合各分群法之特性發展一套更完善之「最大量擷取法 (Maximum Capturing)」，此外，該研究採用「共同頻繁項目集數量之分析 (MC-1)」、「共同頻繁項目集總權重之分析 (MC-2)」及「非對稱性二進制相似度之分析 (MC-3)」等三種不同相似度演算法，以相同資料及上述各分群法，建置文檔之群集並指派群及名稱，以評估分群之績效，於結果中顯示，該研究所提之分群法，將具備最高之績效，當中，MC-1 於中文文檔之分群結果中具備最高之正確性。

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，評比之方式主要以社交網路之概念為基礎，根據虛擬社群中成員之個人檔案，分別運用上述三種分群演算法進行分群，以評估各別成效，首先，該研究建構成員檔案為基之語意框架，當中包含成員檔案之模型化與建置社交網路指標等步驟，待完成上述社交網路分析之預處理步驟後，該研究即分別運用於上述分群法中，並針對分群時之複雜性、分群後群集關聯正確性，以及新成員加入時群集重組效率等項目進行評估，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，該研究將資料劃分為文章、段落、句子、短語與詞彙等五種類型，並以模式化方式，對資料進行標準化語句分析樹之節點轉換，次之，則以「最鄰近學習法」針對節點間之關聯程度進行階層式之劃分，最後，整合上述數據並轉換為邏輯樹狀圖，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，首先，藉該研究以文章類別標籤為基礎，從中萃取得文章概念，並以概念空間模型進行詞袋之表示，以改善詞袋集合中存在過多詞彙之問題，次之，該研究整合「導向式」與「分裂式」推導之概念重要性評估法則，衡量各概念於文章內所佔之權重，以取得文章之核心概念；最終，該研究針對文章間之核心概念進行相似度之解析，並於整合類別資料庫中文章之分類資訊後，以歸納得文章所屬之類別，因此，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。

綜上所述，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，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；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，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，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。

綜合於上述之探討，本研究亦根據「語句及詞彙相關性分析」議題，彙整相關使用技術與分析方式，如表 2.5 所示。

表 2.5、語句及詞彙相關性分析之文獻彙整表

過去研究	使用技術或方法	分析方式與過程
Yan 等人 (2009)	單純貝氏；決策樹；最大熵；語意分析	以單純貝氏、決策樹與最大熵等分類法，檢驗中文詞彙語意解析之準確性，並發展整合型模式以計算詞彙間及語句間語意相關性之程度。
Cilibrasi 與 Vitanyi (2007)	Google 搜尋引擎；語意解析	透過 Google 搜尋引擎，統計兩詞彙中之各獨立詞彙，以及同時包含兩詞彙搜尋所得之結果數，並分析其標準化動態邏輯距離，以視為語意關係距離
Chen 與 Lin (2011)	Google 搜尋引擎；標準化 Google 距離；語意解析；跳躍路由；詞彙關係網路	修正原有之 NGD 演算法，以建構 Google 核心距離演算法，並提出跳躍路由演算法，並藉上述演算法之兩相結合，建構詞彙關係網路。
Oliva 等人 (2011)	語句相似度解析；語法權重衡量	藉語意及語法資訊之整合，以分析語句相似度，並提供語法權重調整之功能，當中，語法資訊乃深層分析多個語句後歸納並彙整所得。
Wenyin 等人 (2010)	語意解析；相似矩陣；關係矩陣	以語料庫建置相似矩陣，並算詞彙與文本之相似度，再根據運算結果建置關係矩陣，於取得文本向量後以分析得文本間之相似程度。
Zheng 等人 (2009)	評估語句關係對分群法績效之影響	於分析並取得最佳之文章表示方法後，以 WordNet 為基之分群演算法進行比較，最終獲致效能由高至低之語句關係依序為上位關係、下位關係、部分—整體關係與整體—部分關係。
Karamolegkos 等人 (2009)	以社交網路概念評估分群法之績效	運用階層式、K-means 與 Spectral 分群成員個人檔案，以評估複雜性、正確性及重組之效率，並發現 Spectral 之複雜性與正確性擁有最佳績效，而階層式於新成員加入時績效最佳。
Zhang 等人 (2010)	評估分群法績效；發展整合型分群法；頻繁項目集；群集名稱定義	評估相關研究並擇優取之，再藉非對稱性二進制相似度分析之整合，發展一套以頻繁項目集為基之分群演算法，以進行文件之分群與群集名稱之定義。
Galitsky 等人 (2012)	句法分析樹；最鄰近學習法；潛藏語意；相似度排名	對資料進行標準化語句分析樹之節點轉換，並以最鄰近學習法劃分得階層式之關聯程度，再藉邏輯節點間之相互匹配解析得關鍵字之語意，並以此排名相似度。
Li 等人 (2011)	文章分類；語意解析；概念萃取；概念權重衡量	藉類別標籤萃取文章概念，並於衡量概念權重後以取得文章核心概念，接著解析核心概念相似度並整合資料庫分類資訊，以歸納得文章所屬類別。

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。

### 2.3.2 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，首先，該研究透過對向量空間模型之應用，根據使用者撰寫與閱讀過之文章建置集合向量，其次，藉由各集合向量間相似度之計算，解析得文章間之關聯程度，並以階層式分群法進行分群，同時藉華德之聯動準則 (Ward's Linkage Criterion) 進行群集間樹狀圖及關係程度之呈現，接著，運用 K 值進行群集相關度檢測，並以檢測過程中之數據建置另一維度集合向量，最終，以社群實踐探測法，將領域相關度較高之使用者予以連結並供知識需求者參考，進而降低判斷知識可用性之時間成本。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，以更為全面之方式，推薦予使用者最為相似之其他使用者、資源與社交網路，進而促進使用者間之互動，以激盪出更具品質及專業程度之知識內容。

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，首先，該研究透過對 PageRank 演算法之應用，發展一套社交網路生成模組，以根據使用者間評價關聯性建置類別關係連結，次之，即以關係連結為基礎，動態更新評價資訊與社交網路，以進行

使用者類別專業性品質之實時衡量，最終，該研究即以使用者類別關係網路呈現分析得之數據，並同時以領域類別為基礎註記具備高專業度，以及低專業度使用者，供所有成員參考並作為知識採納之依據，除提升獲取具品質領域知識可能性外，亦提供知識管理之指標。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，以本體化方式建構個人檔案，以及呈現各階層間模糊正確率，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。

為解決個人檔案所衍伸之時效性問題，Liang 等人 (2008) 提出以語意擴展方法為基個人化領域知識推薦模式，首先，該研究蒐集使用者於社群中評價文章之歷史資訊，次之，根據歷史資訊所屬文章進行關鍵概念萃取，以建置目標使用者領域性概念專業得分比重，接著，根據專業性得分及其所對應之詞彙，以語意擴展網路方式建置概念間語意相關性樹狀網路，最終，整合上述語意網路之數據，該研究除透過模式化之方式建構使用者之個人檔案外，亦推薦與其相符合之專業性領域文章，透過模式之運行，該研究不僅解決使用者自建檔案之問題，更降低使用者檢索興趣文章時所需耗費之時間成本。

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，首先，該機制結合向量空間模型與 TF-IDF 演算法對討論串之語意相似度進行解析，次之，針對成員之個人檔案及其歷史發文進行解析，以評估並建置該成員知識專業程度，接著，即整合與社交網路理論相關之「社會親密與知名度測量方法」，以及「以社交網路為基之馬可夫鏈分析模組」，以分析得各論壇成員間於社群中之關係，於最終，經由整合上述相關之數據後，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，此外，該研究亦藉 LDA 及詞袋之相互結合，發展以提問為導向整合性排名方法論，透過對「Yahoo! Answers」問

答資料之蒐集與整合，藉以推薦符合於成員偏好之領域知識討論串。

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，首先，該研究將使用者查詢與檢索之行為視為專精知識之偏向，並以此建構階層式使用者個人檔案，接著，考量使用者於查詢習慣之差異，藉由對使用者查詢時之關鍵詞、總體頻率最小值、總體頻率最大值與最大頻率路徑值等數據之蒐集，以進行領域偏向影響因子之解析，藉此建構更具參考性之使用者個人檔案，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，首先，該研究蒐集使用者個人檔案及歷史使用紀錄，並分別從中萃取得使用者之代表特徵及知識本體之調適因子，藉此建構結構化使用者個人檔案與個人化之領域知識本體，以分別應用於後續知識之推薦及專業偏向之解析中，次之，該研究藉由對使用者個人化領域知識本體，以及領域知識庫中知識本體契合度之解析，以分析其專業偏向，接折，即根據結構化使用者個人檔案中所屬之特徵因子，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，首先，該研究蒐集成員個人檔案，並以基於本體個人化模式對資料執行預處理，以建置結構化個人檔案集合向量，次之，以個人化語意內容檢索模組針對知識資料庫，以及個人檔案集合向量進行語意權重之衡量，以取得個人檔案中資訊關鍵程度之排序，最後，則進行多層次成員間興趣內容之分群，取得興趣與知識領域背景相似之成員群集，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。

綜合於上述之探討，本研究亦根據「成員專業領域偏向分析」議題，彙整相關使用技術與分析方式，如表 2.6 所示。

表 2.6、成員專業領域偏向分析之文獻彙整表

過去研究	使用技術或方法	分析方式與過程
Vertommen 等人 (2008)	向量空間模型；分群法；華德聯動準則；群集檢測；社群實踐探測法	根據使用者撰寫與閱讀之文章建置集合向量並解析關聯程度，接著以階層式分群法分群，並以華德聯動準則呈現群集間之樹狀關係，再運用 K 值進行群集相關度之檢測。
Meo 等人 (2011)	超圖；使用者資訊相似度解析；自動化個人檔案之建置；使用者推薦	將使用者、資源與社交網路視為節點以超圖進行表示，並於解析會員資格、友誼關係、歷史發文與評價紀錄後，藉建置通用型之個人檔案，以進行相似使用者之推薦。
Caverlee 等人 (2010)	PageRank；自動社交網路生成；成員類別專業品質衡量	改良 PageRank 演算法，以發展社交網路生成模組，並以成員關係連結為基礎進行動態實時之分析，藉此行量成員領域類別之專業性。
Han 與 Chen (2009)	模糊分群法；模糊 C 均值；PHR；OSGP；模擬退火；本體論	蒐集使用者行為記錄以解析專業與興趣偏向，並於聚類相似使用者後以 PHR 與 OSGP 整合型演算法解析分群結果，再以本體化之方式建構個人檔案，並呈現各階層間之模糊正確率。
Liang 等人 (2008)	概念萃取；語意擴展網路；語意解析；知識推薦	蒐集使用者之歷史資訊，並於萃取關鍵概念後分析領域專業得分比重，接著以語意擴展網路建置樹狀網路及個人檔案，再根據個人檔案推薦與其相符合之專業性領域文章。
Li 等人 (2012)	向量空間模型；TF-IDF；社會親密與知名度測量；馬可夫鏈模型	於解析討論串之相似度後，即藉分析使用者個人檔案與歷史發文以評估知識專業程度，接著以社交網路相關理論分析使用者間之關係，再以此形式推薦其感興趣或所需之知識。
Ni 等人 (2012)	潛在狄利克雷分配模型；知識推薦	改良潛在狄利克雷分配模型，以解決因詞彙共現數過少及未考量知識領域，所導致主題萃取可信度不足之問題，並以提問為導向排名方法論推薦符合於使用者偏好之討論串。
Chen 等人 (2013)	自動化個人檔案之建置；使用者專業偏向解析；知識本體；知識推薦	蒐集使用者個人檔案及歷史紀錄，於萃取特徵與調適因子後以建構個人檔案與知識本體，接著解析領域知識庫中知識本體之契合度，取得使用者專業偏向並進行知識之推薦。
Maleszka 等人 (2013)	自動化個人檔案之建置；使用者專業偏向解析；使用者分群；使用者推薦	蒐集使用者於社群中查詢與檢索之行為，以此建構階層式之使用者個人檔案，同時予以分群，以取得具備相似專業背景之使用者群集，藉以推薦適切之協同合作成員。
Cantador 與 Castells (2011)	本體論；個人檔案結構化；語意權重衡量；個人檔案資訊關鍵程度解析；使用者分群	蒐集成員個人檔案後建置結構化個人檔案集合向量，並於衡量語意權重後解析資訊關鍵程度，最後進行多層次成員間興趣內容之分群，以取得興趣與知識領域背景相似之成員群集。

## (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，首先，於去除發文內容中之停用詞後，該研究藉向量空間模型及 TF-IDF 之結合，以計算分享者間專業知識相關性之得分，次之，透過社交網路概念與 PageRank 演算法之結合，以標題感知方式，分析得分享者於社群中之專業代表性程度，最終，整合

所有數據後，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) 結合評價信任度與指向關聯性（即使用者間互動之狀況）等議題，提出以個體及群組可信度為基協同過濾文章推薦模式，首先，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，最終，透過上述數據之整合以計算得成員專業性得分，藉此建構相關之領域專家列表，以推薦相關之專家列表供其參考，進而提升知識需求者取得所需知識之可能性。

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，故該研究以語意為基礎並透過廣義標題時效模式，發展專家尋找方法，首先，該研究以向量方式針對文章標題進行結構化定義，其次，以潛在狄利克雷分配模型進行數據之儲存與表示，接著，該研究透過標題中廣義詞彙語意與文章發布時間之結合，以分析文章與時效間之關聯性，於最終，即整合分析得之時效關聯性，並根據標題內容與語意詞彙庫間之相關性，分析並計算得知知識貢獻者之專業性，於進行專業度之排序後，提供予使用者進行自定義查詢，以呈現符合時效條件專家列表，進而提升專家尋找之精確性。而 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，於此，該研究蒐集成員歷史發文及個人檔案等資訊，並以標準化方式建置專業度指標數據，爾後，運用鏈結分析之方法，包含 PageRank 及 Hyperlink-Induced Topic Search；HITS，解析得目標成員專業度指標於社交網路中與其他成員之鏈結程度，接著，即針對成員間互動行為進行權重衡量，以增添社交網路鏈結程度之精確性，最終，整合並經正規化專業度數據，茲以排名並視其為最終結果。

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

為發展社交網路意見領袖衡量之混合型系統，該研究借鑑於過去研究探討得社交網路意見領袖生成影響因子，進行模式化數據之演算，於專業性與創新性中，依序以潛在狄利克雷分配、TF-IDF 與餘弦定理進行指標之衡量；於影響力及活動性中，以發文覆蓋與回覆率之分項指標進行評估；執著性與集中性中，則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，該系統主要藉前饋式類神經網路，尋找適切專家以進行相關知識評估與更新，而除了知識提供者與知識內容間匹配程度之解析外，該研究亦運用公平指數解析知識提供者專業程度，以界定其貢獻知識後知識品質權重衡量之依據，另一方面，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，首先，該研究將人際關係間之信任劃分為知識型、聲譽型、關係型、相似型與個人型等信任因子，其次，該研究針對社群使用者間互動之行為，衍生之對話及評價等資料進行蒐集，接著，透過「跨產業標準流程」進行資料結構化之處理，並同時針對上述信任因子，定義出使用者知識關係程度、使用者瀏覽行為相似度、評論關聯性、評論品質與社交契合度等定量分析標準，最終，該研究即進行相關量化得分標準之計算，於整合所有數據後，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，發展一套整合型可信度評估模式，首先，該研究將發布者之聲譽以模糊不確定性概念劃分為「信任」、「懷疑」及「不信任」三個等級，以及「個人聲譽」與「公眾聲譽」兩類別，並根據目標發布者與其他使用者間相互之評價，分別計算得「信任」、「懷疑」及「不信任」程度，並藉上述數據之整合，於最終獲致目標發布者可信度，提供知識需求者參考，進而提升其於網路社群中擷取正確知識之效能與效率。

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) 發展以評分數據為基之使用者知識經驗可信任度預測系統，首先，該研究以類別為基礎，根據其被評分結果計算得類別專業度，次之，則針對發布者於之使用行為，包含類別發布文章數、評價數，以及被評價數等數據進行統計，以計算得類別活躍程度，並針對各類別間之關聯性進行解析，以取得其隱含活躍程度，最終，該研究即整合上述數據，且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，除了計算上述類型相似性之程度外，該研究亦針對當中之「類別」，運用 Random Forests 與 Bayesian 分類演算法進行處理，並同時分析各類別間之關聯程度，於最終，運算所得數據之整合，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。

綜上所述，就虛擬社群成員個人資訊解析議題言之，於成員專業領域偏向分析之層面，大多以成員於虛擬社群中之個人檔案，以及歷史使用行為（如歷史發文與瀏覽行為）作為分析之基礎，以判定個體成員之領域偏向，有少數研究考量到成員間存在互動之關係，故融合社會網路與分群之概念於判定成員之領域偏向；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，以取得其於領域中之可信程度，解析之方式多以詞頻為基進行相似度之衡量，以判定成員歷史發文與該領域文章中之相似程度，藉以取得成員知識之可信度。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，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，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，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」。

綜合於上述之探討，本研究亦根據「成員知識可信度分析」議題，彙整相關使用技術與分析方式，如表 2.7 所示。

表 2.7、成員知識可信度分析之文獻彙整表

過去研究	使用技術或方法	分析方式與過程
Wang 等人 (2013)	向量空間模型；TF-IDF；社交網路；PageRank；線性組合；級聯排名；定標策略	結合向量空間模型及 TF-IDF 計算使用者專業知識相關性得分，接著藉社交網路與 PageRank 分析專業性程度，再以線性組合、級聯排名與定標策略進行知識專業程度數據之呈現。
Lai 等人 (2013)	自動化個人檔案之建置；相似度解析與分群；使用者可信度分析；評價資訊可參考度衡量	蒐集使用者評價資訊以建置個人檔案及聚類相似使用者，接著根據參與狀況與群體之共同評價，計算使用者與內容之可信度，以獲致混合型評價資訊可採納之程度。
Liu 等人 (2013)	CKIP；TF-IDF；使用者可信度解析；領域專家之探勘	蒐集使用者之回答內容與評價紀錄，並藉 CKIP 與 TF-IDF 之結合以彙整類別之相似度，接著於解析知識檔案、評價資訊及領域類別後取得專業性得分，並藉此建置領域專家列表。
Daud 等人 (2010)	語意分析；標題時效分析；潛在狄利克雷分配模型	以向量方式針對標題進行定義，再以潛在狄利克雷分配模型進行數據儲存，接著以詞彙語意與文章發布時間之結合，分析並計算得各知識貢獻者之專業性程度。
Schall (2012)	使用者領域專業度分析；PageRank；HITS；社交網路鏈結度	蒐集使用者歷史發文與個人檔案以建置專業度指標，接著藉鏈結分析法，解析得專業度指標之鏈結程度，並於衡量使用者間互動行為之比重後，以取得整體之領域專業度數據。
Li 等人 (2013)	社群意見領袖之探勘；潛在狄利克雷分配；TF-IDF；餘弦定理	透過潛在狄利克雷分配、TF-IDF 與餘弦定理，衡量使用者之專業及創新性，並藉使用者發文與回覆之行為，評估其影響力、活動、執著與集中性，藉以獲致意見領袖得分之數據。
Lykourantzou 等人 (2010)	前饋式類神經網路；公平指數；系統風險管理	以前饋式類神經網路解析使用者與內容間匹配程度，並藉公平指數分析其專業度以衡量知識品質，再以分類、使用者互動、整體風險、推薦與協同合作等指標進行系統風險之管理。
Zolfaghar 與 Aghaie (2012)	上下文分析；使用者信任度評估；跨產業標準流程	蒐集使用者間之對話與歷史評價資訊，以跨產業標準流程進行資料結構化之處理，分析使用者「知識關係程度」、「瀏覽行為相似度」、「評論關聯性」、「評論品質」與「社交契合度」，並於整合所有數據後，取得使用者之可信任程度。
Kim 與 Ahmad (2013)	證據理論；模糊不確定性概念；使用者信任度評估	蒐集使用者間之評價資訊，分別計算個人及公眾聲譽中，信任、懷疑及不信任程度，並於整合所有數據後，以於最終獲致目標知識發布者可信任之程度。
Kim 與 Phalak (2012)	使用者知識經驗可信度預測方法	以類別為基彙整使用者獲得之評分資訊，根據其被評分結果計算類別專業度，再分析使用者歷史使用行為，以計算得類別活躍程度，最終整合所有數據後，將計算得整體可信度。
Korovaiko 與 Thomo (2012)	使用者社交網路潛在可信度評估；Random Forests；Bayesian	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，運用 Random Forests 與 Bayesian 分類演算法分析各類別間關聯程度，以計算得使用者間之信任相關性及隱含信任度。

整體而言，就虛擬社群成員個人資訊解析議題言之，於成員專業領域偏向分析之層面，大多以成員於虛擬社群中之個人檔案，以及歷史使用行為（如歷史發文與瀏覽行為）作為分析之基礎，以判定個體成員之領域偏向，有少數研究考量到成員間存在互動之關係，故融合社會網路與分群之概念於判定成員之領域偏向；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，以取得其於領域中之可信程度，解析之方式多以詞頻為基進行相似度之衡量，以判定成員歷史發文與該領域文章中之相似程度，藉以取得成員知識之可信度。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：Liang 等人（2008）；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」。

## 2.4 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（2.2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，於教育領域層面中，主要以統計技術與量化概念整合之方式，針對提升教育績效方法進行全面性量化資料之蒐集與統計，藉以從中討論得相關教育績效提升之策略；於醫療領域層面中，大多透過量化方式針對醫療型虛擬社群

成功之因素，以及成員分享知識之動機進行深入之探討，提供相關討論得之影響因子予線上醫療服務提供者參考；於其他領域之層面中，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（2.3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，方以進行後續之解析；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之層面，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。

整體而論，本研究發展之「虛擬社群知識群聚度判定模式」分為「虛擬社群問答契合度解析模組」與「虛擬社群問答契合度解析模組」，於前者中，本研究以群眾智慧之影響性為切入點，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模組，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模組，最終，整合於上述兩模組，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，提供相關數據予知識擷取者，以減少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，活絡虛擬社群之利用率，進而催化虛擬社群持續性之發展。此外，為了歸納並比較本研究與過去研究於技術層面中之異同，於此即針對「虛擬社群問答內文契合狀況解析」與「虛擬社群成員個人資訊解析」兩議題進行差異化之比較，如表2.8所示。

表 2.8、本研究與過去文獻差異彙整表

比較研究	分析對象	應用範圍	自動化處理	自動化語料庫	語句上下文解析	詞彙語意解析	社交網路分析	成員互動行為解析	群眾智慧解析	跨語言解析	優勢
本研究	討論串、虛擬社群使用者	虛擬社群	✓	✓	✓	✓		✓	✓		以內文及使用者層面分析討論串可群聚正確知識之程度
Wang等人(2013)	討論串	問答平台				✓		✓	✓		以正面情緒指標判定最佳解答
Ko等人(2010)	問答內容	問答平台	✓			✓					以回歸預測答案成為最佳解之概率
Moreo等人(2012)	問答內容	問答平台	✓			✓					計算關鍵提問詞彙與類別之契合程度
Cao等人(2011)	討論串	虛擬社群	✓	✓	✓						釐清提問間之交互作用，並萃取各種不同形式線上論壇中之內文及解答
Tung與Lu(2010)	部落格文章	部落格		✓	✓						分析發文者於文章中隱含之需求及情緒
Liu等人(2011)	討論串	虛擬社群	✓	✓	✓						萃取最具代表性之評論
Ferrández(2011)	問答內容	問答平台	✓	✓	✓	✓					具備內文及語意問答契合度之分析方法
Sorg與Cimiano(2012)	網頁知識文件	特定領域	✓	✓	✓					✓	分析跨語言文章間之整體相關程度
Chen與Chen(2012)	討論串	特定領域		✓	✓				✓		進行經驗知識之統整，並提供知識之彙整索引
Iwashita等人(2011)	客戶提問	特定領域	✓	✓	✓	✓					語句本體之擴展與類別之定義
Yan等人(2009)	網頁知識文件	特定領域	✓			✓					發展整合型之中文語意相關性分析模式

Cilibrasi 與Vitanyi (2007)	詞彙	一般領域	✓	✓		✓			✓	分析兩詞彙之標準化動態邏輯（語意） 距離
Chen與 Lin (2011)	詞彙	一般領域	✓	✓		✓			✓	建構詞彙關係網路
Oliva等 人(2011)	網頁知識 文件	一般領域	✓		✓	✓				分析語句相似度，並提供語法權重調整 之功能
Wenyin 等人 (2010)	網頁知識 文件	一般領域	✓			✓				分析文本間之相似程度
Zheng等 人(2009)	分群演算 法	特定領域	✓	✓		✓				評估語句關係對分群演算法績效之影 響
Karamole gkos等人 (2009)	分群演算 法	特定領域	✓	✓			✓	✓		以社交網路概念評估分群演算法之績 效
Zhang等 人(2010)	分群演算 法	特定領域	✓	✓	✓					評估分群演算法之績效，並發展整合型 分群演算法
Galitsky 等人 (2012)	網頁知識 文件	一般領域	✓	✓	✓	✓				分析不可測關鍵字之語意
Li等人 (2011)	網頁知識 文件	一般領域	✓	✓		✓				分析文章所屬類別
Vertomm en等人 (2008)	虛擬社群 使用者	特定領域	✓	✓	✓		✓	✓		群集間樹狀關係之呈現與群集相關度 之檢測
Meo等人 (2011)	虛擬社群 使用者	特定領域	✓	✓			✓	✓		使用者、資源與社交網路之超圖表示 法；通用型個人檔案之建置
Caverlee 等人 (2010)	虛擬社群 使用者	虛擬社群	✓				✓	✓		以成員關係連結為基礎進行動態實時 之分析
Han 與 Chen (2009)	虛擬社群 使用者	虛擬社群	✓	✓	✓					整合型模糊分群法之發展；以本體化之 方式建構個人檔案

Liang等人(2008)	虛擬社群使用者	虛擬社群	✓	✓		✓					以語意擴展網路建置樹狀網路及個人檔案，並推薦相關之專業性領域文章
Li等人(2012)	虛擬社群使用者	特定領域	✓	✓	✓		✓	✓			以社交網路相關理論分析使用者間之關係
Ni等人(2012)	討論串、虛擬社群使用者	虛擬社群	✓	✓		✓					以提問為導向排名方法論推薦符合於使用者偏好之討論串
Chen等人(2013)	虛擬社群使用者	虛擬社群	✓			✓					萃取使用者特徵與知識本體之調適因子，藉此建構個人檔案與領域知識本體
Maleszka等人(2013)	虛擬社群使用者	虛擬社群	✓	✓					✓		階層式使用者個人檔案之建置與分群
Cantador與Castells(2011)	虛擬社群使用者	虛擬社群	✓	✓					✓		多層次成員間興趣內容之分群
Wang等人(2013)	虛擬社群使用者	特定領域	✓	✓	✓		✓		✓		以線性組合、級聯排名與定標策略進行知識專業程度數據之呈現
Lai等人(2013)	虛擬社群使用者	虛擬社群	✓	✓				✓	✓		取得混合型評價資訊可採納之程度
Liu等人(2013)	虛擬社群使用者	虛擬社群	✓	✓	✓						建置領域專家列表
Daud等人(2010)	虛擬社群使用者	虛擬社群	✓			✓					以詞彙語意與文章發布時間之結合，計算各知識貢獻者之專業性程度
Schall(2012)	虛擬社群使用者	特定領域	✓				✓	✓			藉鏈結分析法，解析得專業度指標之鏈結程度
Li等人(2013)	虛擬社群使用者	虛擬社群	✓	✓	✓	✓					分析使用者意見領袖得分之數據
Lykourantzou等人(2010)	虛擬社群使用者	特定領域	✓	✓				✓			藉品質指標之自動化衡量，以達系統風險管理之目的
Zolfaghar與Aghaie(2012)	虛擬社群使用者	虛擬社群	✓				✓	✓			以跨產業標準流程進行資料結構化處理，並計算使用者之可信任程度

Kim與 Ahmad (2013)	虛擬社群使用者	虛擬社群	✓					✓			計算信任、懷疑及不信任程度，並整合為目標知識發布者可信任之程度
Kim與 Phalak (2012)	虛擬社群使用者	虛擬社群	✓					✓			根據使用者歷史使用行為，計算類別活躍程度，將分析得整體可信度
Korovaik o 與 Thomo (2012)	虛擬社群使用者	虛擬社群	✓				✓	✓			分析各類別間關聯程度，以計算得使用者間之信任相關性及隱含信任度



### 第三章、虛擬社群知識群聚度判定模式

本研究建構之「虛擬社群知識群聚度判定模式」，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，乃以虛擬社群成員之「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，當中群集間相似程度之解析，以取得相關領域性群集間之內文與語意關係距離，此外，根據上述分析目標虛擬成員所得之數據，以及其他同領域虛擬社群成員所得之數據，藉「正規化」之方式取得相互比較之基準點，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。綜上所述，本研究所提「虛擬社群知識群聚度判定模式」之整體運作情況，如圖 3.1 所示。

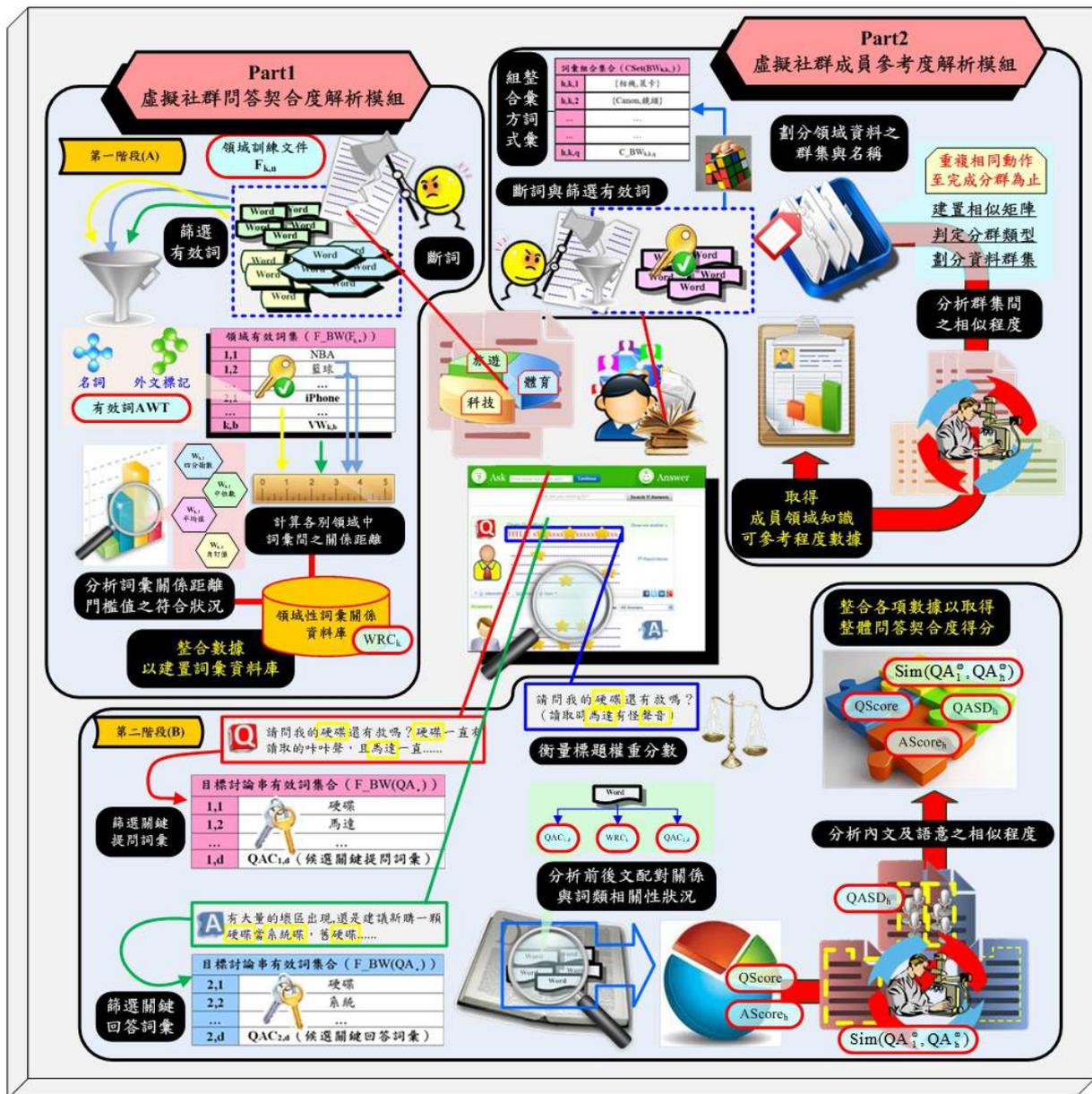


圖 3.1、虛擬社群知識群聚度判定模式之流程架構圖

### 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，同時藉本研究制定之「四分位數」、「中位數」、「平均值」，以及「自訂門檻值」等四種篩選方式，以更為精確之方式挑選得實具高關聯程度之領域詞彙組合，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 3.2 所示。



圖 3.2、領域性關鍵詞彙集合之建立運作示意圖

說明本推論模式之第一階段(A)前，將第一階段(A)採用之符號定義如下：

$F_{k,n}$	第 k 類之第 n 筆訓練文件
$BW(F_{k,\bullet})$	第 k 類訓練文件中，所有訓練文件經斷詞後之結果
$F\_BW(F_{k,\bullet})$	第 k 類訓練文件中，所有訓練文件經斷詞及有效詞篩選後之集合，當中，有效詞係指具備實質意義之詞彙(如「手機」與「iPhone」等名詞及外文標記皆為有效詞；「的」及「你」等虛詞為無效詞)
AWT	詞類屬於名詞或外文標記之詞彙，當中名詞包含普通名詞、專有名稱、地方詞與位置詞，外文標記不包含單純之數值(如 1、2、3...等)，此外，若外文標記之前或後方存在數字，則此外文標記與數字將組合為一個有效詞(如 600D、S5200 與 F1.4D)
$RD(AVW_{k,i}, BVW_{k,j})$	第 k 類訓練文件中，主要有效詞 i 與次要有效詞 j 之關係距離
$AVW_{k,i}$	第 k 類訓練文件中之第 i 筆主要有效詞
$BVW_{k,j}$	第 k 類訓練文件中之第 j 筆次要有效詞
$GSH(AVW_{k,i})$	第 k 類訓練文件內，第 i 筆主要有效詞經 Google 搜尋引擎站內搜尋功能搜尋後之結果數
$GSH(BVW_{k,j})$	第 k 類訓練文件內，第 j 筆次要有效詞經 Google 搜尋引擎站內搜尋功能搜尋後之結果數
$GSH(AVW_{k,i}, BVW_{k,j})$	第 k 類訓練文件內，同時包含第 i 筆主要及第 j 筆次要有效詞經 Google 搜尋引擎站內搜尋功能搜尋後之結果數
$TF_k$	第 k 類訓練文件所包含之文件總數
$ARD_{k,p}$	第 k 類訓練文件中，第 p 組領域有效詞集之整體關係距離
$W_{k,t}$	第 k 類訓練文件中，經第 t 種方法篩選後，高關聯程度領域詞彙組關係距離之篩選門檻值，t=1,2,3,4，篩選方法分別為「四分位數」、「中位數」、「平均值」與「自訂門檻值」
Q	第 k 類訓練文件中，主要有效詞 i 與次要有效詞 j 關係距離組合之第三四分位數指標
$N(AVW_{k,\bullet}, BVW_{k,\bullet})$	第 k 類訓練文件中，以兩個有效詞為一組合所定義之詞彙組合總數
$HRD_k$	第 k 類訓練文件中，符合高關聯程度詞彙組合門檻值之詞彙組合集合

$S\_HRD_{k,w}$	第 k 類訓練文件中，符合門檻之詞彙組合，經分割後之第 w 個獨立詞彙，如運動領域中符合門檻之詞彙組合為「(籃球,球場)」 (( $AVW_{k,i}, BVW_{k,j}$ ))，則 $S\_HRD_{k,1}$ 為「籃球」， $S\_HRD_{k,2}$ 為「球場」
$S\_HRD_{k,\bullet}$	第 k 類訓練文件中，已存在之獨立詞彙
$N(HRD_k, S\_HRD_{k,w})$	第 k 類訓練文件中，符合門檻之第 w 個獨立詞彙，於詞彙組合集合 ( $HRD_k$ ) 內出現之次數
$WRC_k$	第 k 類領域性關鍵詞彙集合

### 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合之類型，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  $F_{k,\bullet}$  進行斷詞，以取得類別訓練文件中詞彙之分佈狀況  $BW(F_{k,\bullet})$  (如公式(3.1)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集  $F\_BW(F_{k,\bullet})$  (如公式(3.2)所示)。

$$BW(F_{k,\bullet}) = \{BW_{k,1}, BW_{k,2}, \dots, BW_{k,a}\} \quad (3.1)$$

$$F\_BW(F_{k,\bullet}) = \{VW_{k,1}, VW_{k,2}, \dots, VW_{k,b} \mid VW_{k,b} \in AWT\} \quad (3.2)$$

### 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  $F\_BW(F_{k,\bullet})$  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<sub>k,i</sub>,BVW<sub>k,j</sub>))」, 以及領域訓練文件之總數 TF<sub>k</sub>, 計算得領域有效詞間之關係距離 RD(AVW<sub>k,i</sub>,BVW<sub>k,j</sub>) (如公式(3.3)所示), 其結果整理如表 3.1 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。

$$RD(AVW_{k,i}, BVW_{k,j}) = \frac{\text{Max}\{\log \text{GSH}(AVW_{k,i}), \log \text{GSH}(BVW_{k,j})\} - \log \text{GSH}(AVW_{k,i}, BVW_{k,j})}{\log TF_k - \text{Min}\{\log \text{GSH}(AVW_{k,i}), \log \text{GSH}(BVW_{k,j})\}} \quad (3.3)$$

表 3.1、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> )

### 步驟(A3)—建立領域性關鍵詞彙集合

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<sub>k,i</sub>, 以及次要有效詞 BVW<sub>k,j</sub> 為一組合之領域詞彙組關係距離 RD(AVW<sub>k,i</sub>,BVW<sub>k,j</sub>) (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合。首先, 於開始建置領域性關鍵詞彙集合前, 必須先行計算各有效詞彙組關係距離 RD(AVW<sub>k,i</sub>,BVW<sub>k,j</sub>) 之「整體關係距離」, 亦即將有效詞彙組關係距離, 以及所有關係距離加總後之數據, 進行相除以計算得有效詞彙組之整體關係距離 ARD<sub>k,p</sub> (如公式(3.4)所示)。

$$ARD_{k,p} = \frac{RD(AVW_{k,i}, BVW_{k,j})}{\sum_{\text{all } i,j} RD(AVW_{k,i}, BVW_{k,j})} \quad (3.4)$$

其次，為使領域性關鍵詞彙集合更具準確性，於此步驟中，本研究亦制定「四分位數」、「中位數」、「平均值」，以及「自訂門檻值」等四種篩選方式，藉此以更為精確之方式挑選得實具高關聯程度之領域詞彙組合  $HRD_k$ （如公式(3.5)、公式(3.6)、公式(3.7)與公式(3.8)所示）。相關之篩選法則與詳細作法如下所述：

### 1. 領域詞彙組關係距離四分位數

為使領域性關鍵詞彙集合更具準確性，本研究即制定領域詞彙組關係距離四分位數  $W_{k,1}$ （於此係於第三位之四分位數進行篩選），透過更為精確之方式挑選得實具高關聯程度之領域詞彙組合，以防止極端值對挑選過程所造成之影響。於此計算領域詞彙組關係距離  $RD(AVW_{k,i}, BVW_{k,j})$  中第三位之四分位數指標  $Q$ ，用以界定第三位之領域詞彙組關係距離四分位數  $W_{k,1}$ ，並以此訂定高關聯程度詞彙組之四分位數門檻值（如公式(3.5)所示）。當中， $N(AVW_{k,\bullet}, BVW_{k,\bullet})$  為詞彙組合之總數，此外，若該有效詞彙組之整體關係距離  $ARD_{k,p}$  小於（或相等）四分位數  $W_{k,1}$ ，即表示該詞彙組屬於高關聯程度之領域詞彙組合  $HRD_k$ 。

$$\begin{aligned}
 Q &= N(AVW_{k,\bullet}, BVW_{k,\bullet}) \times 75\% \\
 W_{k,1} &= \begin{cases} ARD_{k,p \times 75\%} & \text{IF } Q \notin \{X : |X| \in N\} \\ \frac{1}{2} \times (ARD_{k,p \times 75\%} + ARD_{k,p \times 75\% + 1}) & \text{IF } Q \in \{X : |X| \in N\} \end{cases} \quad (3.5) \\
 \text{If } ARD_{k,p} \leq W_{k,1} &\text{ Then } (AVW_{k,i}, BVW_{k,j}) \in HRD_k
 \end{aligned}$$

### 2. 領域詞彙組關係距離中位數

除上述「領域詞彙組關係距離四分位數」外，本研究亦制定領域詞彙組關係距離中位數  $W_{k,2}$ ，以提供另一種挑選實具高關聯程度領域詞彙組合之方法，藉此防止極大值與極小值對建構領域性關鍵詞彙集合造成之影響，進而使欲建構之領域性關鍵詞彙集合更具準確度與可信度（如公式(3.6)所示）。

$$\begin{aligned}
 W_{k,2} &= \begin{cases} ARD_{k, \frac{p+1}{2}} & \text{IF } N(AVW_{k,\bullet}, BVW_{k,\bullet}) \text{ is odd} \\ \frac{1}{2} \times (ARD_{k, \frac{p+1}{2}} + ARD_{k, \frac{p+1}{2} + 1}) & \text{IF } N(AVW_{k,\bullet}, BVW_{k,\bullet}) \text{ is even} \end{cases} \quad (3.6) \\
 \text{If } ARD_{k,p} \leq W_{k,2} &\text{ Then } (AVW_{k,i}, BVW_{k,j}) \in HRD_k
 \end{aligned}$$

### 3. 領域詞彙組關係距離平均值

「領域詞彙組關係距離平均值」代表領域詞彙組關係之整體平均距離，若領域詞彙組關係距離低於整體平均值，則表示此領域詞彙組對應之關係距離未能達到關係距離之整體標準，而未達到整體平均之領域詞彙組關係距離，亦不適合用於建構領域性關鍵詞彙集合，故於此計算領域詞彙組關係距離之平均值  $W_{k,3}$ ，以挑選實具高關聯程度之領域詞彙組合（如公式(3.7)所示）。

$$W_{k,3} = \frac{\sum_{\text{all } p} \text{ARD}_{k,p}}{N(\text{AVW}_{k,\bullet}, \text{BVW}_{k,\bullet})}, \quad (3.7)$$

If  $\text{ARD}_{k,p} \leq W_{k,3}$  Then  $(\text{AVW}_{k,i}, \text{BVW}_{k,j}) \in \text{HRD}_k$

### 4. 領域詞彙組關係距離自訂門檻值

於此亦可自行制訂領域詞彙組關係距離門檻值  $W_{k,4}$ ，以進行領域性關鍵詞彙集合篩選門檻之設定（如公式(3.8)所示）。

$$\text{If } \text{ARD}_{k,p} \leq W_{k,4} \text{ Then } (\text{AVW}_{k,i}, \text{BVW}_{k,j}) \in \text{HRD}_k \text{ where } 0 \leq W_{k,4} < 1 \quad (3.8)$$

最後，為使具備高關聯程度之領域詞彙組合得以於合理之狀況下，各自分離成獨立之詞彙，以用於建構領域性關鍵詞彙集合，首先，必須先行將符合門檻之領域詞彙組合予以分割，以視開始篩選適合建構領域性關鍵詞彙集合詞彙前之預處理動作（如公式(3.9)所示），當中，若分割所得之獨立詞彙已存在，則該詞彙不再重複儲存。其次，依據所有詞彙組關係距離中，符合門檻值詞彙組內，獨立詞彙（ $S_{\text{HRD}_{k,w}}$ ）達成門檻值之次數（ $N(\text{HRD}_k, S_{\text{HRD}_{k,w}})$ ），進行相關之統計，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  $\text{WRC}_k$  中之詞彙（如公式(3.10)所示）。

$$\begin{aligned} &\text{If } (\text{AVW}_{k,i}, \text{BVW}_{k,j}) \in \text{HRD}_k \\ &\text{Then } S_{\text{HRD}_{k,w}} = \text{AVW}_{k,i} \text{ and } S_{\text{HRD}_{k,w+1}} = \text{BVW}_{k,j} \\ &\text{Where } S_{\text{HRD}_{k,w}} \notin S_{\text{HRD}_{k,\bullet}} \end{aligned} \quad (3.9)$$

$$\text{If } N(\text{HRD}_{k,w}, S_{\text{HRD}_{k,w}}) \geq \frac{\sum_{\text{all } w} N(\text{HRD}_k, S_{\text{HRD}_{k,w}})}{N(S_{\text{HRD}_{k,\cdot}})} \text{ Then } S_{\text{HRD}_{k,w}} \in \text{WRC}_k \quad (3.10)$$

經由上述三步驟之推論及運行，將可建構多個實具領域概念特性之「領域性關鍵詞彙集合」。此階段「領域性關鍵詞彙集合之建立」係藉具專業領域知識訓練文件之挹注，並予以斷詞及過濾不具意義之詞彙後，計算出領域有效詞間之關係距離，爾後，經本研究制定之「四分位數」、「中位數」、「平均值」，以及「使用者自訂門檻值」等四種篩選方式，以更為精確之方式挑選得實具高關聯程度之領域詞彙組合，且藉由符合門檻詞彙組合之分割及篩選，於最終建構相關之「領域性關鍵詞彙集合」，以視此階段最終目的。

### 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 3.3** 所示。

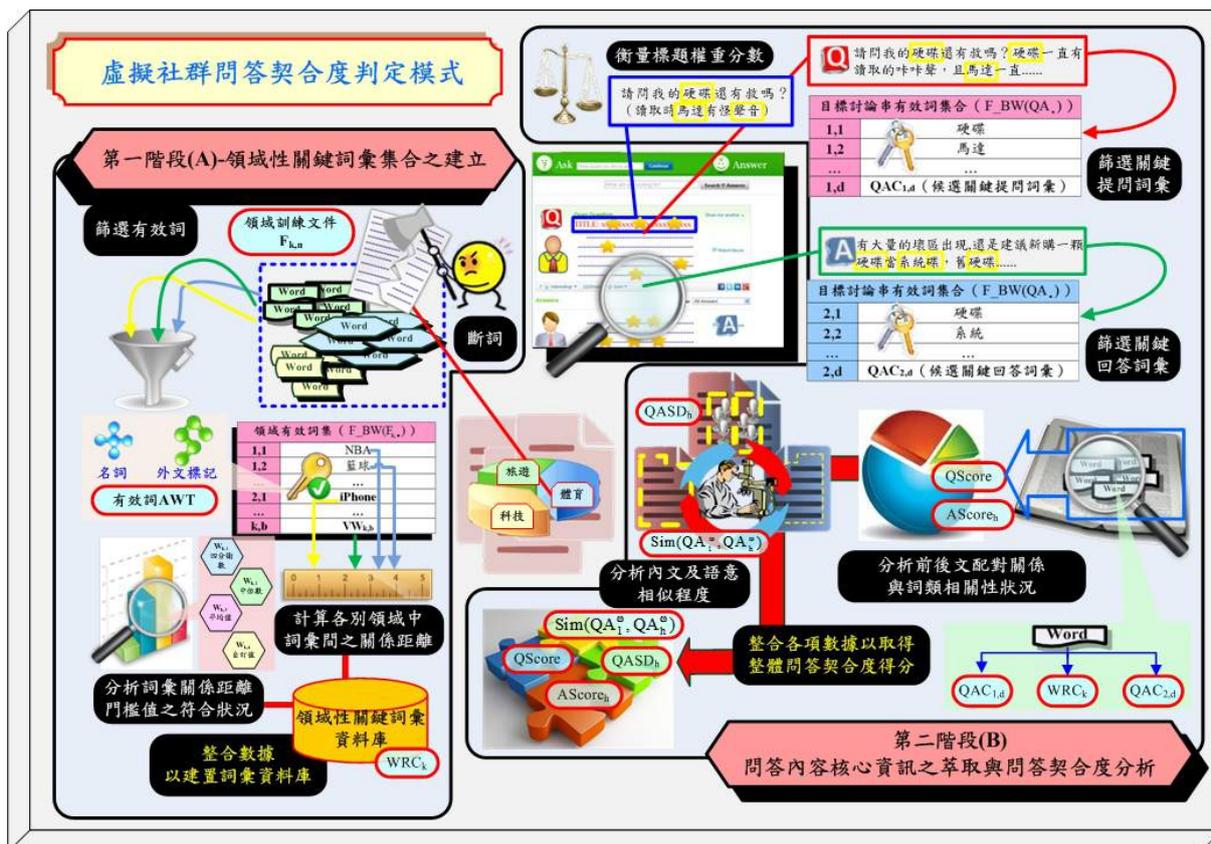


圖 3.3、問答內容核心資訊之萃取與問答契合度分析之運作示意圖

說明本推論模式之階段(B)前，將階段(B)採用之符號定義如下：

- $QA_h$  目標虛擬社群討論串中之第  $h$  筆資料，當中「資料」分別係指討論串中之提問與回答， $h=1,2,\dots$ ， $QA_1$  為「提問」、 $h \geq 2$  為「其他回答」（於同一筆討論串中）
- $BW(QA_h)$  目標虛擬社群討論串中，所有資料經斷詞後所得詞彙之集合
- $QAW_{h,c}$  目標虛擬社群討論串中，第  $h$  筆資料經斷詞後所得之  $c$  個詞彙（第  $c$  個斷詞結果）
- $AWT$  詞類屬於名詞或外文標記之詞彙，當中名詞包含普通名詞、專有名稱、地方詞與位置詞，外文標記不包含單純之數值（如 1、2、3...等），此外，若外文標記之前或後方存在數字，則此外文標記與數字將組合為一個有效詞（如 600D、S5200 與 F1.4D）
- $F\_BW(QA_h)$  目標虛擬社群討論串中，所有資料經斷詞及有效詞篩選後之集合，當中，有效詞係指具備實質意義之詞彙（如「手機」與「iPhone」等名詞及外文標記皆為有效詞；「的」及「你」等虛詞為無效詞）

$QAC_{h,d}$	目標虛擬社群討論串中，第 h 筆資料經斷詞及有效詞篩選後所得之第 d 個候選關鍵提問或回答之詞彙（依資料所屬之種類而定）
$OScore(QAC_{h,d})$	目標虛擬社群討論串中，第 h 筆資料內第 d 個候選關鍵提問或回答詞彙之前後文配對及標題加權整體分數
$QAC_{h,d,a}^P$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙前段語句內（即前一個標點符號間之語句）之第 a 個觸發詞
$QAC_{h,d,b}^N$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙後段語句內（即後一個標點符號間之語句）之第 b 個觸發詞
$FQ(QAC_{h,d,a}^P, QAC_{h,d}^P)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙前段語句內，第 a 個觸發詞於所屬語句中出現之頻率
$FQ(QAC_{h,d,b}^N, QAC_{h,d}^N)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙後段語句內，第 b 個觸發詞於所屬語句中出現之頻率
$FQ(QAC_{h,d,a}^P, QA_h)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙前段語句內，第 a 個觸發詞於所屬資料中出現之頻率
$FQ(QAC_{h,d,b}^N, QA_h)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問或回答詞彙後段語句內，第 b 個觸發詞於所屬資料中出現之頻率
$N(QAC_{h,\bullet}^P)$	目標虛擬社群討論串中，第 h 筆資料內所有候選關鍵提問或回答詞彙前段語句內觸發詞之總數
$N(QAC_{h,\bullet}^N)$	目標虛擬社群討論串中，第 h 筆資料內所有候選關鍵提問或回答詞彙後段語句內觸發詞之總數
$N(QAC_{h,\bullet})$	目標虛擬社群討論串中，第 h 筆資料內候選關鍵提問或回答詞彙之總數
$FQ(QAC_{h,d}, Title)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問詞彙於討論串標題中出現之頻率
$FQ(QAC_{h,d}, QA_h)$	目標虛擬社群討論串中，第 h 筆資料之第 d 個候選關鍵提問詞彙，於所屬資料中出現之頻率
$Set(BQA_{h,\bullet,ARV(POS.)})$	目標虛擬社群討論串中，第 h 筆資料內各個候選關鍵提問與回答詞彙向前推算 ARV(POS.) 個詞彙後，當中所含詞彙之集合

$POS_m$	訓練文章所含之第 $m$ 個詞類
$N(POS_m)$	訓練文章所含之詞類總數
$ARV(POS_s)$	詞類距離採納值， $0 \leq ARV(POS_s) \leq N(POS_m)$
$BQA_{h,d,n}$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙向前推算之第 $n$ 個詞彙， $n \leq ARV(POS_s)$
$PWN(QAC_{h,d})$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙可向前推算之詞彙總數
$RScore(QAC_{h,d})$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙之詞類相關性平均分數
$FQ(\text{Set}(BQA_{h,d,n}))$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙，向前推算之第 $n$ 個詞彙於集合 $\text{Set}(BQA_{h_s, ARV(POS_s)})$ 中出現之頻率， $n \leq ARV(POS_s)$
$FQ(BQA_{h,d,n}, QA_h)$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙，向前推算之第 $n$ 個詞彙於所屬資料中出現之頻率， $n \leq ARV(POS_s)$
$RFQ(\text{Set}(BQA_{h,d}))$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙於集合 $\text{Set}(BQA_{h_s, ARV(POS_s)})$ 中，重複出現之頻率
$IScore(QAC_{h,d})$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙之代表性（重要性）得分
$OScoreP(QAC_{h,d})$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙之前後文配對及標題加權整體分數比重值
$RScoreP(QAC_{h,d})$	目標虛擬社群討論串中，第 $h$ 筆資料之第 $d$ 個候選關鍵提問或回答詞彙之詞類相關性平均分數比重值
$D_T$	目標虛擬社群討論串所屬之領域區塊（由發問者提問所處之領域而定），如發問者於體育類型之討論區提問，則該提問即屬「體育領域」之區塊
$WRC_k$	第 $k$ 類領域性關鍵詞彙集合
$QV\_QWM_{k,y}$	目標虛擬社群討論串中，第 $y$ 筆與第 $k$ 類別領域性關鍵詞彙集合（ $WRC_k$ ），相符合候選關鍵提問詞彙（ $QWM_{k,y}$ ）之關鍵提問詞彙代表性（重要性）得分

$QScore_k$	目標虛擬社群討論串中，提問之領域性提問品質得分
$FQ(QWM_{k,y}, QA_1)$	第 $y$ 筆與第 $k$ 類別領域性關鍵詞彙集合 ( $WRC_k$ )，相符合候選關鍵提問詞彙於該提問中出現之頻率
$QV\_AWM_{h,z}$	目標虛擬社群討論串中，第 $z$ 筆回答與提問有效詞集合 ( $F\_BW(QA_1)$ )，相符合候選關鍵回答詞彙 ( $AWM_{h,z}$ ) 之關鍵回答詞彙代表性 (重要性) 得分
$AScore_h$	目標虛擬社群討論串中，第 $h$ 筆資料之回答品質得分
$FQ(AWM_{h,z}, QA_h)$	目標虛擬社群討論串中，第 $z$ 筆相符合候選關鍵回答詞彙於該回答中出現之頻率
$QA_h^0$	目標虛擬社群討論串中，以所有候選關鍵提問或回答詞彙為基礎所建置第 $h$ 筆資料之集合向量
$Sim(QA_1^0, QA_h^0)$	目標虛擬社群討論串中，提問與第 $h$ 筆資料間問答內文契合度之得分
$GSH(QAC_{1,i}^Q)$	目標虛擬社群討論串中，提問內第 $i$ 筆候選關鍵提問或回答詞彙經 Google 搜尋引擎搜尋後之結果數
$GSH(QAC_{h,j}^A)$	目標虛擬社群討論串中，第 $h$ 筆資料 (回答) 內第 $j$ 筆候選關鍵提問或回答詞彙經 Google 搜尋引擎搜尋後之結果數
$GN$	Google 搜尋引擎所包含之網頁總數
$N(QAC_{h,\cdot})$	目標虛擬社群討論串中，第 $h$ 筆資料內候選關鍵提問或回答詞彙之總數
$OASD_h$	目標虛擬社群討論串中，提問與第 $h$ 筆資料間之問答語意關係距離
$OMScore_h$	目標虛擬社群討論串中，提問與第 $h$ 筆資料間之整體問答契合度得分

### **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 ( $QA_h$ ) 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  $BW(QA_{\cdot})$  (如公式(3.11)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  $F\_BW(QA_h)$  (如公式(3.12)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」。

$$BW(QA_h) = \{QAW_{h,1}, QAW_{h,2}, \dots, QAW_{h,c}\} \quad (3.11)$$

$$F\_BW(QA_h) = \{QAC_{h,1}, QAC_{h,2}, \dots, QAC_{h,d} \mid QAC_{h,d} \in AWT\} \quad (3.12)$$

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  $QAC_{h,d}^P$  與後段語句  $QAC_{h,d}^N$  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(3.13)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  $h = 1$

$$\begin{aligned}
 \text{Then OScore}(\text{QAC}_{h,d}) &= \left[ \sum_{\text{alla}} \left( \frac{\text{FQ}(\text{QAC}_{h,d,a}^P, \text{QAC}_{h,d}^P)}{N(\text{QAC}_{h,\bullet}^P)} + \frac{\text{FQ}(\text{QAC}_{h,d,a}^P, \text{QAC}_h^P)}{N(\text{QAC}_{h,\bullet})} \right) \right. \\
 &\quad \left. \sum_{\text{all b}} \left( \frac{\text{FQ}(\text{QAC}_{h,d,b}^N, \text{QAC}_{h,d}^N)}{N(\text{QAC}_{h,\bullet}^N)} + \frac{\text{FQ}(\text{QAC}_{h,d,b}^N, \text{QAC}_h^N)}{N(\text{QAC}_{h,\bullet})} \right) \right] \\
 &\quad \text{FQ}(\text{QAC}_{h,d}, \text{Title}) \\
 &\quad \cdot \left[ \frac{1}{N(\text{QAC}_{h,\bullet}^P)} \frac{1}{N(\text{QAC}_{h,\bullet}^N)} \frac{1}{\text{FQ}(\text{QCW}_{h,d}, \text{QCW}_h)} \right] \\
 \text{Else OScore}(\text{QAC}_{h,d}) &= \left[ \sum_{\text{alll}} \left( \frac{\text{FQ}(\text{QAC}_{h,d,a}^P, \text{QAC}_{h,d}^P)}{N(\text{QAC}_{h,\bullet}^P)} + \frac{\text{FQ}(\text{QAC}_{h,d,a}^P, \text{QAC}_h^P)}{N(\text{QAC}_{h,\bullet})} \right) \right] \\
 &\quad \left[ \sum_{\text{all m}} \left( \frac{\text{FQ}(\text{QAC}_{h,d,b}^N, \text{QAC}_{h,d}^N)}{N(\text{QAC}_{h,\bullet}^N)} + \frac{\text{FQ}(\text{QAC}_{h,d,b}^N, \text{QAC}_h^N)}{N(\text{QAC}_{h,\bullet})} \right) \right] \\
 &\quad \cdot \left[ \frac{1}{N(\text{QAC}_{h,\bullet}^P)} \frac{1}{N(\text{QAC}_{h,\bullet}^N)} \right]
 \end{aligned} \tag{3.13}$$

### 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(3.11)** 經斷詞後所得之詞彙集合  $\text{BW}(\text{QA}_\bullet)$ ，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  $\text{ARV}(\text{POS}_\bullet)$  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  $\text{Set}(\text{BQA}_{h,\bullet,\text{ARV}(\text{POS}_\bullet)})$ 」(如 **公式(3.14)** 所示)。

$$\text{Set}(\text{BQA}_{h,\bullet,\text{ARV}(\text{POS}_\bullet)}) = \begin{bmatrix} \text{BQA}_{h,1,1} & \text{BQA}_{h,2,1} & \cdots & \text{BQA}_{h,d,1} \\ \text{BQA}_{h,1,2} & \text{BQA}_{h,2,2} & \cdots & \text{BQA}_{h,d,2} \\ \vdots & \vdots & & \vdots \\ \text{BQA}_{h,1,n} & \text{BQA}_{h,2,n} & \cdots & \text{BQA}_{h,d,n} \end{bmatrix} \tag{3.14}$$

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  $FQ(\text{Set}(BQA_{h,d,n}))$  與  $RFQ(\text{Set}(BQA_{h,d}))$ ，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(3.15)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  $ARV(POS.)$ 」之情況時，將以  $PWN(QAC_{h,d})$  取代原本之詞類採納距離值。

IF  $PWN(QAC_{h,d}) < ARV(POS.)$

$$\text{Then } RScore(QAC_{h,d}) = \frac{\sum_{\text{all } n} \frac{FQ(\text{Set}(BQA_{h,d,n}))}{FQ(BQA_{h,d,n}, BQA_h)}}{PWN(QAC_{h,d}) \cdot N(QAC_{h,\bullet}) \cdot RFQ(\text{Set}(BQA_{h,d})) + 1} \quad (3.15)$$

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

### **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(3.16)所示）。

$$IScore(QAC_{h,d}) = OScore(QAC_{h,d}) \cdot OScoreP(QAC_{h,d}) + RScore(QAC_{h,d}) \cdot RScoreP(QAC_{h,d}) \quad (3.16)$$

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

### **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群聚度關鍵因素之一，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(3.17)及公式(3.18)所示)。

$$\begin{aligned} & \text{IF } D_T \in WRC_k \text{ and } QAC_{1,d} \in WRC_k \\ & \text{Then } QV\_QWM_y = IScore(QAC_{1,d}) \end{aligned} \quad (3.17)$$

$$QScore = \sum_{\text{all } y} QV\_QWM_y + \sum_{\text{all } y} \frac{QV\_QWM_y \cdot (FQ(QWM_y, QA_1) - 1)}{FQ(QWM_y, QA_1)} \quad (3.18)$$

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(3.19)及公式(3.20)所示)。

$$\begin{aligned} & \text{IF } QAC_{h,d} \in F\_BW(QA_1) \\ & \text{Then } QV\_AWM_z = IScore(QAC_{h,d}) \\ & \text{Where } h > 1 \end{aligned} \quad (3.19)$$

$$AScore_h = \sum_{\text{all } z} QV\_AWM_z + \sum_{\text{all } z} \frac{QV\_AWM_z \cdot (FQ(AWM_z, QA_h) - 1)}{FQ(AWM_z, QA_h)} \quad (3.20)$$

Where  $h > 1$

### 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  $Sim(QA_1^{\circ}, QA_h^{\circ})$ ，並視其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(3.21)所示)。

$$\begin{aligned}
QA_1^o &= [QAW_{1,1}^Q, QAW_{1,2}^Q, \dots, QAW_{1,p}^Q]^T \\
QA_h^o &= [QAW_{h,1}^A, QAW_{h,2}^A, \dots, QAW_{h,q}^A]^T \\
\text{Sim}(QA_1^o, QA_h^o) &= \frac{QA_1^o \cdot QA_h^o}{\|QA_1^o\| \times \|QA_h^o\|}
\end{aligned} \tag{3.21}$$

Where  $h > 1$

### 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(3.22)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。

$$QASD_h = \frac{\sum_{\text{all } i} \frac{\text{Max}\{\log \text{GSH}(QAC_{1,i}^Q), \log \text{GSH}(QAC_{h,j}^A)\} - \log \text{GSH}(QAC_{1,i}^Q, QAC_{h,j}^A)}{\log \text{GN} - \text{Min}\{\log \text{GSH}(QAC_{1,i}^Q), \log \text{GSH}(QAC_{h,j}^A)\}}}{N(QAC_{1,\bullet}^Q) \cdot N(QAC_{h,\bullet}^A)} \tag{3.22}$$

Where  $h > 1$

### 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  $\text{Sim}(QA_1^o, QA_h^o)$ 」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(3.23)所示)。

$$OMScore_h = \frac{QScore \cdot AScore_h + \frac{\text{Sim}(QA_1^o, QA_h^o)}{QASD_h}}{\sum_{\text{all } h} QScore \cdot AScore_h + \frac{\text{Sim}(QA_1^o, QA_h^o)}{QASD_h}} \tag{3.23}$$

Where  $h > 1$

本研究所提之「虛擬社群問答契合度解析模組」，主要分成兩階段之方法論，以解析目標虛擬社群討論串中提問與回答之契合程度，於第一階段中，主要之目的係建構「領域性關鍵詞彙集合」，以提供相關第二階段問答契合度解析中，領域性問答品質計算之依據，而於第二階段中，則分成三大面項之主軸進行問答內容核心資訊之萃取與問答契合度之解析，第一部份由步驟(B1)至步驟(B4)構成，當中之目的乃取得提問品質與回答品質得分，第二部分係透過步驟(B5)及步驟(B6)，針對問答內文契合及語意關聯程度進行解析，而第三部份則藉「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。綜上所述，本模組以 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，並視為本模組之最終目的。

### 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 3.4 所示。



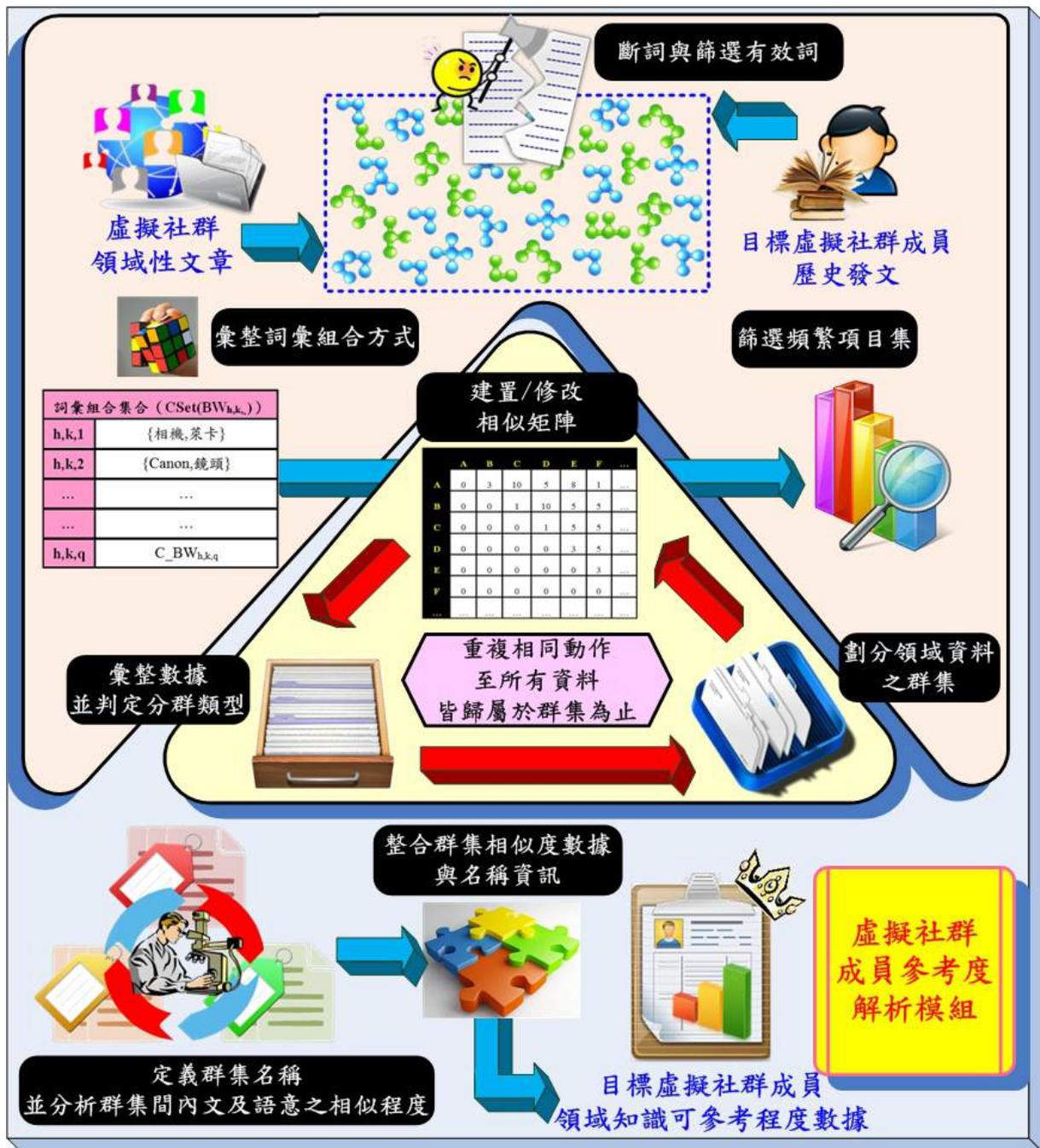


圖 3.4、虛擬社群成員參考度解析模組之運作示意圖

虛擬社群成員參考度之解析可分數個步驟進行之，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。

說明本推論模式之前，將採用之符號定義如下：

- $D_k$  第  $k$  個虛擬社群中之領域（如電腦、運動及旅遊等之領域）
- $M_T$  目標虛擬社群成員
- $Set_{h,k}$  第  $h$  個對象所包含或發佈第  $k$  個領域內資料之集合，當中「資

料」分別係指該對象所包含或發佈之討論串與歷史發文（依所屬對象而定）， $h=1,2,3,\dots$ ， $Set_{1,k}$  為「虛擬社群」、 $Set_{2,k}$  為「目標虛擬社群成員」所包含或發佈第  $k$  個領域內資料之集合，而  $h \geq 3$  之對象係指「其他虛擬社群成員」（不含目標成員）

$AH_{h,k,i}$	第 $h$ 個對象所包含或發佈第 $k$ 個領域內之第 $i$ 筆資料
AWT	詞類屬於名詞或外文標記之詞彙，當中名詞包含普通名詞、專有名稱、地方詞與位置詞，外文標記不包含單純之數值（如 1、2、3...等），此外，若外文標記之前或後方存在數字，則此外文標記與數字將組合為一個有效詞（如 600D、S5200 與 F1.4D）
$BW(AH_{h,k,\bullet})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，所有資料經斷詞後所得詞彙之集合
$BW_{h,k,p}$	第 $h$ 個對象所包含或發佈第 $k$ 個領域內第 $p$ 個詞彙
$F\_BW(AH_{h,k,\bullet})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，所有資料經斷詞及有效詞篩選後之集合，當中，有效詞係指具備實質意義之詞彙（如「手機」與「iPhone」等名詞及外文標記皆為有效詞；「的」及「你」等虛詞為無效詞）
$VW_{h,k,q}$	第 $h$ 個對象所包含或發佈第 $k$ 個領域內第 $q$ 個有效詞
$CSet(BW_{h,k,\bullet})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，所有不考量順序詞彙之組合，如存在三個詞彙 $a$ 、 $b$ 、 $c$ ，則有「 $a$ 」、「 $b$ 」、「 $c$ 」、「 $ab$ 」、「 $ac$ 」、「 $bc$ 」、「 $abc$ 」等七種組合
$C\_BW_{h,k,e}$	第 $h$ 個對象所包含或發佈第 $k$ 個領域內第 $e$ 種詞彙組合
$I(C\_BW_{h,k,e}, AH_{h,k,i})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，第 $i$ 筆資料內出現第 $q$ 種詞彙組合之指標函數；若 $I(C\_VW_{h,k,e}, AH_{h,k,i})=1$ ，即代表詞彙組合 $C\_VW_{h,k,e}$ 出現於該資料中；否則其值為 0
$FI(AH_{h,k,\bullet})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，所有資料內詞彙之頻繁項目集（Frequent Itemsets）
$N(AH_{h,k,\bullet})$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，所有資料之總數
MinSup	最小支持度，用以篩選詞彙組合為頻繁項目集
$Sim(AH_{h,k,a}^A, AH_{h,k,b}^B)$	第 $h$ 個對象所包含或發佈第 $k$ 個領域中，第 $a$ 筆資料與第 $b$ 筆資料之相似度數據

$\text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,a}^A)$	第 h 個對象所包含或發佈第 k 個領域中，屬於第 a 筆資料詞彙組合之集合
$\text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,b}^B)$	第 h 個對象所包含或發佈第 k 個領域中，屬於第 b 筆資料詞彙組合之集合
$\text{SM}[AH_{h,k,\bullet}]$	第 h 個對象所包含或發佈第 k 個領域中，以所有資料間之相似度所建置之相似矩陣 (Similarity Matrix)
$\text{MaxSet}(\text{SM}[AH_{h,k,\bullet}])$	相似矩陣中具備最大相似度資料對之集合，當中，「資料對」係指相似度判定中之兩資料，如 $\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)$ 則資料對即為 $(AH_{h,k,a}^A, AH_{h,k,b}^B)$
$\text{CMin}(\text{SM}[AH_{h,k,\bullet}])$	相似矩陣中符合條件之最小相似度數值
CType	資料分群時所採用之分群類型，CType=1,2,3，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 CType=1；最大相似度等於符合條件之最小相似度，則 CType=2；若最大相似度為 0，則 CType=3
$\text{CR}'_{h,k,\bullet}$	第 h 個對象所包含或發佈第 k 個領域中，已存在之群集 (於先前經分群時所保留之群集)
$\text{CR}_{h,k,c}$	第 h 個對象所包含或發佈第 k 個領域中，經合併後之第 c 個群集
$\text{CR}_{h,k,c+n}$	第 h 個對象所包含或發佈第 k 個領域中，經分群後所得之第 c+n 個新群集
$\text{N}(\text{CR}_{h,k,c})$	第 h 個對象所包含或發佈第 k 個領域中，第 c 個群集內資料之總數
$\text{ML}(\text{FI}(AH_{h,k,\bullet}), \text{CR}_{h,k,c})$	第 h 個對象所包含或發佈第 k 個領域中，第 c 個群集內具備最長長度之頻繁項目，如群集內存在「a」、「bc」與「abc」等三個頻繁項目，則「abc」為最長長度之頻繁項目
$\text{MN}(\text{FI}(AH_{h,k,\bullet}), \text{CR}_{h,k,c})$	第 h 個對象所包含或發佈第 k 個領域中，同時出現於第 c 個群集內其他資料中次數最高之頻繁項目，如群集內存在 A 與 B 兩筆資料，且此兩筆資料分別由「a、b、c」及「b、f」等有效詞構成，於 b 為頻繁項目的前提下，「b」即為次數最高之頻繁項目

Topic <sub>h,k,c</sub>	第 h 個對象所包含或發佈第 k 個領域中，第 c 個群集之名稱
CTSet_Topic <sub>h,k,c</sub>	第 h 個對象所包含或發佈第 k 個領域中，第 c 個群集之候選群集名稱集合
Topic' <sub>h,k,•</sub>	第 h 個對象所包含或發佈第 k 個領域中，已存在之群集名稱
AH <sup>A<math>\omega</math></sup> <sub>1,k,a</sub>	虛擬社群所包含第 k 個領域中，以第 a 筆討論串經斷詞後之詞彙為基礎所建置之集合向量
AH <sup>B<math>\omega</math></sup> <sub>2,k,b</sub>	目標虛擬社群成員所發佈第 k 個領域中，以第 b 筆歷史發文經斷詞後之詞彙為基礎所建置之集合向量
SimVS(CR <sup>X</sup> <sub>1,k,x</sub> , CR <sup>Y</sup> <sub>2,k,y</sub> )	虛擬社群及目標虛擬社群成員所包含與發佈第 k 個領域中，第 x 個群集與第 y 個群集間資料之內文相似度數據
N(AH <sup>A</sup> <sub>1,k,•</sub> )	虛擬社群所包含第 k 個領域中所有討論串之總數
N(AH <sup>B</sup> <sub>2,k,•</sub> )	目標虛擬社群成員所發佈第 k 個領域中所有歷史發文之總數
SimNGD(CR <sup>X</sup> <sub>1,k,x</sub> , CR <sup>Y</sup> <sub>2,k,y</sub> )	虛擬社群及目標虛擬社群成員所包含與發佈第 k 個領域中，第 x 個群集與第 y 個群集間資料之語意相似度數據
GSH(VW <sup>A</sup> <sub>1,k,m</sub> , AH <sup>A</sup> <sub>1,k,a</sub> )	虛擬社群所包含第 k 個領域中，第 a 筆討論串內第 m 個有效詞經 Google 搜尋引擎搜尋後之結果數
GSH(VW <sup>B</sup> <sub>2,k,n</sub> , AH <sup>B</sup> <sub>2,k,b</sub> )	目標虛擬社群成員所發佈第 k 個領域中，第 b 筆歷史發文內第 n 個有效詞經 Google 搜尋引擎搜尋後之結果數
GN	Google 搜尋引擎所包含之網頁總數
N(VW <sup>A</sup> <sub>1,k,•</sub> )	虛擬社群所包含第 k 個領域中所有有效詞之總數
N(VW <sup>B</sup> <sub>2,k,•</sub> )	目標虛擬社群成員所發佈第 k 個領域中所有有效詞之總數
MDR(D <sub>k,x</sub> , (M <sub>T</sub> , DK <sub>k,y</sub> ))	目標虛擬社群成員第 k 大類內第 y 子項之領域知識，於虛擬社群第 x 子項中領域知識可參考程度之數據

### 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<sub>h,k</sub> 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<sub>h,k,•</sub>) (如公式(3.26)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(3.42)。

$$BW(AH_{h,k,\bullet}) = \{BW_{h,k,1}, BW_{h,k,2}, \dots, BW_{h,k,p}\} \quad (3.26)$$

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(3.27)所示)，而此部份所得之資訊將運用於步驟(C7)中之公式(3.43)。

$$F\_BW(AH_{h,k,\bullet}) = \{VW_{h,k,1}, VW_{h,k,2}, \dots, VW_{h,k,q} \mid VW_{h,k,q} \in AWT\} \quad (3.27)$$

### 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(3.28)所示)。

$$CSet(BW_{h,k,\bullet}) = \{C\_BW_{h,k,1}, C\_BW_{h,k,2}, \dots, C\_BW_{h,k,e}\} \quad (3.28)$$

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(3.29)所示)。

$$I(C\_BW_{h,k,e}, AH_{h,k,i}) = \begin{cases} 1, & \text{If } C\_BW_{h,k,e} \text{ exist in } AH_{h,k,i} \\ 0, & \text{Otherwise} \end{cases} \quad (3.29)$$

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合，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(3.30)所示)。

$$FI(AH_{h,k,\bullet}) = \{C\_BW_{h,k,1}, C\_BW_{h,k,2}, \dots, C\_BW_{h,k,e}\} \quad (3.30)$$

$$|\sum_{\text{all } i} I(C\_BW_{h,k,i}, AH_{h,k,i}) \geq N(AH_{h,k,\bullet}) \cdot \text{MinSup}\}$$

### 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(3.31)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。

$$\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) = N(\text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,a}^A) \cap \text{Set}(C\_BW_{h,k,\bullet}, AH_{h,k,b}^B)) \quad (3.31)$$

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(3.30)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(3.31)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(3.30)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(3.32)）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 $\text{Sim}(AH_{h,k,1}^A, AH_{h,k,2}^B)$ 與 $\text{Sim}(AH_{h,k,2}^A, AH_{h,k,1}^B)$ ）之處理方式，其作法亦同於上。

$$SM[AH_{h,k,\bullet}] = \begin{bmatrix} 0 & \text{Sim}(AH_{h,k,1}^A, AH_{h,k,2}^B) & \text{Sim}(AH_{h,k,1}^A, AH_{h,k,3}^B) & \cdots & \text{Sim}(AH_{h,k,1}^A, AH_{h,k,b}^B) \\ 0 & 0 & \text{Sim}(AH_{h,k,2}^A, AH_{h,k,3}^B) & \cdots & \text{Sim}(AH_{h,k,2}^A, AH_{h,k,b}^B) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (3.32)$$

#### 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  $\text{MaxSet}(SM[AH_{h,k,\bullet}])$  (如公式(3.33)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  $\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)$  則資料對即為  $(AH_{h,k,a}^A, AH_{h,k,b}^B)$ 。

$$\text{MaxSet}(SM[AH_{h,k,\bullet}]) = \left\{ \begin{array}{l} (AH_{h,k,1}^A, AH_{h,k,2}^B), (AH_{h,k,1}^A, AH_{h,k,3}^B), (AH_{h,k,2}^A, AH_{h,k,3}^B), \dots, (AH_{h,k,a}^A, AH_{h,k,b}^B) \\ \text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) \in \text{Max}(\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)) \\ \text{and } AH_{h,k,a}^A \text{ or } AH_{h,k,b}^B \text{ has not been assigned to any cluster} \end{array} \right\} \quad (3.33)$$

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  $\text{CMin}(SM[AH_{h,k,\bullet}])$  (如公式(3.34)所示)。

$$\begin{aligned} \text{CMin}(SM[AH_{h,k,\bullet}]) &= \text{Min}(\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)) \\ \text{Where } \text{Min}(\text{Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B)) &\neq 0 \end{aligned} \quad (3.34)$$

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(3.35)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  $\text{CType}=3$ 。

$$CType = \begin{cases} 1, & \text{If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) \neq \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ & \text{and } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) > \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ 2, & \text{If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) = \text{CMin}(\text{SM}[\text{AH}_{h,k,\bullet}]) \\ 3, & \text{If } \text{Max}(\text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B)) = 0 \end{cases} \quad (3.35)$$

### 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 (CType) 而定，此外，當該批次 (MaxSet(SM[AH<sub>h,k,•</sub>])) 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 CType=1 時，若資料對 ((AH<sub>h,k,a</sub><sup>A</sup>, AH<sub>h,k,b</sub><sup>B</sup>)) 中，兩筆資料皆不屬於已存在之群集 (CR<sub>h,k,•</sub>'), 則針對此些符合條件之資料對，各別增加新的群集 CR<sub>h,k,c+n</sub> (如公式(3.37)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。

$$CR_{h,k,c+n} = \left\{ \begin{array}{l} (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B), (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,3}^B), (\text{AH}_{h,k,2}^A, \text{AH}_{h,k,3}^B), \dots, (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \\ | \text{CType} = 1 \text{ and } \text{AH}_{h,k,a}^A \notin \text{CR}'_{h,k,\bullet} \text{ and } \text{AH}_{h,k,b}^B \notin \text{CR}'_{h,k,\bullet} \end{array} \right\} \\ \text{and } \text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) = 0 \forall a, b \\ \text{Where } (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \in \text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}]) \quad (3.36)$$

- 當 CType=1 時，若於資料對 ((AH<sub>h,k,a</sub><sup>A</sup>, AH<sub>h,k,b</sub><sup>B</sup>)) 中，有其中一筆資料 (AH<sub>h,k,a</sub><sup>A</sup> 或 AH<sub>h,k,b</sub><sup>B</sup>) 屬於已存在之群集 (CR<sub>h,k,•</sub>'), 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<sub>h,k,c</sub> (如公式(3.36)所示)，當中，若有任一筆資料屬於新群集 (CR<sub>h,k,c+n</sub>'), 則此條件不成立 (亦即不將此資料對進行合併)。

$$CR_{h,k,c} = \left\{ \begin{array}{l} (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,2}^B), (\text{AH}_{h,k,1}^A, \text{AH}_{h,k,3}^B), (\text{AH}_{h,k,2}^A, \text{AH}_{h,k,3}^B), \dots, (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \\ | \text{CType} = 1 \text{ and } \text{AH}_{h,k,a}^A \in \text{CR}'_{h,k,\bullet} \text{ or } \text{AH}_{h,k,b}^B \in \text{CR}'_{h,k,\bullet} \\ \text{and } \text{AH}_{h,k,a}^A \notin \text{CR}_{h,k,c+n} \text{ and } \text{AH}_{h,k,b}^B \notin \text{CR}_{h,k,c+n} \end{array} \right\} \cup \text{CR}'_{h,k,c} \\ \text{and } \text{Sim}(\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) = 0 \forall a, b \\ \text{Where } (\text{AH}_{h,k,a}^A, \text{AH}_{h,k,b}^B) \in \text{MaxSet}(\text{SM}[\text{AH}_{h,k,\bullet}]) \quad (3.37)$$

- 當 CType=2 時，則針對此批資料中 (MaxSet(SM[AH<sub>h,k,•</sub>]))，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<sub>h,k,c+n+1</sub> (如公式(3.38)所示)。

$$\begin{aligned}
 CR_{h,k,c+n+1} &= \left\{ \begin{array}{l} AH_{h,k,1}^A, AH_{h,k,2}^B, \dots, AH_{h,k,a}^A, AH_{h,k,b}^B \\ | CType = 2 \text{ and } AH_{h,k,a}^A \notin CR'_{h,k,•} \text{ or } AH_{h,k,b}^B \notin CR'_{h,k,•} \end{array} \right\} \\
 \text{and Sim}(AH_{h,k,a}^A, AH_{h,k,b}^B) &= 0 \forall a, b \\
 \text{Where } (AH_{h,k,a}^A, AH_{h,k,b}^B) &\in \text{MaxSet}(SM[AH_{h,k,•}])
 \end{aligned} \tag{3.38}$$

- 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<sub>h,k,c+n+2</sub> (如公式(3.39)所示)。

$$CR_{h,k,c+n+2} = \left\{ AH_{h,k,1}, AH_{h,k,2}, \dots, AH_{h,k,i} \mid CType = 3 \text{ and } AH_{h,k,i} \notin CR'_{h,k,•} \right\} \tag{3.39}$$

### 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 Topic<sub>h,k,c</sub> (如公式(3.40)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 (CTSet\_Topic<sub>h,k,c</sub>) 中，並於後續公式(3.41)中進行更進一步之處理。

$$\begin{aligned}
 &\text{If } N(CR_{h,k,c}) \neq 1 \text{ and } FI(AH_{h,k,•}) \in ML(FI(AH_{h,k,•}), CR_{h,k,c}) \\
 &\text{and } FI(AH_{h,k,•}) \in MN(FI(AH_{h,k,•}), CR_{h,k,c}) \\
 &\text{Then } FI(AH_{h,k,•}) = \text{Topic}_{h,k,c} \\
 &\text{Else } FI(AH_{h,k,•}) \in \text{CTSet\_Topic}_{h,k,c} \\
 &\text{Where } FI(AH_{h,k,•}) \in CR_{h,k,c} \text{ and } \text{CTSet\_Topic}_{h,k,c} \notin \text{Topic}'_{h,k,•}
 \end{aligned} \tag{3.40}$$

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(3.41)所示)。

$$\begin{aligned} &\text{If } \text{FI}(\text{AH}_{h,k,\bullet}) \in \text{MN}(\text{FI}(\text{AH}_{h,k,\bullet}), \text{CR}_{h,k,c}) \text{ and } \text{FI}(\text{AH}_{h,k,\bullet}) \notin \text{Topic}_{h,k,\bullet} \\ &\text{Then } \text{FI}(\text{AH}_{h,k,\bullet}) = \text{Topic}_{h,k,c} \\ &\text{Where } \text{FI}(\text{AH}_{h,k,\bullet}) \in \text{CTSet\_Topic}_{h,k,c} \end{aligned} \quad (3.41)$$

### 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 ( $\text{CR}_{h,k,c}$ )，分別劃分並定義為「 $\text{CR}_{1,k,x}^X$ 」與「 $\text{CR}_{2,k,y}^Y$ 」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(3.42)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同向量係根據步驟(C1)中公式(3.26)所得之斷詞結果所建置。

$$\begin{aligned} \text{AH}_{1,k,a}^{A\omega} &= [\text{BW}_{h,k,a,1}^A, \text{BW}_{h,k,a,2}^A, \dots, \text{BW}_{h,k,a,v}^A]^T \\ \text{AH}_{2,k,b}^{B\omega} &= [\text{BW}_{h,k,b,1}^B, \text{BW}_{h,k,b,2}^B, \dots, \text{BW}_{h,k,b,w}^B]^T \\ \text{SimVS}(\text{CR}_{1,k,x}^X, \text{CR}_{2,k,y}^Y) &= \frac{\sum_{\text{AH}_{1,k,a}^A \in \text{CR}_{1,k,x}^X} \sum_{\text{AH}_{2,k,b}^B \in \text{CR}_{2,k,y}^Y} \frac{\text{AH}_{1,k,a}^{A\omega} \cdot \text{AH}_{2,k,b}^{B\omega}}{\|\text{AH}_{1,k,a}^{A\omega}\| \cdot \|\text{AH}_{2,k,b}^{B\omega}\|}}{N(\text{AH}_{1,k,\bullet}^A) + N(\text{AH}_{2,k,\bullet}^B)} \quad (3.42) \\ \text{Where } \text{AH}_{1,k,\bullet}^A &\in \text{CR}_{1,k,x}^X \text{ and } \text{AH}_{2,k,\bullet}^B \in \text{CR}_{2,k,y}^Y \end{aligned}$$

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(3.43)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(3.42)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彙，即藉步驟(C1)中公式(3.27)有效詞篩選之結果所取得。

$$\text{SimNGD}(\text{CR}_{1,k,x}^X, \text{CR}_{2,k,y}^Y) = \frac{\sum_{\text{all } m,n} \frac{\text{Max}\{\log \text{GSH}(\text{VW}_{1,k,m}^A, \text{AH}_{1,k,a}^A), \log \text{GSH}(\text{VW}_{2,k,n}^B, \text{AH}_{2,k,b}^B)\} - \log \text{GSH}((\text{VW}_{1,k,m}^A, \text{AH}_{1,k,a}^A), (\text{VW}_{2,k,n}^B, \text{AH}_{2,k,b}^B))}{\log \text{GN} - \text{Min}\{\log \text{GSH}(\text{VW}_{1,k,m}^A, \text{AH}_{1,k,a}^A), \log \text{GSH}(\text{VW}_{2,k,n}^B, \text{AH}_{2,k,b}^B)\}}}{\frac{N(\text{VW}_{1,k,\bullet}^A) \cdot N(\text{VW}_{2,k,\bullet}^B)}{N(\text{AH}_{1,k,\bullet}^A) + N(\text{AH}_{2,k,\bullet}^B)}}$$

Where  $\text{VW}_{1,k,m}^A \in \text{AH}_{1,k,a}^A$  and  $\text{VW}_{2,k,n}^B \in \text{AH}_{2,k,b}^B$  and  $\text{AH}_{1,k,\bullet}^A \in \text{CR}_{1,k,x}^X$  and  $\text{AH}_{2,k,\bullet}^B \in \text{CR}_{2,k,y}^Y$

(3.43)

最終，本步驟將公式(3.42)與公式(3.43)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  $\text{MDR}(\text{D}_{k,x}, (\text{M}_T, \text{DK}_{k,y}))$  (如公式(3.44)所示)。

$$\text{MDR}(\text{D}_{k,x}, (\text{M}_T, \text{DK}_{k,y})) = \frac{\frac{\text{SimVS}(\text{CR}_{1,k,x}^X, \text{CR}_{2,k,y}^Y)}{\text{SimNGD}(\text{CR}_{1,k,x}^X, \text{CR}_{2,k,y}^Y)}}{\sum_{\text{all } h,y} \frac{\text{SimVS}(\text{CR}_{1,k,x}^X, \text{CR}_{h,k,y}^Y)}{\text{SimNGD}(\text{CR}_{1,k,x}^X, \text{CR}_{h,k,y}^Y)}}$$

(3.44)

Where  $h > 1$

綜上所述，本研究所提之「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。

根據本研究所發展之「虛擬社群知識群聚度判定模式」而言，本研究乃採用諸多演算法進行相關之解析，針對各項演算法之應用、擴展、改良與整合狀況，本研究將之彙整如表 3.2。



表 3.2、參考文獻延伸與本研究發展彙整表

探索主題	發展方法	演算法	參考文獻	本研究發展與延伸	
問答契合度 解析	領域性關鍵詞彙 集合建立	門檻篩選法則； 文章斷詞法； 詞彙語意解析法 (擴展既有研究方法)	中研院斷詞系統	<ul style="list-style-type: none"> <li>● 整合中研院平衡語料庫詞類標記集，將斷詞法擴展為有效詞篩選法</li> <li>● 將原為兩詞彙間語意解析方法，擴展為虛擬社群領域討論串中，領域性關鍵詞彙篩選方法</li> <li>● 結合四分位、中位數與平均值統計法，劃分領域關鍵詞彙，並建置領域性關鍵詞彙資料庫</li> </ul>	
			Cilibrasi 與 Vitanyi (2007)		
			四分位、中位數與平均值統計法		
	問答核心資訊萃取	語句上下文分析法 (改良並擴展既有研究方法)	中研院斷詞系統	Tung 與 Lu (2012)	<ul style="list-style-type: none"> <li>● 將原須人工篩選關鍵詞彙並建置語料庫之方法，改良為自動化篩選關鍵詞彙與建置語料庫</li> <li>● 擴展部落格文章關鍵詞彙資訊萃取法，為問答核心內容資訊萃取法</li> <li>● 結合斷詞法、領域性關鍵詞彙集合概念及語句上下文解析法，計算問答詞彙代表性得分</li> </ul>
				--	
	問答品質衡量	啟發式演算法 (創新研究方法)	--	--	<ul style="list-style-type: none"> <li>● 結合詞彙重要性分佈及領域性關鍵詞彙集合概念，發展問答品質衡量演算法</li> </ul>
	問答內文 契合度分析	內文相似度解析法 (應用既有研究方法)	Salton 等人 (1975)	--	<ul style="list-style-type: none"> <li>● 轉化文章內文相似度解析法，為提問與回答內文契合度分析法</li> </ul>
問答語意 關係距離解析	詞彙語意解析法 (擴展既有研究方法)	Cilibrasi 與 Vitanyi (2007)	--	<ul style="list-style-type: none"> <li>● 將原為兩詞彙間語意解析方法，擴展為問答內容間語意關係距離之分析法</li> </ul>	
整體問答契合度 解析法	內文相似度解析法； 詞彙語意解析法 (擴展並整合既有研究方法)	Salton 等人 (1975)	--	<ul style="list-style-type: none"> <li>● 將內文相似度解析法，以及擴展後詞彙語意解析法，進行整合，並結合問答品質衡量、問答內文契合度，以及問答語意關係距離解析法，發展整體問答契合度解析法，以解析得最終提問與回答間之整體契合程度</li> </ul>	
			Cilibrasi 與 Vitanyi (2007)		
成員個人資訊 解析	成員資料分群法	以頻繁項目集為基分群法 (擴展既有研究方法)	Zhang 等人 (2010)	<ul style="list-style-type: none"> <li>● 結合斷詞與有效詞篩選法，將原文章分群法，延伸並擴展為虛擬社群成員資料分群法</li> <li>● 以內文及語意相似度解析法為基，並輔以群集相似度計算法，以擴展為群集內文及語意相似度解析法，藉以應用於成員參考度之分析</li> </ul>	
	成員參考度分析	內文相似度解析法； 詞彙語意解析法 (擴展並整合既有研究方法)	Salton 等人 (1975)		
			Cilibrasi 與 Vitanyi (2007)		
			Zhang 等人 (2010)		

**備註：**「擴展既有研究方法」係指將過去言就尚未考量之處，透過本研究所發展之啟發式演算法進行擴展；「改良既有研究方法」係以過去演算法為基，進行改良，以提升解析之效能與效率（正確率）；「創新研究方法」係全為本研究所發展之啟發式演算法；「應用既有研究方法」係指經簡易轉化後，所直接應用之演算法；「整合既有研究方法」係以本研究發展之啟發式演算法，針對各項技術進行整合，所發展得之整合型演算法。

## 第四章、系統架構

根據第三章所發展之方法論與模式，本研究乃開發一套虛擬社群知識群聚度判定系統，以確認方法論與模式之可行性。本研究將系統主推論模組劃分為「虛擬社群問答契合度解析」與「虛擬社群成員參考度解析」等兩模組，以分別針對上述兩面向之議題進行解析，於前者中，主要以發問者之「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，以判定「提問品質」、「回答品質」、「問答內文契合度」與「問答語意契合度」，並於整合數據後分析與計算得「目標虛擬社群討論串提問與回答之整體問答契合度得分」；於後者中，則以社群成員之「歷史發文」為基礎，透過對目標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，當中群集間相似程度之解析，以取得相關領域性群集間之內文與語意關係距離，並於正規化後以獲致實具參考價值之目標社群成員領域知識可參考程度。藉由上述兩參考數據之提供，於社群管理者之層面，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，於知識需求者之層面，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。本章即針對本研究所提之「虛擬社群知識群聚度判定系統」，分別以系統核心架構、系統功能架構、資料模式定義、系統流程與系統開發工具進行說明。

### 4.1 虛擬社群知識群聚度判定系統之核心架構

本研究所開發之「虛擬社群知識群聚度判定系統」依其運作流程可分為「討論串資料上傳」、「領域文件上傳」、「成員歷史發文上傳」、「資料預處理」、「問答契合度之解析」，以及「成員參考度之解析」等六大階段，本系統之運作流程架構如圖4.1所示，各功能層次之詳細流程說明如下。

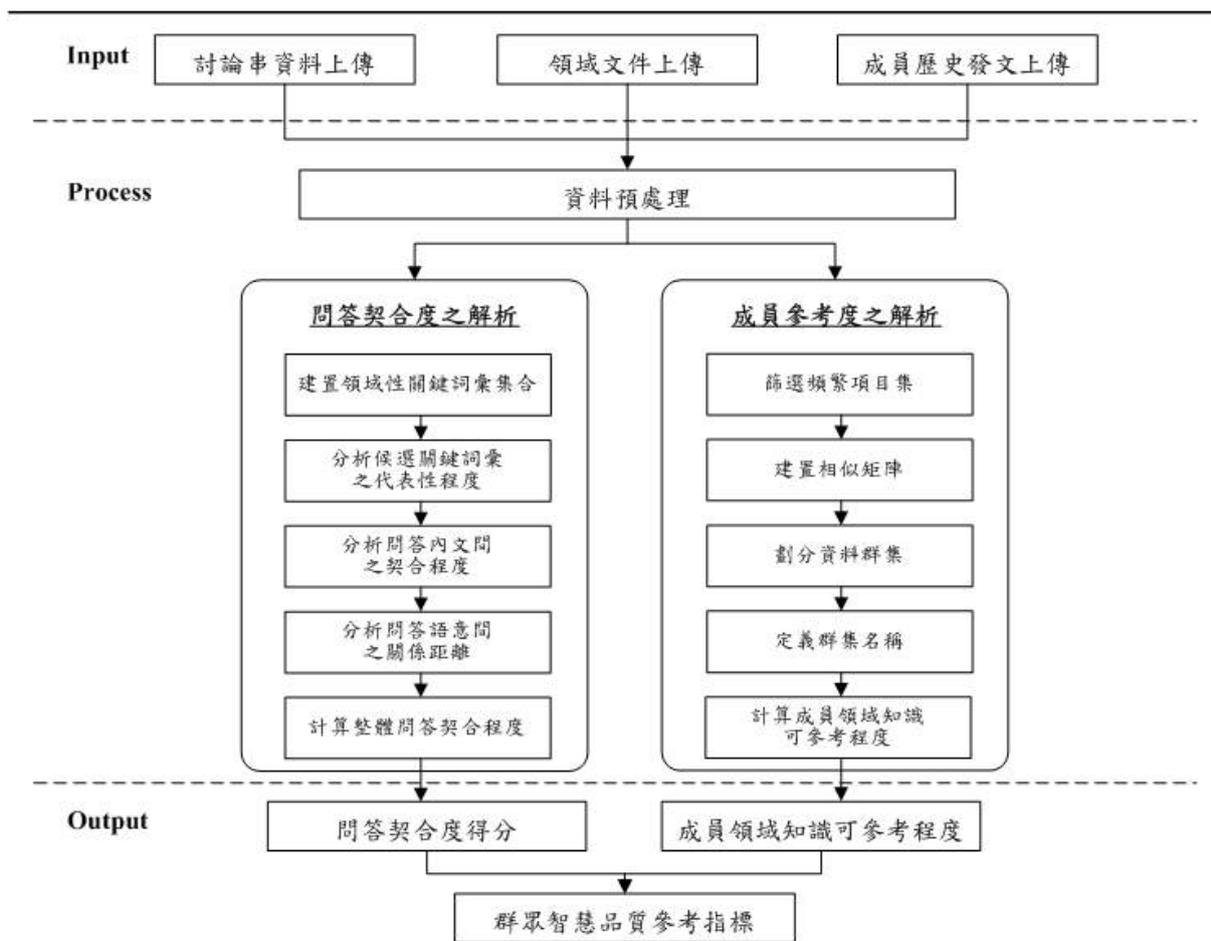


圖 4.1、虛擬社群知識群聚度判定系統之流程架構

- 領域文件上傳

系統使用者可蒐集相關領域之訓練文件（如虛擬社群領域性文章），並將文件上傳至系統，系統即擷取目標領域文件之內文，於透過「資料預處理」後，將以「虛擬社群問答契合度解析模組」建立領域性關鍵詞彙集合，並作為「虛擬社群成員參考度解析模組」中成員領域知識可參考程度解析之依據。

- 討論串資料上傳

系統使用者將討論串資料上傳至系統，系統即擷取此目標討論串之分析資訊，包含討論串標題、討論串所屬領域、提問內容及回答內容等，於透過「資料預處理」後，將以「虛擬社群問答契合度解析模組」進行問答契合度之解析，以取得目標討論串中各筆回答內容與提問間之問答契合度得分。

- 成員歷史發文上傳

系統使用者將虛擬社群成員之歷史發文資料上傳至系統，系統將擷取此目標成員之歷史發文資訊，包含歷史發文內容及歷史發文所屬領域，於透過「資料預處理」後，將以「虛擬社群成員參考度解析模組」進行成員參考度之解析，以取得目標成員領域知識可參考程度之數據。

- 資料預處理

此部份乃透過中研院所開發之 CKIP 斷詞系統，針對討論串資料、領域文件，以及成員歷史發文中之內文進行斷詞，並於進行有效詞之篩選後，將相關斷詞及有效詞篩選結果予以儲存，以進行後續之應用，並視為相關資料預處理之作業。

- 問答契合度之解析

虛擬社群問答契合度解析模組可分為「領域性關鍵詞彙集合建立」與「問答契合度解析」等兩大功能。於「領域性關鍵詞彙集合建立」功能中，可藉領域文件之匯入，於分析詞彙間之語意關係距離，並進行門檻之篩選後，以建置相關領域之關鍵詞彙集合；於「問答契合度解析」功能中，將於分析討論串中各詞彙之代表性程度後，以內文相似度與語意關係距離等兩層面，進行整合性整體問答契合度之計算，以取得實具參考性提問及回答間之契合度分數。

- 成員參考度之解析

虛擬社群成員參考度解析模組可分為「成員資料分群」，以及「成員參考度解析」等三大功能。於「成員資料分群」功能中，將根據相似矩陣取得分群類型，並透過頻繁項目集間之關係，以劃分資料之群集並定義群集之名稱；於「成員參考度解析」中，將於分析並整合群集間內文相似度及語意關係距離後，以取得成員領域知識可參考程度之數據。

## 4.2 系統功能架構

本研究所建置虛擬社群知識群聚度判定系統乃架構於網際網路環境下。使用者可透過網際網路登入本系統，並使用本系統所提供之各項功能。當使用者登入系統後，系統即根據使用者帳號判斷該使用者於系統中之功能權限。

在本系統平台之權限管理架構下乃將系統使用者區分為一般使用者與系統管理者，以下即分別針對此兩種不同身份使用者所能使用之功能加以說明：

### 一般使用者

1. 可新增與上傳未進行問答契合度解析之討論串內容、未進行成員參考度解析之社群成員，及其歷史發文資料至系統。
2. 可瀏覽符合使用者上傳之所有討論串，以及成員歷史發文之內容。
3. 可瀏覽符合使用者所上傳討論串與社群成員，經分析後所得問答契合度，以及結果成員參考度解析之結果。

### 系統管理者

1. 可上傳未經領域性關鍵詞彙分析之領域文件至系統
2. 可查詢、新增、修改或刪除領域文件內容
3. 可瀏覽/編輯系統資料庫內之所有討論串、社群成員，以及歷史發文等資料
4. 可執行領域性關鍵詞彙集合建立
5. 可修改/查詢系統參數與門檻值
6. 可執行問答契合度解析
7. 可執行成員資料分群
8. 可執行成員參考度解析

本系統所開發之重點模組共有「虛擬社群資料維護模組」、「虛擬社群問答契合度解析模組」、「虛擬社群成員參考度解析模組」，以及「系統參數設定模組」等四大模組；

圖4.2即表示虛擬社群知識群聚度判定系統之核心模組架構。

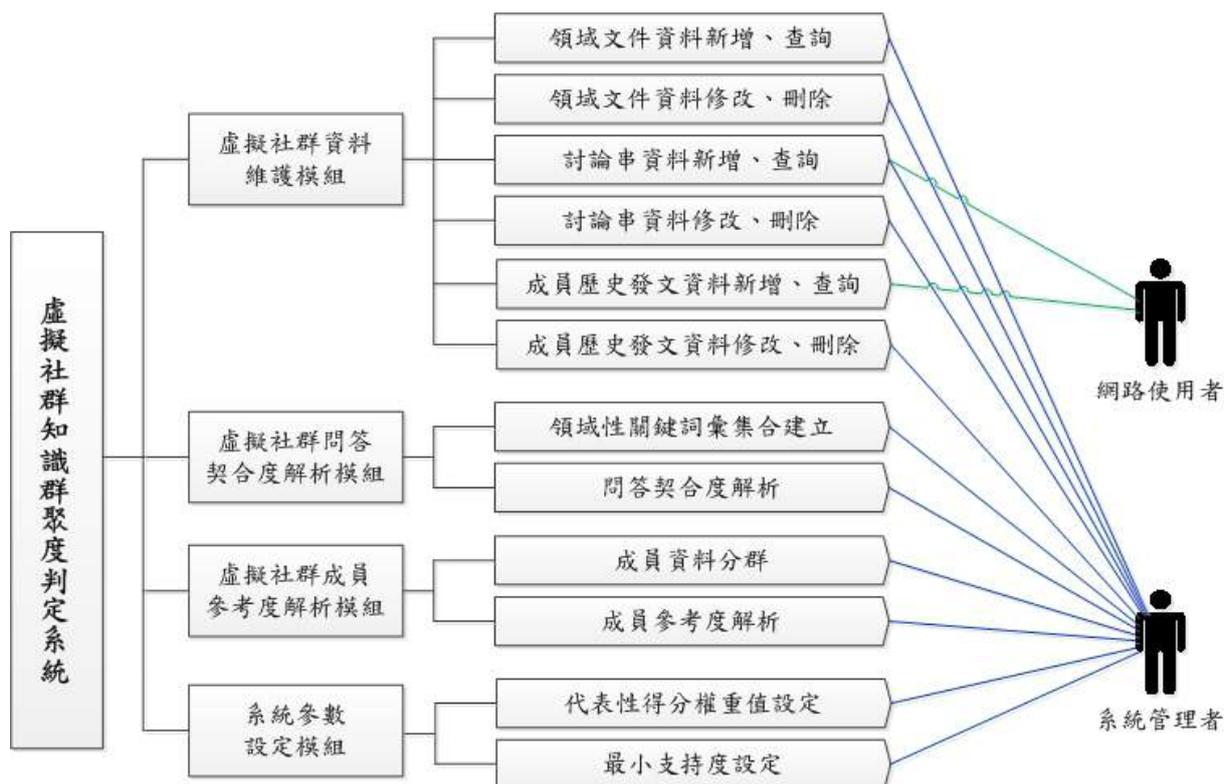


圖 4.2、虛擬社群知識群聚度判定系統之功能架構

針對上述系統架構所包含之基本功能模組說明如下：

(一) 虛擬社群資料維護模組

- 領域文件資料新增：提供系統管理者將未經領域性關鍵詞彙分析之領域文件，匯入並維護於系統資料表中。
- 領域文件資料查詢：提供系統管理者進行已維護領域文件資料之查詢。
- 討論串與成員歷史發文資料新增：提供系統管理者與一般使用者將未判定討論串，以及成員歷史發文資料，匯入並維護於系統資料表中。
- 討論串與成員歷史發文資料查詢：提供系統管理者與一般使用者進行已維護討論串，以及成員歷史發文資料之查詢。
- 領域文件、討論串與成員歷史發文資料修改：提供系統管理者修改錯誤之領域文件、討論串，以及成員歷史發文資料。
- 領域文件、討論串與成員歷史發文資料刪除：提供系統管理者刪除錯誤之領域文件、討論串，以及成員歷史發文資料。

## (二) 虛擬社群問答契合度解析模組

- 領域性關鍵詞彙集合建立功能：系統管理者可針對所有領域類別，進行領域性關鍵詞彙集合之建立。於管理者選定欲建置之領域類別後，即可根據目標類別中之所有領域文件，分析詞彙間之語意關係距離，並進行門檻之篩選，以建置相關領域之關鍵詞彙集合。
- 問答契合度解析功能：系統管理者可輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後選擇欲解析之討論串，即可針對目標討論串中之所有提問與回答，進行整合性整體問答契合度之計算，以取得問答契合度分數。

## (三) 虛擬社群成員參考度解析模組

- 成員資料分群功能：系統管理者可輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後選擇欲解析之成員，即可針對目標成員之所有歷史發文內容，以相似矩陣與頻繁項目集為基礎，劃分群集並定義群集之名稱。
- 成員參考度解析功能：系統管理者可輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後選擇欲解析之成員，即可針對已完成分群目標成員之群集，進行群集間內文相似度及語意關係距離之分析，以取得成員領域知識可參考程度之數據。

## (四) 系統參數設定模組

- 代表性得分權重值參數設定：提供權限內使用者進行詞類距離採納值、前後文配對及標題加權整體分數比重值，以及詞類相關性平均分數比重值等參數之設定，以提升「問答契合度解析」功能，提問與回答間契合程度分析之準確性。
- 最小支持度參數設定：提供權限內使用者進行最小支持度參數之設定，以提升「成員資料分群」功能，目標成員歷史發文資料分群之準確性。

本系統之使用者可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。針對一般使用者可執行討論串，以及成員歷史發文等資料之新增與查詢功能（如圖4.3之一般使用者所示），而系統管理者則可進行領域文件、討論串，以及成員歷史發文等資料之維護（包含新增、查詢、修改與刪除）、系統參數設定、虛擬社群問答契合度解析，以及虛擬社群成員參考度解析等（如圖4.3之系統管理者所示）。

首先，當使用者上傳未解析之討論串，以及成員歷史發文資料後，系統管理者乃需完成參數設定以及領域文件上傳，待設定與上傳完畢後，系統管理者即可執行虛擬社群文達契合度解析模組以建置領域性關鍵詞彙集合，並根據關鍵詞彙集合中詞彙間之關聯，以及虛擬社群討論串中問答間之關係，以代表性程度、問答內文契合程度，以及問答語意契合程度等方法論，透過系統自動分析並計算得整體問答契合程度；另一方面，系統管理者則透過虛擬社群成員參考度解析模組，先行進行虛擬社群成員之資料進行分群，再根據分群得各群集間，以及虛擬社群討論串間之關係，以解析並計算得虛擬社群成員參考度；最後，使用者即藉由虛擬社群資料維護模組之查詢功能，檢視系統中分析得討論串之整體問答契合程度，以及虛擬社群成員之成員參考度等結果。綜合上述，本研究將虛擬社群知識群聚度判定系統之運作架構繪製如圖4.3。

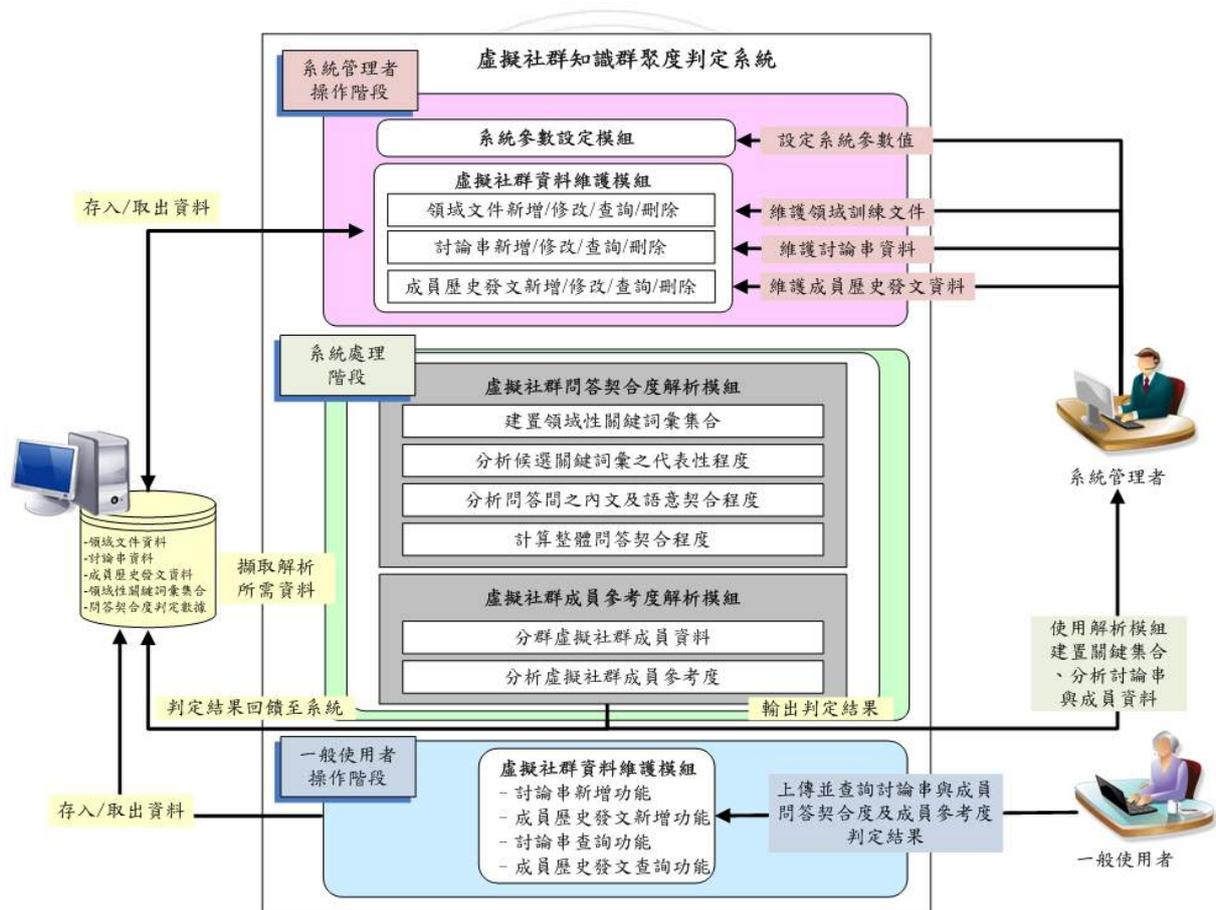


圖 4.3、虛擬社群知識群聚度判定系統運作架構

### 4.3 資料模式定義

本研究發展之虛擬社群知識群聚度判定系統乃以網際網路環境為基礎，並配合資料庫技術以開發系統之各項功能，期使虛擬社群資料管理、虛擬社群問答契合度解析與虛擬社群成員參考度解析等任務可即時完成。依據系統運作之需要，將虛擬社群知識群聚度判定系統之資料分為「領域性關鍵詞彙集合資料」、「虛擬社群問答契合度資料」、「虛擬社群成員參考度資料」與「系統參數資料」等四大部分，以下即就各部分所包含之資料表說明其資料定義。

#### (一) 領域性關鍵詞彙集合資料

此資料之目的乃記錄領域文件相關資料內容、領域類別相關資訊內容，以及解析時所需關鍵資訊，以有效進行領域資料管理、領域類別資訊維護，以及建置領域性關鍵詞彙集合；其所屬之資料表及其相關定義說明如下：

- 領域文件基本資料表：記錄領域文件之基本資料，如領域文件編號、領域文件標題、領域文件內容，以及領域類別編號等資訊。
- 領域文件斷詞結果資料表：記錄經斷詞所得領域文件之斷詞結果。
- 領域文件有效詞資料表：記錄經有效詞篩選所得領域文件之有效詞集。
- 領域類別資料表：記錄所有領域類別之編號，以及名稱等資訊。
- 有效詞彙組資料表：記錄有效詞彙組之相關資料，如有效詞彙組編號、主要有效詞名稱、主要有效詞搜尋結果數、次要有效詞名稱、次要有效詞搜尋結果數、同時包含之搜尋結果數，以及領域類別編號等資訊。
- 語意/整體關係資料表：記錄經分析所得有效詞彙組之語意關係距離，以及整體關係距離，以作為建置領域性關鍵詞彙集合之依據。
- 領域關鍵集合資料表：記錄經判定所得符合門檻領域詞彙之編號、名稱與所屬類別編號，以作為虛擬社群問答契合度解析模組應用之基礎。

#### (二) 虛擬社群問答契合度資料

此資料之目的乃記錄虛擬社群問答契合度相關之資料內容，包含討論串資料、問答資料，以及當中解析時所需之關鍵資訊，以有效進行分析資料之管理，以及問答契合度之解析；其所屬之資料表及其相關定義說明如下：

- 討論串/提問資料表：記錄討論串中提問之基本資料，如討論串編號、提問標題、提問內容，以及領域類別編號等資訊。
- 提問斷詞結果資料表：記錄經斷詞所得提問之斷詞結果。
- 提問有效詞資料表：記錄經有效詞篩選所得提問之有效詞集。
- 關鍵提問詞彙資料表：記錄經比對後，符合領域性關鍵詞彙集合，關鍵提問詞彙之編號、名稱，以及討論串編號等資訊。
- 關鍵提問詞彙前後文資料表：記錄以關鍵提問詞彙為基礎，經分割且斷詞後所得前後文之分割內容，包含提問前後文編號、提問前文內容、提問前文斷詞結果、提問後文內容、提問後文斷詞結果，以及提問前推算集合，且同時記錄經分析後所得之提問前後文整體分數，以及提問詞類相關性平均分數，以作為計算提問品質得分之依據。
- 提問品質得分資料表：記錄經計算後所得提問之提問品質得分，以作為解析整體問答契合程度之基礎。
- 回答資料表：記錄討論串中回答之基本資料，如回答編號、回答內容，以及所屬之討論串編號等資訊。
- 回答斷詞結果資料表：記錄經斷詞所得回答之斷詞結果。
- 回答有效詞資料表：記錄經有效詞篩選所得回答之有效詞集。
- 關鍵回答詞彙資料表：記錄經比對後，符合領域性關鍵詞彙集合，關鍵回答詞彙之編號、名稱，以及所屬討論串編號等資訊。
- 關鍵回答詞彙前後文資料表：記錄以關鍵回答詞彙為基礎，經分割且斷詞後所得前後文之分割內容，包含回答前後文編號、回答前文內容、回答前文斷詞結果、回答後文內容、回答後文斷詞結果，以及回答前推算集合，且同時記錄經分析後所得之回答前後文整體分數，以及回答詞類相關性平均分數，以作為計算回答品質得分之依據。
- 回答品質得分資料表：記錄經計算後所得回答之回答品質得分，以作為解析整體問答契合程度之基礎。
- 問答內文契合度資料表：記錄以問答斷詞結果間之關係為基礎，所計算得之問答內文契合度得分，以作為解析整體問答契合程度之基礎。
- 問答語意關係資料表：記錄以問答有效詞集合間之關係距離為基礎，所分析得之問答語意關係，以作為解析整體問答契合程度之基礎。

- 整體問答契合程度資料表：記錄以提問品質得分、回答品質得分、問答語意關係距離，以及問答語意關係距離為基礎，所解析得之整體問答契合程度。

### (三) 虛擬成員參考度資料

此資料之目的乃記錄虛擬社群成員參考度相關之資料內容，包含社群成員資料、歷史文章資料、領域討論串資料，以及當中解析時所需之關鍵資訊，以有效進行分析資料之管理，以及成員參考度之解析；其所屬之資料表及其相關定義說明如下：

- 領域討論串資料表：記錄領域討論串之基本資料，如領域討論串編號、領域討論串標題、領域討論串內容，以及領域類別編號等資訊。
- 討論串斷詞結果資料表：記錄經斷詞所得領域討論串之斷詞結果。
- 討論串有效詞資料表：記錄經有效詞篩選所得領域討論串之有效詞集。
- 社群成員基本資料表：記錄社群成員之基本資料，如社群成員編號，以及社群成員名稱等資訊。
- 歷史文章資料表：記錄經斷詞所得歷史文章之斷詞結果。
- 歷史文章斷詞結果資料表：記錄經斷詞所得領域討論串之斷詞結果。
- 歷史文章有效詞資料表：記錄經有效詞篩選所得歷史文章之有效詞集。
- 歷史文章/群集資料表：記錄經分群後所得文章與群集間之關係，亦即文章編號所對應之群集編號。
- 群集名稱資料表：記錄以群集間詞彙之關係為基礎，所分析所得群集編號所屬之群集名稱。
- 社群成員/參考度資料表：記錄以群集間各文章之內文及語意關係為基礎，所分析得目標社群成員之領域知識可參考程度，該領域所屬之領域類別編號。

### (四) 系統參數資料

此資料之目的乃記錄系統參數之資料，如分析比重值、採納值與支持度門檻等資料，透過設定以有效提高虛擬社群問答契合度解析，以及虛擬社群成員參考度解析之準確率；其所屬資料表及其相關定義如下：

- 系統參數資料表：記錄系統參數值，以影響判定模組數據結果之準確率。

上述各資料乃為系統中各功能模組所需使用或產生之各項資訊，並依其所規劃之資料表形式記錄於資料庫中，用以支援系統各功能模組執行其任務。此外，透過各項資料表間之關聯性（Entity Relationship Model；ER Model）設計，使本研究所發展之虛擬社群知識群聚度判定系統可方便地進行虛擬社群資料，以及分析資料控管，並有效提升系統之彈性、效率性與正確性。各資料表間之關聯性如圖4.4所示。



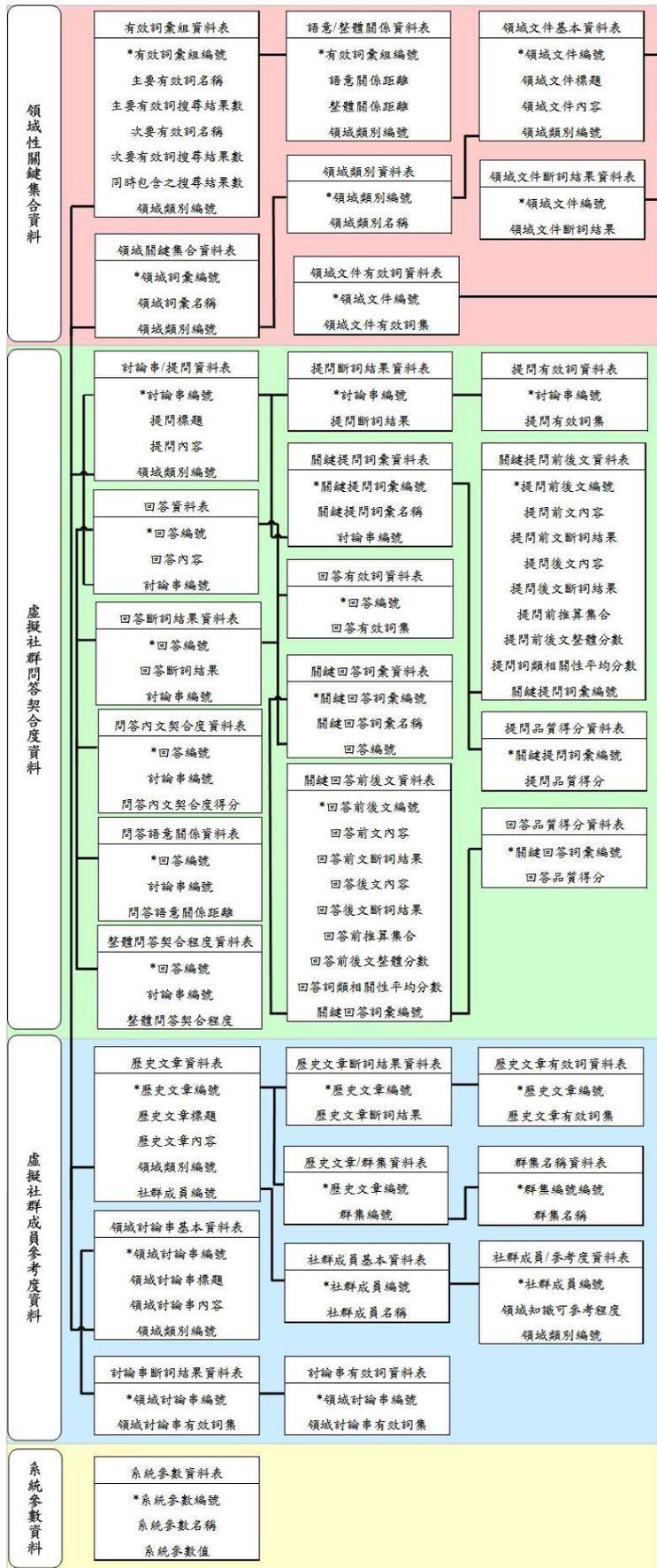


圖 4.4、虛擬社群知識群聚度判定系統之資料關聯

## 4.4 系統流程

本節乃針對「系統功能流程」與「系統資料流程」兩部分進行說明；其中，系統功能流程將介紹使用者於各功能模組之功能流程規劃，而系統資料流程則介紹系統內各項資料傳遞之流程關係。

### 4.4.1 系統功能流程

如 4.2 節所述，本系統實際運作乃依不同功能進行區分，包括「虛擬社群資料維護模組」、「虛擬社群問答契合度解析模組」、「虛擬社群成員參考度解析模組」及「系統參數設定」等四大模組，以下即說明各系統功能之流程規劃。

#### 虛擬社群資料維護模組

此模組可供權限內使用者上傳領域訓練文件、討論串與成員歷史發文，以作為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之分析資料。此外，權限內使用者亦可根據系統中所維護之領域訓練文件、討論串與成員歷史發文等內容，透過領域文件維護、討論串維護，以及成員歷史發文維護等功能中，各項子功能包含新增、查詢、修改與刪除等子功能，進行相關資料之維護動作；其流程設計概念分別如圖 4.5、圖 4.6 與圖 4.7 所示。

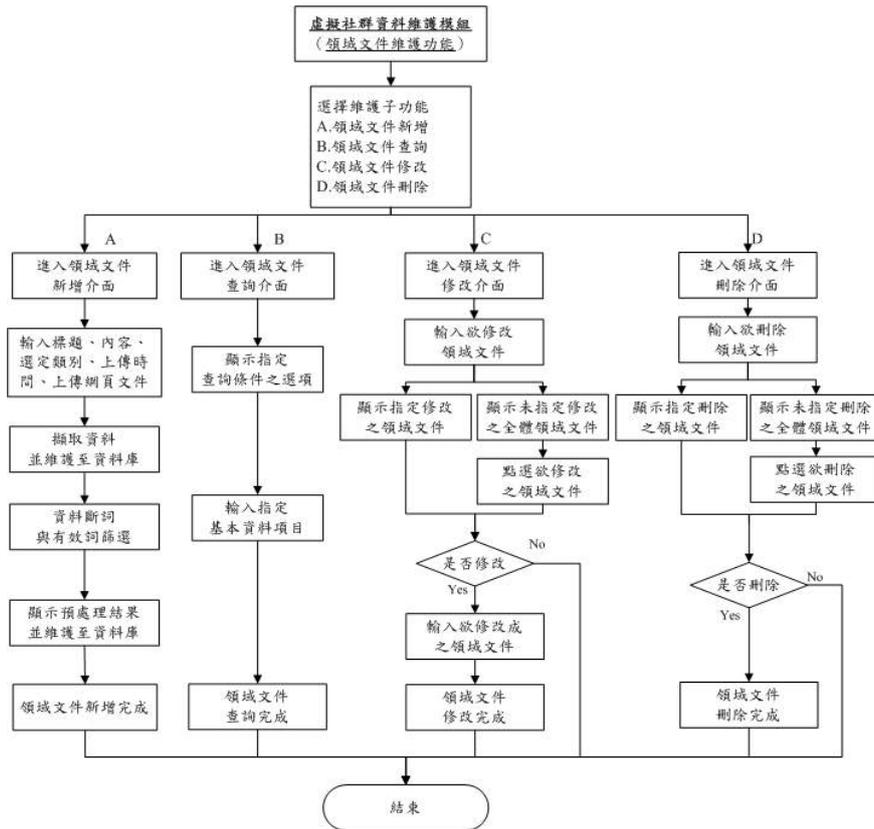


圖 4.5、虛擬社群資料維護模組—領域文件資料維護之功能流程

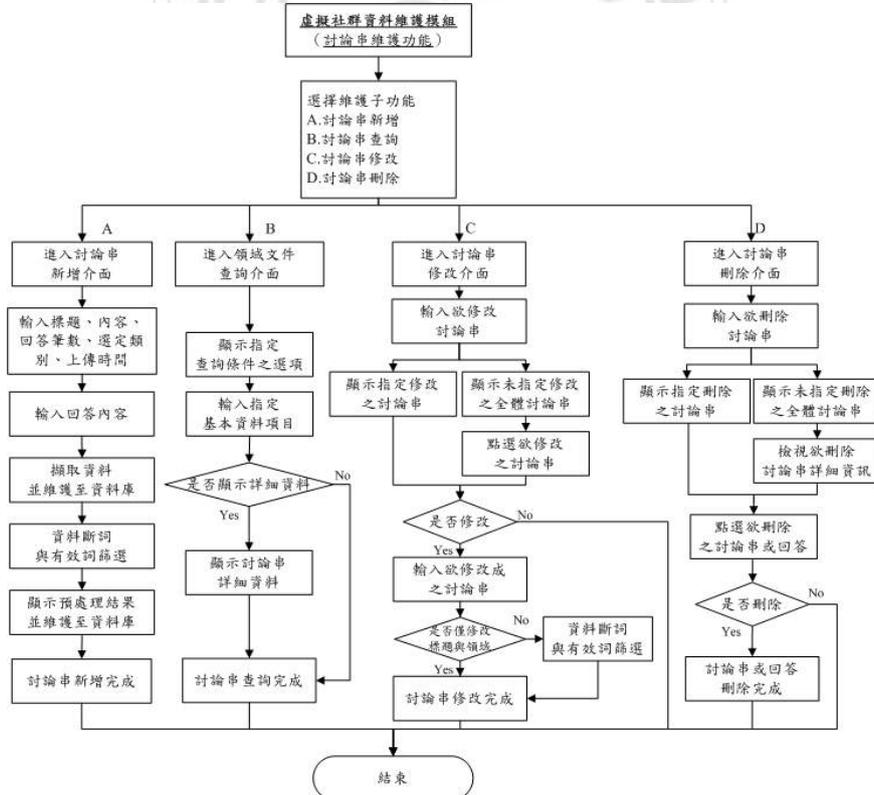


圖 4.6、虛擬社群資料維護模組—討論串維護之功能流程

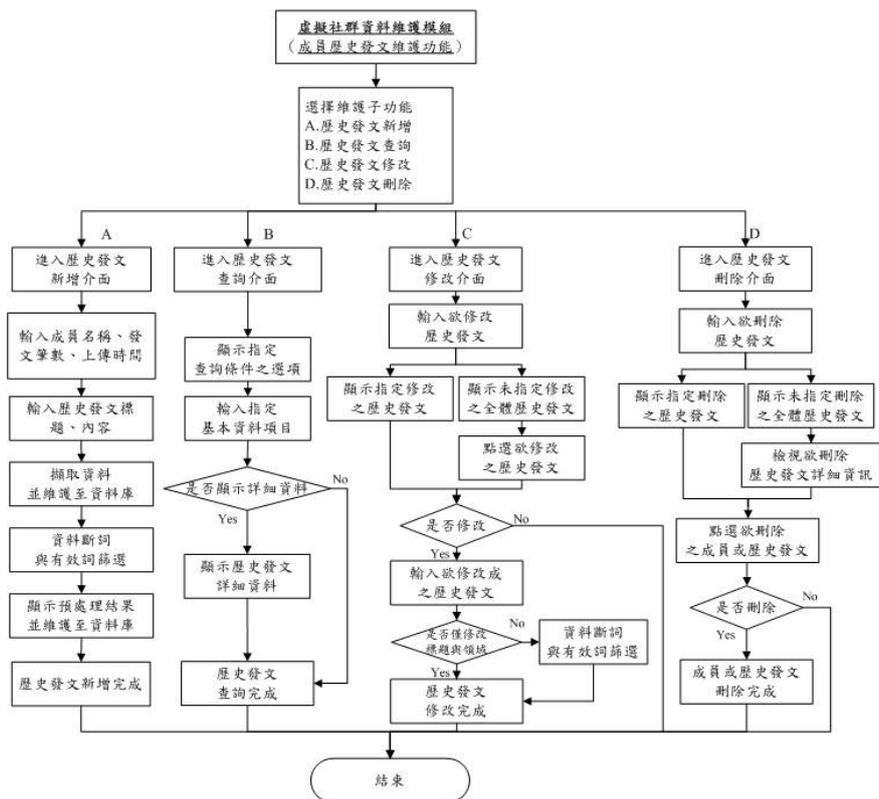


圖 4.7、虛擬社群資料維護模組—成員歷史發文維護之功能流程

### 虛擬社群問答契合度解析模組

本模組乃將此模組劃分為「領域性關鍵詞彙集合建立功能」與「問答契合度解析功能」兩大功能。於「領域性關鍵詞彙集合建立功能」中，首先，本功能根據蒐集之領域文件資料，進行內文斷詞及領域有效詞篩選之自動化資料預處理作業，次之，本功能將以領域有效詞之篩選結果為依據，計算各領域有效詞組間之關係距離，接著，即根據上述計算得之關係距離數據，分別以四分位數、中位數、平均值與使用者自訂等門檻標準，分析並審視領域有效詞組達成之狀況，並於最終將符合門檻之領域有效詞彙組，各自分離成獨立之詞彙，以用於建構領域性關鍵詞彙集合，整體運作流程如圖 4.8 左半部所示。

於「問答契合度解析功能」中，主要係解析已維護於資料庫之討論串，首先，本功能針對討論串進行內文斷詞、有效詞篩選，以及關鍵詞彙篩選（以領域性關鍵詞彙集合為依據）等資料預處理作業，次之，本功能以關鍵詞彙為中心進行前後文語句之分割與斷詞，並計算相關之前後文配對及標題加權整體分數，接著，即依據詞類距離採納值建置前推算集合，分析當中詞類相關性平均分數，並根據設置之比重值計算得關鍵詞彙之代表性得分，且同時以出現頻率為基礎解析得提問與回答之品質，爾後，本功能除分析問答內文之契合度外，亦針對問答間之語意關係距離進行解析，最終，整合於提問品質、

問答品質、問答內文契合度，以及問答語意關係距離等數據，本功能即針對目標提問與回答計算得整體問答契合度得分，並於正規化後獲致整體問答之契合程度，整體運作流程如圖 4.8 中右半部所示。

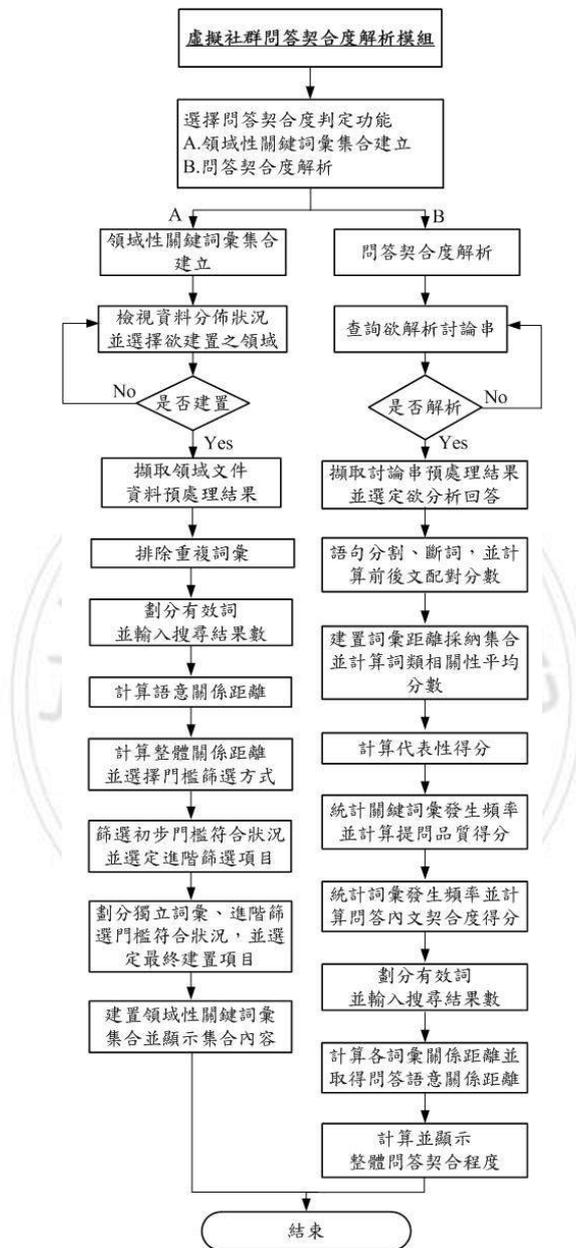


圖 4.8、虛擬社群問答契合度解析模組之功能流程

### 虛擬社群成員參考度解析模組

本模組乃以目標虛擬社群成員「歷史發文」資料，以及「歷史發文所屬領域」為基礎，解析目標成員領域發文間群集之相似度，並同時藉其他成員間相似數據之比較，以於最終透過正規化之方式，取得目標虛擬社群成員領域知識之可參考程度，因此，本研

究將此模組劃分為「成員資料分群功能」與「成員參考度解析功能」兩大功能；其流程設計概念分別如圖 4.9 所示。

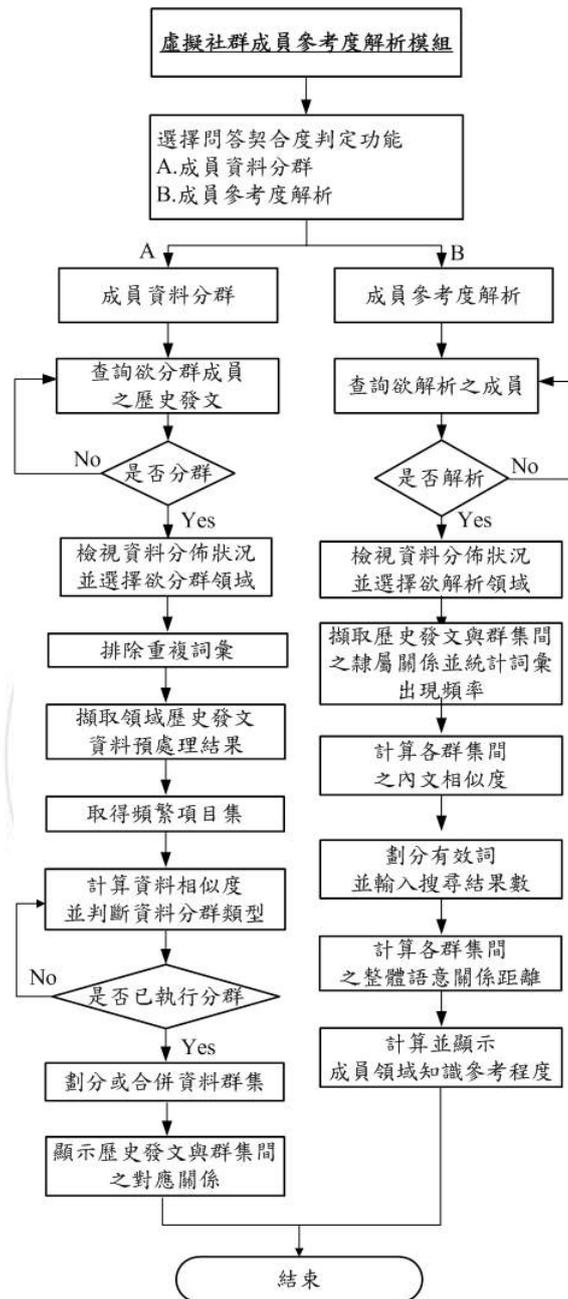


圖 4.9、虛擬社群成員參考度解析模組之功能流程

### 系統參數設定模組

為使系統管理者方便維護各系統相關資料，此模組乃提供系統管理者於線上修改各系統參數資料，包含詞類距離採納值、前後文配對及標題加權整體分數比重值、詞類相關性平均分數比重值，以及最小支持度，進而保持系統門檻值與權重值之正確性，其流

程設計概念如圖 4.10 所示。

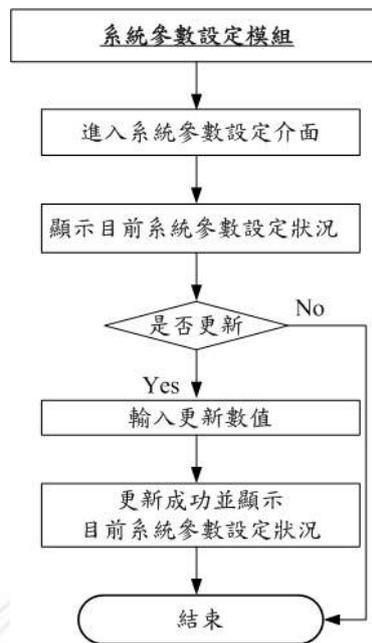


圖 4.10、系統參數設定模組之功能流程

#### 4.4.2 系統資料流程

本系統運作之初，系統管理者將未解析之領域文件匯入系統，並進行領域性關鍵詞彙集合之建立，待系統管理者完成各領域關鍵詞彙集合之建置後，須先行針對問答契合度與成員參考度解析之系統參數進行設定，爾後，一般使用者即可將未解析之討論串資料匯入系統，接著，系統管理者即針對使用者上傳之討論串，執行虛擬社群問答契合度解析模組，以分析並取得各討論串中提問與回答之契合程度，其次，一般使用者亦可上傳未經解析之成員歷史發文資料至系統，待上傳完畢後，系統管理者即針對使用者上傳之成員資料，執行虛擬社群成員參考度解析模組，以計算並判定得虛擬社群成員領域知識可參考之程度，最後，上述分析得之所有判定結果亦將同時儲存於系統資料庫中。上述步驟完成後，一般使用者即可以討論串及成員歷史發文查詢功能進行查詢，以取得系統判定問答契合度及成員參考度之最終結果，本系統相關資料之存取與傳遞情形如圖 4.11 所示。

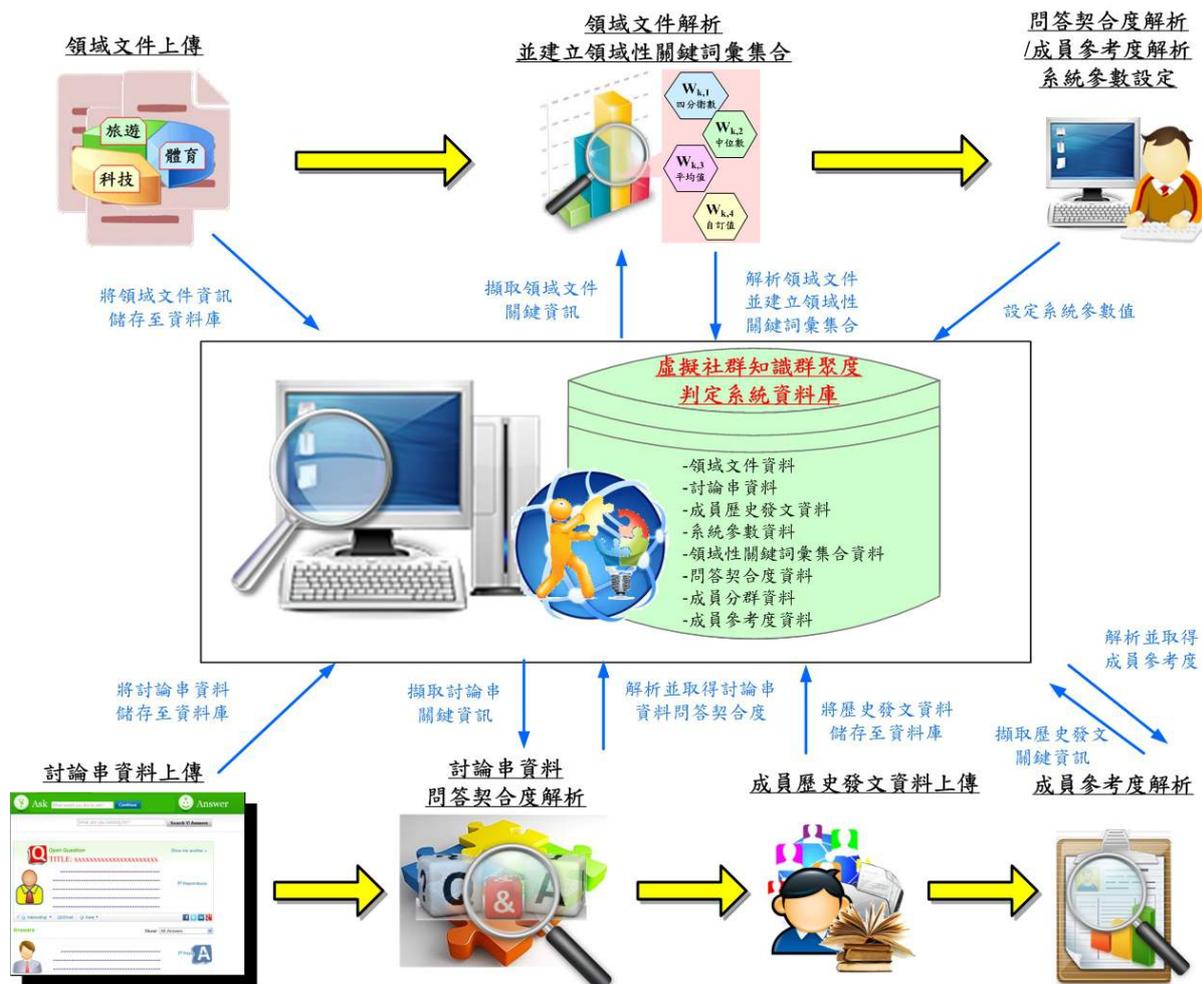


圖 4.11、系統資料流程

#### 4.5 系統開發工具

本系統乃建置於 Microsoft Windows 7 作業系統上，並以 Microsoft SQL Server 2008 資料庫系統儲存系統運作過程相關資料。系統之操作介面與內部演算法則採用 JSP (Java Server Pages) 語法進行開發，並利用 SQL Server 2008 來存取資料庫，輔助系統進行知識文件分析。以下即分別介紹系統開發時所使用之工具。

##### ➤ JSP (Java Server Pages)

JSP 是由 Sun Microsystem 公司所倡導之網站伺服器描述語言程式，其乃以 Java 程式語言為基礎，並集結其他公司所共同建立之動態網頁技術標準，故具有 Java 支援跨平台與跨網站伺服器之優點，而使網頁設計更具彈性。

當使用者透過瀏覽器向伺服器端 (Server) 要求開啟 JSP 網頁時，架設於伺服器端之 JSP 引擎乃先將 JSP 網頁轉譯為 Servlet 程式，其次再將 JSP 執行後所產生之文件資料

傳送至用戶端 (Client)，並同時顯示執行結果於瀏覽器上。此外，JSP還具有下列特性 (詹權恩，2004)：

- 瀏覽者端環境：各種網頁瀏覽器均可，如IE、Google Chrome、Fire Fox、KK Man。
- 模組程式的可重用性：JSP元件 (Enterprise JavaBeans) 可跨平台重複使用於任何地方。Enterprise JavaBeans 元件可存取傳統資料庫，並能以分散式系統模式於Unix 和 Windows 平台工作，減少程式開發之時間並可增加程式之彈性。
- 保護原始程式碼：延伸名為\*.jsp的JSP程式碼並未顯示於Client端之瀏覽器上。
- 跨平台性：JSP可執行於任何具有Web伺服器之環境，並支援多數作業系統。
- 標籤可擴充性：由於JSP技術兼容XML標籤技術，故JSP開發者可擴展JSP標籤或制訂標籤庫，以減少對Scripting語言之依賴，並降低網頁製作者製作網頁和擴充網頁功能之複雜程度。
- 伺服器端環境：Windows XP，並加上「J2SDK」Java程式編譯工具與Tomcat等JSP伺服器。與HTML緊密整合：將JSP融入HTML標籤中，不僅提高便利性亦減少I/O問題，兼具可取代傳統CGI等直譯式語言。
- 平台和伺服器的獨立性：JSP技術一次寫入之後，可以在任何具有符合JavaTM語法結構的環境下執行。
- 伺服器端搭配資料庫：SQL Server資料庫系統。

#### ➤ 關聯式資料庫-Microsoft SQL Server 2008

Microsoft SQL Server 2008為一種關聯式資料庫 (Relational Database Management Systems, RDBMS) 管理系統，其擁有高彈性與多元化之架構，可安裝於主從式架構之作業系統平台上或獨立伺服器主機。關聯式資料庫乃將資料分類儲存於多個二維表格中，此表格通稱為資料表。之後再利用兩資料表間之關聯以查詢相關資料。其優點在於各個資料表均可獨立運作，當進行資料之新增、修改或刪除時，亦不會互相影響。系統管理員可透過應用程式進入伺服器，更改資料型態，並管理及處理伺服器資源。此種資料庫常使用SQL (Structured Query Language) 語法進行資料查詢，SQL語法可用以查詢資料庫、建立新表格、更新與刪除資料，並設定資料庫權限。

綜上所述，本研究乃利用上述工具進行系統開發工作，並將系統架構於Web環境下，以開發4.2節所述之各項系統功能。

## 第五章、系統實作與案例分析

根據第三章所發展之方法論與模式，以及第四章所提出之系統架構與規劃，本研究乃開發一套虛擬社群知識群聚度判定系統，以確認方法論與系統之可行性，當中，各功能模組之操作說明乃彙整於附錄，而為驗證本研究發展方法論與系統於實務應用之可行性及績效，本研究以「搜搜問問」虛擬社群之真實問答資料作為驗證（測試）資料之樣本，並以該虛擬社群中之精華知識文章（不與驗證資料相互重疊），作為領域訓練資料之樣本，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（詳述於 5.2 節中），以驗證本研究發展方法論與系統之可行性及績效。

### 5.1 系統案例之應用流程

為驗證本系統於實務應用環境中之可行性，本研究乃分別針對「虛擬社群問答契合度解析」與「虛擬社群成員參考度解析」兩模組進行真實案例應用之驗證。於整體系統測試資料樣本中，本研究乃針對「搜搜問問」虛擬社群中之真實問答資料，以及成員歷史評價數據，進行測試資料之蒐集，從中隨機蒐集並匯入多筆資料於系統中，藉由本系統問答契合度與成員參考度解析等核心模組之運行，以評估本系統於判定虛擬社群知識群聚度，包含問答契合度及成員參考度之正確性；於領域訓練資料樣本中，本研究則針對「搜搜問問」之精華知識文章進行蒐集，透過本系統「領域性關鍵詞彙集合建置」功能之運行，由領域訓練資料中擷取關鍵詞彙，建置相關領域性關鍵詞彙集合，以作為判定虛擬社群知識群聚度之解析基礎，另一方面，本研究亦以蒐集得之文章，作為領域性群集關係距離分析之依據，以判定虛擬成員領域知識可參考程度。

首先，系統管理者必須蒐集領域訓練資料，將領域訓練資料上傳至系統中，以透過「領域性關鍵詞彙集合建置」功能建立相關之領域性關鍵詞彙集合，此外，系統管理者亦須針對系統中之各項參數值進行相關之設定。其次，一般使用者乃上傳多筆未判定之討論串（包含提問與回答）內容至系統中。接著，系統管理者將執行核心模組，以判定討論串內容中之問答契合度，以及成員之可參考程度，待分析完畢後，系統即輸出相關判定結果輸出予管理者，並同時將這些結果回饋並儲存於系統中。最後，使用者即可查詢得系統中已完成判定討論串中問答之契合度，以及成員之可參考程度。上述完整運作之架構如圖 5.1 所示，以下即進行系統應用情境之詳細說明。



圖 5.1、虛擬社群知識群聚度判定系統之應用流程

### ■ 系統管理者蒐集領域訓練文件

開始執行本系統前，系統管理者須蒐集相關領域之訓練文件，於此系統管理者針對「搜搜問問」進行領域訓練文件之蒐集（如圖 5.2 所示），並於擷取領域訓練文件中之問答內容後，輸入於系統中，以透過「領域性關鍵詞彙集合建置」功能建立相關之領域性關鍵詞彙集合，以作為判定虛擬社群知識群聚度之解析基礎，另一方面，領域訓練文件亦可作為領域性群集關係距離分析之依據，以判定虛擬成員領域知識可參考程度。如表 5.1 所示乃列出系統管理者所蒐集之部份精華知識文章（以電腦/數碼領域為例）。



圖 5.2、「搜搜問問」之精華知識文章

表 5.1、系統管理者所蒐集之部份精華知識文章

文件編號	領域類別	標題	內容(發問及回答)
1	電腦/數碼	低頭玩手機會帶來哪些傷害?	輻射, 手機屏幕小...
2	電腦/數碼	優盤與 U 盤是一種東西嗎?	1998 年, 作為朗科公司...
3	電腦/數碼	什麼樣的密碼是最安全的?	用大寫字母 + 小寫字母...
4	電腦/數碼	為什麼鍵盤上的鍵不是上下對齊的?	補充一句鍵盤上的...
5	生活家居	寄快遞, 早上寄出去, 從下沙到杭州市區...	你好, 你可以試一下...
6	生活家居	為什麼河南今年到現在都不覺得...	河南今年一定會下雪...

■ 一般使用者蒐集虛擬社群討論串資料

一般使用者使用本系統前, 先行蒐集未判定虛擬社群討論串問答資料, 以透過本系統幫助其判定討論串中間答契合度, 以及成員可參考程度。於案例驗證中, 本研究驗證方式乃蒐集「搜搜問問」之討論串資料(如圖 5.3 所示), 作為驗證資料樣本, 以測試系統之可行性與績效。如表 5.2 與表 5.3 所示乃列出一般使用者所蒐集之部份討論串資料。



圖 5.3、「搜搜問問」之討論串資料

表 5.2、一般使用者所蒐集之部份討論串資料(應用於「問答契合度解析」)

討論串編號	領域類別	討論串標題	發問及回答內容
1	電腦/數碼	電腦老師自己斷網...	檢查 MODEM 散熱...
2	電腦/數碼	電腦顯示缺少文件...	建議你下載安裝驅動精靈...
3	電腦/數碼	想買個 40 寸電視當顯示器...	可以。希望是帶有 HDMI...
4	電腦/數碼	不小心卸載了音頻設備...	從硬件設備管理器...
5	社會/人文	請問基督徒, 受洗的...	要知道為什麼受洗...
6	社會/人文	新年怎樣拜年?	拜年是中國民間的傳統習俗...

表 5.3、一般使用者所蒐集之部份討論串資料（應用於「成員參考度解析」）

成員編號	成員名稱	領域類別	發文內容
1	胡夢龍	教育/科學	G 帶錶物體所受重力...
		生活家居	拿戶口本買全票
		教育/科學	自己寫吧，抄也沒用...
2	阿里⑧⑧	地區問題	阻止要阻的巧妙...
		社會/人文	對強迫性的性騷擾...
		遊戲	目前 QQ 音速已经...

### 1. (B1)系統管理者上傳領域訓練文件

當系統管理者進入此功能時，系統即呈現領域文件標題與領域文件內容欄位，供使用者輸入如標題「電腦老師自己斷網...」及內容「有沒有什麼檢測...」等基本資料，此外，系統亦提供領域文件所屬類別與上傳時間之下拉式選單，供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013 年 12 月 20 日」，使用者亦可點擊「瀏覽...」，以選擇並上傳自定義路徑如「D:\data」中「\*mht」領域文件之原始檔案，於此，當管理者完成相關資料之維護並點擊「確認」後（如圖 5.4 所示），系統將針對使用者上傳之原始檔案，進行斷詞與有效詞篩選之資料預處理作業，以分別取得如「有沒有(ADV) 什麼(DET) 檢測(Nv)...」，以及「方法；家；路由器...」等內文斷詞與有效詞篩選結果，於此，系統亦同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中（如圖 5.5 所示）。



圖 5.4、領域訓練文件基本資料之輸入



圖 5.5、領域訓練文件資料之預處理與新增

## 2. (B2)系統管理者執行領域性關鍵詞彙集合建立

「領域性關鍵詞彙集合建立」功能主要目的乃針對系統管理者所上傳之領域訓練文件，進行核心資訊之萃取，以建置相關領域性關鍵詞彙集合，藉此用於後續之判定。以下乃分別以步驟方式說明本功能之應用流程。

### Step(1)—對訓練文件斷詞並篩選有效詞

管理者於選定其上傳資料領域「電腦/數碼」並執行功能後，系統即對所有領域訓練文件之斷詞及有效詞篩選結果進行彙整，並於過濾重複詞彙後取得「有效詞彙彙整資訊」（如圖 5.6 至圖 5.8 所示），其次，系統將對領域有效詞集，以一個主要有效詞如「筆記本」，次要有效詞如「頻」等詞彙為一組，進行詞彙組之劃分，並提供管理者搜尋結果數輸入之介面（如圖 5.9 所示）。





圖 5.9、有效詞劃分與搜尋結果輸入畫面

### Setp(2)—計算領域有效詞間之關係距離

當管理者完成所需資訊之輸入並送出資料後，系統即針對各有效詞彙組，進行語意關係距離之計算，並藉小數點後五位進行四捨五入後，以呈現相關計算結果之數據如「方法」與「遊戲」之語意關係距離為「1.0」（如圖 5.10 所示）。



圖 5.10、語意關係距離計算畫面

### Setp(3)—建立領域性關鍵詞彙集合

待完成領域有效詞間關係距離之解析後，系統將針對詞彙組合間之整體關係距離進行計算，並提供予管理者「四分位數」、「中位數」、「平均值」及「自訂門檻值」等門檻篩選方式，待管理者選定所有篩選方式並送出資料後（如圖 5.11 所示），系統將計算得

相應之門檻值，並進行門檻符合狀況之篩選（如圖 5.12 所示），接著，管理者根據門檻篩選狀況，於選定分割獨立詞彙項目並送出後，系統將以「達成門檻之次數」為基，以劃分為獨立詞彙，並呈現符合門檻狀況之資訊（如圖 5.13 所示），於最終，管理者點擊「建置領域性關鍵詞彙」後，將完成電腦/數碼類領域性關鍵詞彙集合之建構（如圖 5.14 所示）。



圖 5.11、整體關係距離計算與門檻選擇畫面



圖 5.12、門檻值計算與詞彙組門檻符合狀況篩選畫面



圖 5.13、獨立詞彙劃分與門檻符合狀況篩選畫面



圖 5.14、領域性詞彙關係網路建立完成畫面

### 3. (B3)系統管理者執行系統參數值設定

待系統管理者完成領域性關鍵詞彙集合之建置後，管理者須點選系統參數設定模組，設定各項系統參數之權重值與門檻值，以保持後續系統主功能判斷之正確性（如圖 5.15 與圖 5.16 所示）。



圖 5.15、系統參數設定



圖 5.16、系統參數設定成功畫面

#### 4. (A1)一般使用者上傳未判定討論串資料

當一般使用者進入此功能時，系統即呈現提問標題與內容，以及回答筆數等欄位，供使用者輸入如提問標題「各路大仙...」、提問內容「各路大仙 看看我的電腦網速...」，以及回答筆數「5」等基本資料，此外，系統亦提供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013 年 12 月 22 日」，於此，當使用者完成相關資料之輸入並點擊「確認」後（如圖 5.17 所示），系統將呈現使用者所輸入提問之基本資料，並根據使用者方才輸入之回答筆數「5」，創建 5 個回答內容輸入之欄位，使用者根據討論串中回答之順序，依序輸入「網速和網絡帶寬...」、「看看有多少進程」、「看看有沒有偷網用的...」、「網速是和運營商...」與「下個魯大師」等回答內容，並於點擊「確認」（如圖 5.18 所示），系統將針對領域討論串「1034646889」執行斷詞與有效詞篩選等資料欲處理動作，以分別

取得如「各(DET) 路(M) 大仙(N)...」，以及「大仙；我；電腦...」等內文斷詞與有效詞篩選結果，並同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中，接著，當使用者點擊「下一步」後（如圖 5.19 所示），系統即針對各筆回答執行資料預處理動作，以分別取得如「網速(N) 和(C) 網絡(N)...」，以及「網速；網絡；帶寬...」等資料預處理結果，亦同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中（如圖 5.20 所示）。



圖 5.17、討論串基本資料之輸入



圖 5.18、討論串各筆回答資料之輸入



圖 5.19、討論串提問資料之預處理與新增



圖 5.20、討論串各筆回答資料之預處理與新增

## 5. (B4)系統管理者執行問答契合度解析

「問答契合度解析」功能主要目的乃解析使用者上傳虛擬社群討論串之問答資料，當中提問與各筆回答間之契合程度，以提供使用者評估該筆知識內容採納價值之相關依據。以下乃分別以步驟方式說明本功能之應用流程。

### Setp(1)—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

系統管理者執行此功能並以條件式查詢方式，查詢得一般使用者上傳討論串「1034646889」之基本資料，當管理者選取此討論串並送出資料後，將透過系統取得「斷詞」、「有效詞篩選」與「關鍵詞彙」等資訊，接著，管理者勾選其欲分析回答之編號

「1034646889」並送出後（如圖 5.21 與圖 5.22 所示），系統將對問答執行語句分割，以取得「前文語句」、「後文語句」，並計算得如關鍵提問詞彙如「電腦」之前後文配對及標題加權整體分數為「0.2437」（如圖 5.23 所示）。



圖 5.21、討論串查詢與分析選擇介面



圖 5.22、目標提問與各筆回答資料之預處理畫面



圖 5.23、前後文配對及標題加權整體分數計算畫面

### Setp(2)—計算候選關鍵提問與回答詞彙之詞類相關性平均分數

系統管理者將可藉此步驟之運行，根據管理者先前設定之詞類距離採納值，以建置並取得關鍵提問詞彙如「電腦」之前推算集合為「的；我；看看...」，並計算得詞類相關性平均分數為「0.5」（如圖 5.24 所示）。



圖 5.24、詞類相關性平均分數計算畫面

### Setp(3)—取得候選關鍵提問與回答詞彙之代表性得分

待取得關鍵問答詞彙之「前後文配對及標題加權整體分數」，以及「詞類相關性平均分數」後，系統管理者執行此步驟時，系統將根據參數設定中之比重值，以計算得關鍵提問詞彙如「電腦」之代表性得分為「0.34622」（如圖 5.25 所示）。



圖 5.25、關鍵詞彙代表性得分解析畫面

#### Setp(4)—計算目標虛擬社群討論串中提問與回答品質之得分

於此部份，系統管理者透過步驟之運行，以根據「代表性得分」、「關鍵問答詞彙出現頻率」與「關鍵問答詞彙總數」等數據，藉重要性分佈之概念，以解析得「0.73966」及「1.49634」等提問與回答品質得分（如圖 5.26 所示）。



圖 5.26、提問與回答品質得分計算畫面

#### Setp(5)—計算目標虛擬社群討論串中之問答內文契合度得分

系統將自動彙整提問與回答之斷詞詞彙，並於統計詞彙發生頻率後（如詞彙「各」於提問及回答語句分別出現「1」與「0」次），透過向量空間模型方法論之應用，以計算得問答內文契合度得分「0.52」（如圖 5.27 與圖 5.28 所示）。



圖 5.27、斷詞詞彙詞頻資訊彙整畫面



圖 5.28、問答內文契合度得分計算畫面

### Setp(6)—計算目標虛擬社群討論串中之問答語意關係距離

當系統管理者於有效詞彙組合劃分介面中完成「獨立詞彙搜尋結果數」及「詞彙組搜尋結果數」等數據之輸入並送出後，即透過系統以改良後標準化 Google 距離方法論，計算得所有詞彙組合之「語意關係距離」（如圖 5.29 與圖 5.30 所示），並於後續根據上述語意關係數據，以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 5.31 所示）。



圖 5.29、提問與回答之有效詞組合劃分畫面



圖 5.30、各詞彙組合語意關係距離之計算畫面



圖 5.31、提問與回答有效詞組合之劃分畫面

## Setp(7)—計算目標虛擬社群討論串中之整體問答契合度得分

於最終，系統管理者透過系統彙整「提問與回答品質得分」、「問答內文契合度得分」與「問答語意關係距離」等數據後，將計算得整體問答契合度得分「3.62117」，以及正規化之整體問答契合程度分數「0.42771」（數值落於0~1間）（如圖 5.32 所示），相關之分析結果亦同時儲存於資料庫中。



圖 5.32、整體問答契合度得分計算畫面

## 6. (A2)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 5.33 與圖 5.34 所示，使用者得知回答編號「1763146692」與討論串中提問之整體問答契合度為「0.42771」，使用者可參考此數據，以決定是否採納該筆回答之知識內容，而若此使用者為該社群之管理者，其亦可參考此契合狀況，以執行相關之管理措施。



圖 5.33、討論串資料之查詢



圖 5.34、檢視詳細討論串資訊

## 7. (A3)一般使用者上傳未判定成員歷史發文資料

當一般使用者進入此功能時，系統即呈現成員名稱與歷史發文筆數欄位，供使用者輸入如成員名稱「galant7072」，以及歷史發文筆數「8」等基本資料，此外，系統亦提供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013年12月22日」，於此，當使用者完成相關資料之輸入並點擊「確認」後（如圖 5.35 所示），系統將呈現使用者所輸入提問之基本資料，並根據使用者方才輸入之成員名稱「galant7072」，以及歷史發文筆數「8」，建置新成員資料，並創建 8 個歷史發文內容輸入之欄位，使用者根據目標成員之領域歷史發文，依序輸入標題「東京景點分享」，以及「這次是第一次...」、「為什麼都顯示網址...」、「感謝樓上大大...」...等歷史發文內容，並於點擊「確認」後（如圖 5.36

所示)，系統將針對目標成員「galant7072」之所有領域發文資料，執行斷詞與有效詞篩選等資料處理動作，以分別取得如「這(DET) 次(M) 是(Vt)...」，以及「航空；酷航；空姐...」等資料預處理結果，亦同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中（如圖 5.37 所示）。



圖 5.35、成員歷史發文基本資料之輸入



圖 5.36、目標成員各筆歷史發文資料之輸入



圖 5.37、目標成員各筆歷史發文資料之預處理與新增

## 8. (B5)系統管理者執行成員資料分群

「成員資料分群」功能主要目的乃根據虛擬社群成員之歷史發文，進行領域歷史發文資料之分群，以提供後續成員參考度解析之相關依據。以下乃分別以步驟方式說明本功能之應用流程。

### Setp(1)—斷詞與篩選有效詞

系統管理者針對欲分群成員「galant7072」之歷史發文（如圖 5.38 與圖 5.39 所示），按下「顯示成員歷史發文領域分佈狀況」後，系統將呈現目標成員歷史發文之領域分佈狀況，接著，管理者選取「電腦/數碼」領域（如圖 5.40 所示），以進行「斷詞」與「有效詞篩選」等資料預處理動作，並於排除重複詞彙後，以彙整目標成員歷史發文之有效詞集資訊（如圖 5.41 與圖 5.42 所示）。



圖 5.38、查詢成員歷史發文資料



圖 5.39、檢視目標成員詳細歷史發文



圖 5.40、統計成員歷史發文領域分佈



圖 5.41、成員領域歷史發文之資料預處理



圖 5.42、彙整領域有效詞資訊

## Setp(2)—篩選頻繁項目集

待完成資料之預處理後，系統將透過管理者所設定系統參數中之「最小支持度」，以篩選得如「{淺草寺}」之頻繁項目集（如圖 5.43 所示）。



圖 5.43、篩選頻繁項目集

### Setp(3)—建置相似矩陣並判定分群類型

待取得頻繁項目集後，系統將根據目標成員與其於領域中之所有發文，以頻繁項目集之交集結果為基礎，分析得各發文間之相似程度，如發文編號「1」與「2」之相似度為 0，並同時彙整得最大相似度為「1」，以及不為 0 之最小相似度為「1」，藉此判定得目前所屬之分群類型（CType）為「2」（如圖 5.44 所示）。



圖 5.44、建置相似矩陣並判定分群類型

### Setp(4)—劃分領域資料之群集

接著，系統根據前一步驟中判定得之分群類型，將發文編號「4」與「10」劃分為同一群集，並將相似度數據由原先之「Sim(4,10)=1」取代為「Sim(4,10)=0」，於此同時，

系統亦針對經取代相似度數據後之新相似矩陣，彙整得最大相似度為「1」，以及不為 0 之最小相似度為「null（無符合條件之數值）」，並藉此判定得分群類型（CType）為「3」（如圖 5.45 所示），此外，由於部分成員資料仍未完成分群，故管理者於此點擊「繼續執行分群」，系統將根據前步驟中之分群類型，再次執行分群之動作，以將「1」、「3」、「5」、「6」、「7」與「11」等發文編號劃分為同一群集，於此，由於全部之資料皆已完成群集之劃分，故使用者於點擊「彙整最終分群結果」後（如圖 5.46 所示），於最終，系統將彙整目標社群成員「galant7072」於「電腦/數碼領域」中，歷史發文之分群結果，並呈現如發文編號「1」之發文內容、隸屬於「群集 2」，以及目標社群成員所有發文之分群狀況等資訊，以提供使用者參考，此外，當使用者點擊「顯示群集分佈圖」時，系統將以彈跳新視窗方式，呈現使用者領域歷史發文群集分佈結果，如分群結果中包含「Cluster1」與「Cluster2」兩群集，而「Cluster1」中乃包含編號「4」及「10」等歷史發文（如圖 5.47 所示）。



圖 5.45、劃分領域資料之群集(1)



圖 5.46、劃分領域資料之群集(2)



圖 5.47、彙整最終分群結果

## 9. (B6)系統管理者執行成員參考度解析

成員參考度解析功能主要目的乃根據上一步驟中分群得之成員資料群集，進行成員參考度解析，以判定得成員領域知識可參考程度。以下分別說明本功能應用流程。

### Setp(1)—定義群集名稱

系統管理者針對欲分群成員「galant7072」之歷史發文(如圖 5.48 與圖 5.49 所示)，按下「顯示成員歷史發文領域分佈狀況」後，系統將呈現目標成員歷史發文之領域分佈狀況(如圖 5.50 所示)，接著，當管理者選取「電腦/數碼」領域後，系統將彙整目標成員已完成分群之資訊，並根據頻繁項目集與群集間之關係，定義並取得如群集編號「1」

之群集名稱為「淺草寺、餐廳」(如圖 5.51 與圖 5.52 所示)。



圖 5.48、查詢成員歷史發文資料

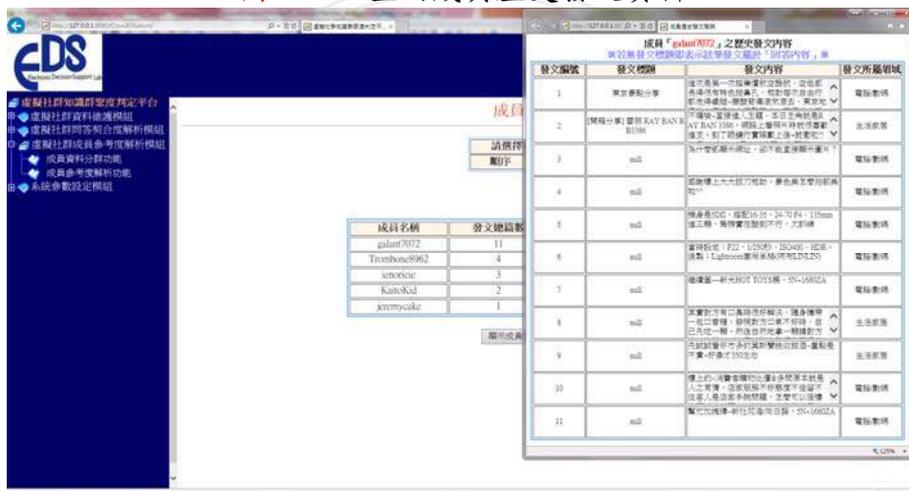


圖 5.49、檢視目標成員詳細歷史發文



圖 5.50、統計成員歷史發文領域分佈



圖 5.51、彙整成員領域歷史發文之分群資訊



圖 5.52、定義群集名稱

### Step(2)—取得群集名稱

待完成群集名稱之定義後，系統將根據使用者選定之領域類別，進行目標領域中所有領域訓練討論串之彙整，並同時執行斷詞與篩選有效詞等資料預處理動作（如圖 5.53 所示），接著，系統即以向量空間模型之方式，計算得如群集編號「1」與討論串編號「1」之內文相似度為「0.81626」，並計算得整體內文相似度為「0.35933」（如圖 5.54 所示），另一方面，系統將以群集為單位提供有效詞彙組合輸入之介面，於此，待管理者完成全數群集搜尋結果數據之維護後（如圖 5.55 與圖 5.56 所示），並執行下一步，系統將計算得「0.28336」與「0.3641」等語意關係距離，以及整體語意關係距離「0.32373」（如圖 5.57 所示），於最終，系統將整合上述所有資訊，以解析得目標社群成員「galant7072」於「電腦/數碼」領域類別中之領域知識可參考性分數為「1.10997」，以及領域知識可參考程度為「0.21268」（如圖 5.58 所示）。



圖 5.53、彙整並預處理領域討論串之資訊



圖 5.54、計算領域討論串與群集間之內文相似度



圖 5.55、劃分群集與討論串之有效詞組合(1)

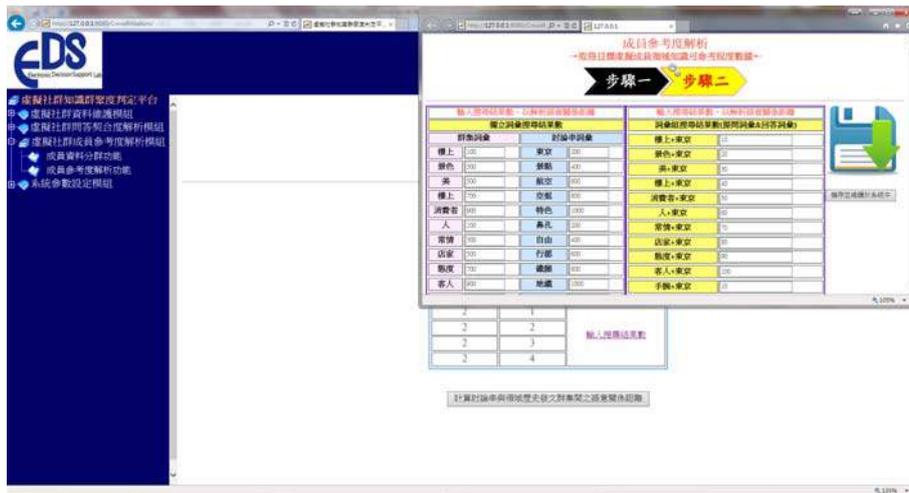


圖 5.56、劃分群集與討論串之有效詞組合(2)



圖 5.57、計算領域討論串與群集間之語意關係距離



圖 5.58、計算目標社群成員領域知識可參考程度

## 10. (A4)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，一般使用者可透過「成員歷史發文資料查詢」功能，以查詢得欲了解目標成員領域知識可參考度解析之結果，如圖 5.59 與圖 5.60 所示，使用者得知目標成員「galant7072」於「電腦/數碼」領域中之領域知識可參考程度為「0.21268」，使用者可參考此數據，以決定是否採納該成員之知識內容。



圖 5.59、成員歷史發文資料之查詢



圖 5.60、檢視詳細成員歷史發文資訊

## 5.2 系統案例驗證與評估

本研究乃以「虛擬社群知識群聚度判定」為基，開發一套「虛擬社群知識群聚度判定系統」，於「虛擬社群知識群聚度判定」之主題中乃包含「虛擬社群問答契合度解析」

與「虛擬社群成員參考度解析」兩項子議題，故本研究將分別針對上述兩議題，進行獨立之驗證，確保兩相議題間之驗證結果不相互影響，以獲致更具參考性之系統驗證與評估結果，另一方面，為了進行更具嚴謹系統績效之評估，本研究亦針對「虛擬社群知識群聚度判定」主題，以案例為導向探討整合「問答契合度解析」及「成員參考度解析」兩層面所發展之「虛擬社群知識群聚度判定系統」，於應用層面之實質效用，以及相關之管理意涵。整體而言，本研究於5.2節中之規劃如圖5.61所示。

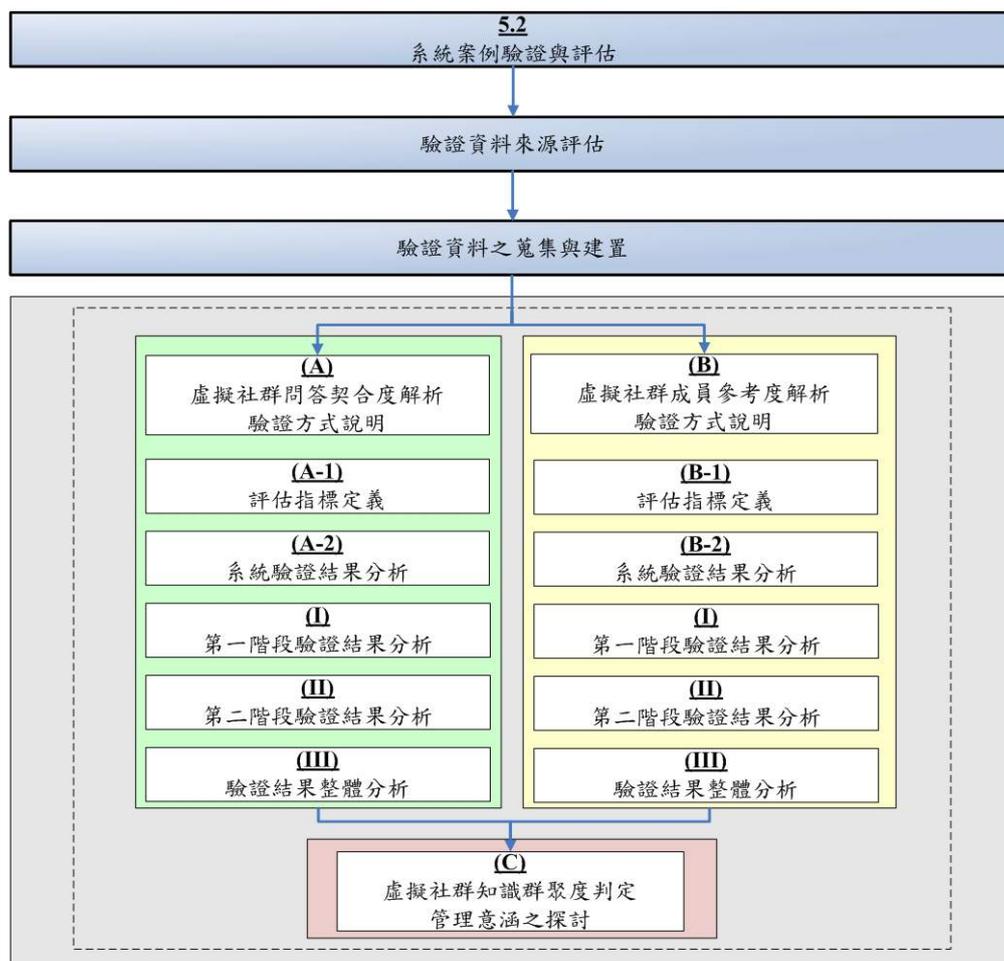


圖 5.61、系統驗證與評估之架構

### 驗證資料來源評估

本研究係參考 Toba 等人 (2014) 所設計高品質回答內容萃取演算法之績效驗證方式，於驗證本系統之績效與可行性，該研究認為以社群成員分享為基之問答虛擬社群，其本質應趨向於成員自發性分享「正確」知識，亦即正確之回答內容於整體虛擬社群中應佔較大之比例，因此，該研究由「Yahoo! Answers」中，蒐集 5854 筆涵蓋 12 個領域

類別之問答內容，且所有蒐集資料約有 80% 為具備良好品質之問答內容，當中，該研究評斷回答品質之方式係以兩面向，包含人工審核（由兩位自願者審核），以及參考最佳解答（由提問者指定）等面向，進行 Cohen's Kappa 統計（信度一致性測量），將兩位審核者視為兩個觀測者，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 之間，進行相應資料之蒐集。待完成資料之蒐集後，該研究即以「正確率」之方式，針對該研究開發之模式與人工審核之比值進行計算，以於最終獲致相關績效之評估結果。

承接於上述，Toba 等人 (2014) 驗證資料之來源為「Yahoo! Answer」（語言別為英文），此知識分享虛擬社群屬 Yahoo 公司（企業）所提供，與「Yahoo 奇摩知識+」（語言別為中文）相較之，雖整體介面不盡相同，但就功能與性質而言皆大致相同，而本研究係以分析中文問答文章為主，故本研究於此即以「Yahoo 奇摩知識+」為代表，進行相關驗證資料來源之評估。就 Yahoo 奇摩知識+ 中使用者所評論問答內容之品質，以及其所訂定知識等級衡量標準而言，至今少有研究針對上述兩者評定標準之合理性與正確性進行探討，因此，Toba 等人 (2014) 所提之驗證方式固然合理，但仍須考量該虛擬社群知識等級衡量標準之精確性與嚴謹性。有鑑於此，為使驗證資料更具可信度，本研究即針對「Yahoo 奇摩知識+」、「百度知道」、「搜搜問問」與「愛問知識人」等目前較具代表性之中文知識型虛擬社群進行質化之比較，以評估驗證資料來源之嚴謹性與可用度，相關之比較彙整如表 5.3。

表 5.3、中文知識型虛擬社群質化之比較

虛擬社群 比較項目	Yahoo 奇摩知識+	百度知道	搜搜問問	愛問知識人
知識量（問與答）	20,923,012 筆	242,459,614 筆	222,952,612 筆	無統計數據
知識等級制度	✓	✓	✓	✓
領域專家制度		✓	✓	
其他使用者可評價	✓	✓	✓	
知識管理員		✓	✓	✓

就比較結果而言，於知識量之層面，「百度知道」與「搜搜問問」乃當中具備龐大知識量之虛擬社群；於知識等級制度之層面，當中所有之虛擬社群皆具備知識此制度；於領域專家制度部份，僅「百度知道」與「搜搜問問」具備此制度；於其他使用者可評價面向，除「愛問知識人」以外，其他虛擬社群皆具備此功能；知識管理員之機制，除

「Yahoo 奇摩知識+」以外，其餘虛擬社群皆具備此機制。

針對於上述質化比較言之，可從中發現「百度知道」與「搜搜問問」係當中知識量較龐大，且具備最多項管理制度之虛擬社群，故作為驗證資料之來源，將較其他虛擬社群更為適切，而本研究於針對上述兩者進行進一步之探討後，從中發現，兩虛擬社群中「知識等級制度」、「其他使用者可評價」，以及「知識管理員機制」等運作方式皆相似，差異較大之層面為「領域專家制度」，雖兩者領域專家之審核，皆由人工之方式進行審核，但「百度知道」僅針對特定領域之專家進行審核，而「搜搜問問」則提供更廣泛之領域專家審核，此外，「百度知道」對領域專家審核之制度較不透明化（僅公開須符合「回答品質高」、「專業從業人員」與「具備證書」等條件），而「搜搜問問」除公開上述百度知道中之條件外，亦制訂一套知識採納率標準，以及成為領域專家後續之獎懲規範（若無達成規範，將取消專家資格）。綜上而論，雖「搜搜問問」知識量小於「百度知道」，但考量驗證資料應「質」大於「量」之因素，本研究最終決定將以「搜搜問問」作為驗證資料之來源。

### **驗證資料之蒐集與建置**

有鑑於上述之比較結果，本研究乃選定「搜搜問問」作為驗證資料之來源，進行本研究「虛擬社群知識群聚度」判定結果績效之衡量，以驗證本系統之可行性，而於驗證系統績效前本研究乃須訓練並建置「領域性關鍵詞彙集合」，以使後續之判定能順利運行。虛擬社群「搜搜問問」中之知識頻道，存在數量龐大一般社群成員所分享，以及官方管理者所定期發佈各領域之「精華知識」，由於此些領域知識以較符合虛擬社群成員口語化之方式進行知識之分享，且因知識內容與領域之分類乃藉多數成員之人工審核建置而成，故「精華知識」乃具備相當知識之品質與分類之正確性，有鑑於上述之優勢，本研究乃蒐集「搜搜問問」知識頻道內之「精華知識」（如圖5.62與圖5.63所示），並參考當中知識之分類，作為領域訓練資料，以建構符合虛擬社群成員口語化表達特性之「領域性關鍵詞彙集合」，進而於後續取得更具準確性判定與驗證之結果。



圖 5.62、「搜搜問問」中之知識頻道



圖 5.63、「搜搜問問」知識頻道內之精華知識

### 5.2.1—(A) 虛擬社群問答契合度解析之驗證

#### ➤ 系統驗證設計說明

於「虛擬社群問答契合度解析模組」中，本研究參考 Toba 等人 (2014) 所設計高品質回答內容萃取演算法之績效驗證方式，於驗證本研究開發模組與系統之績效與可行性。Toba 等人 (2014) 之主要目的係面向社群成員分享為基之問答虛擬社群，運用階層式分類法進行回答內容品質之分類，以從中萃取得高品質之回答內容。

本研究「問答契合度解析模組」之目標雖為解析提問與回答契合程度，但廣義言之，

Toba 等人 (2014) 於目標問答討論串中萃取得之高品質解答 (最佳解答)，亦可視為與目標提問具備高契合程度之回答，就本質而言，本研究與該研究之最終目標具備高度之相關性，因此，該研究所提實具合理性之驗證方式，可為本研究借鑑之。

然而，雖 Toba 等人 (2014) 最終目的與本研究具備相當之契合性，但該研究仍未考量「群眾智慧」於虛擬社群中所賦予之影響，即問答契合之評估雖為關鍵，但針對社群成員所凝聚群眾智慧之層面亦須納入考量，尤以本研究視知識群聚乃虛擬社群形塑具品質知識之關鍵而言，群眾智慧之因子亦須納入系統驗證之考量。為增添系統驗證對群眾智慧因子之考量，本研究將改良並運用 Toba 等人 (2014) 所採之系統驗證方式，針對本研究發展之「虛擬社群問答契合度解析模組」進行相關系統之驗證

### ➤ 系統驗證實作說明

首先，本研究乃於「搜搜問問」中蒐集680筆具備相當品質，且具備領域概念之問答內容 (如表5.4所示例)，而評斷之標準將以兩面向，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之間；其次，由上述680份問答內容討論串中隨機挑選出200份問答內容討論串作為訓練資料，視為第一階段驗證，將其逐一匯入系統中，以獲取「領域性關鍵詞彙集合」，亦由680份問答內容討論串中，隨機挑選20筆領域問答內容 (各類別2筆，如表5.5所示) 作為測試資料，進而判定問答契合度語料庫之分析基準，並透過對「正確率」與「召回率」指標之應用，藉此階段性地評估本系統對於「虛擬社群問答契合度解析」之績效；爾後，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份網頁知識文件分為6週期持續匯入系統中，每週期分別匯入80份網頁知識文件，最後，於各週期中利用前述所選定之20筆領域問答內容 (測試資料) 重新進行「虛擬社群問答契合度解析」之推論，以分析系統於不同訓練問答討論串數量下之長期學習趨勢。

表 5.4、虛擬社群問答契合度解析訓練資料表（部分資料）

編號	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
1	奧運體育	WWE 中的 8B 去哪裡了？	Batista (巴蒂斯塔) 已經在去年 5 月份就離開 WWE 了，目前正在從事電影行業，最近還在...	13/0	高品質知識	非高品質知識
2	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。一種說法是古希臘人創造的。傳說古希臘人注意到...	1491/4	高品質知識	高品質知識
3	電腦/數碼	外出時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。平時上下班都還湊合，但是一遇到稍微...	82/51	非高品質知識	高品質知識
4	藝術/文學	公主和格格的區別是什麼？	烏一、親王之女，稱為"和碩格格"，漢名為"郡主"；二、世子及郡王之女...	1662/0	高品質知識	高品質知識

表 5.5、虛擬社群問答契合度績效判定測試資料彙整表（已轉換為繁體中文）

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？	1。電腦的配置一般是指電腦的硬件配件的高檔程度、性價比等，電腦的性能好壞主要決定於 (1)CPU：決定運行速度，比如賽揚 D2.66G，其中“2.66G”是指它...
	諾基亞 5230 好不好用？	優點：價格實惠,外觀時尚,支持 GPS 導航功能,支持 WCDMA 的 3G 網絡.缺點：沒有 WAPI 無線功能,攝像頭像素較低,機身較厚.優點：機子不錯，小半年了，沒有...
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，而且貴重木材很少使用實木。例如實木餐椅比較常見，但一般是中高檔的用進口櫟木，中檔的用楓木、樺木、國產櫟木，低檔的...
	洗衣機多久清潔消毒一次比較好？	洗衣機洗衣機內乍看非常乾淨，但洗衣筒外還有個套筒，洗衣水就在這兩夾層中間來回流動，時間一長，夾層里便會附著大量的污垢。這些污垢不僅在洗衣時嚴重...
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水中，水不僅阻力大，而且導熱性能也非常好，散熱速度快，因而消耗熱量多。就好比一個剛煮熟的雞蛋，在空氣中的冷卻速度...
	適合都市人的有氧運動有哪些？	跳繩：不間斷地跳繩 10 分鐘，和慢跑 30 分鐘消耗的熱量差不多，是一種低耗時高耗能的有氧運動。長期堅持可以令雙腿變得緊緻。下蹲：能明顯改善梨形身材...
休閒/愛好	麻將是怎麼發明的？	麻將也稱“麻雀”，這個由我國古代發明，在中國擁有億萬愛好者的博戲，目前在國外許多國家都有了忠實粉絲。不過，麻將究竟始於何時？它是由誰發明的？這個問題...
	從音樂喜好中可以看出人的哪些品質？	“從品味這扇窗，能看到別人的生活。”德克薩斯大學的心理學家山姆格斯靈說，“品味能揭示一個人的價值觀和需求，這對生活來說太重要了。”而對書籍、音樂...
藝術/文學	米開朗琪羅留在佛羅倫	《勝利者》在佛羅倫薩國家美術館中，有一座米開朗琪羅稱

	莎畫室的作品是什麼	之為《勝利者》的雕像。這是一個擁有健美身軀的裸體的青年，低低的額上垂下曲發。他跪在一個滿臉...
	金庸筆下的韋小寶史上真的有嗎？	說到韋小寶，人們自然會聯想到金庸武俠小說《鹿鼎記》中的那個溜鬚拍馬、油嘴滑舌的韋小寶了。金庸在塑造韋小寶這個人物，簡直直插到人入心入肺，連看客都有點...
社會/人文	外星人接觸過瑪雅古文明嗎？	據國外媒體報導，在 1947 年的夏天，據傳位於新墨西哥州羅斯威爾附近墜毀了一個不明飛行物，並且被軍方控制。由此事件展開了對不明飛行物真相的調查。根據電影...
	喝酒為什麼要碰杯	喝酒為什麼要碰杯？目前有兩種說法。一種說法是古希臘人創造的。傳說古希臘人注意到這樣一個事實，在舉杯飲酒之時，人的五官都可以分享到酒的樂趣：鼻子能...
教育/科學	開學為何讓人如此恐懼？	原因一：黑白顛倒，身體跟不上 開學第一周，學生很容易患上“開學綜合徵”。早晨起不來床、上課無法集中精力、時常感到頭暈目眩、渾身沒勁，這些都可能...
	銀針真可以用來驗毒嗎？	在民間，銀器能驗毒的說法廣為流傳。早在宋代著名法醫學家宋慈的《洗冤集錄》中就有用銀針驗屍的記載。時至今日，還有些人常用銀筷子來檢驗食物中是否...
健康/醫療	喝什麼溫度水解渴？	人在出汗很多、感到很渴的情況下，不顧一切地喝涼水或冷飲，結果只會越喝越渴，造成反射性出汗，使體內失水更多，對身體健康危害甚大。大量出汗還會導致體內...
	怎麼預防人體感染 H7N9？	日常預防 1、加強體育鍛煉，注意補充營養，保證充足的睡眠和休息，以增強抵抗力。2、盡可能減少與禽類和寵物狗，雪貂，的不必要的接觸，尤其是...
商業/理財	有哪些受用一輩子的理財好習慣？	青春的你，有很多主動權：你有大把時間肆意揮霍，也可以時不時氾濫一下情感。可是，在對待腦袋和口袋的態度上，你必須謹小慎微。俗話說，經濟基礎...
	吸引金錢法則是什麼？	當大多數美國人長到二十多歲的時候，他們就會深深地陷入債務危機。無論是他們的學生貸款，信用卡債務，還是汽車費用，都離不開債務。研究表明，大多數人...
娛樂/明星	有哪些明星客串《愛情公寓 4》？	《愛情公寓》第 4 季將開播，除了四對主角情侶的情感問題，更有眾多大牌鼎力“打醬油”。目前已經知悉的是，湖南衛視快樂大本營成員何炅、謝娜、杜海濤也將...
	《夏洛克 3》的看點是什麼？	“福華粉”們終於等到了英國 BBC 電視劇《神探夏洛克》第三季的第一集!看完後腦殘粉們有的繼續腦殘，有的卻大呼“上當!”皆因這一集以“卷福”和“花生”...

### (A-1)評估指標定義

於「虛擬社群問答契合度解析模組」績效之驗證中，本研究乃須蒐集高品質之問答討論串，將之視為高度契合之問答內容，以評估本模組於判定問答契合度層面之正確性，因此，為衡量所蒐集問答討論串之品質，本研究乃針對人工審核（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）兩面向，進行Cohen's Kappa統計，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之間（參考Toba等人（2014））。

另一方面，為了評估本模組之績效，即問答契合度解析之準確性，本研究亦利用「問答契合度召回率」與「問答契合度正確率」，此兩項量化指標進行系統績效驗證，藉以檢視本模組所判定問答契合度之結果，以及實際問答契合狀況之相符程度。

#### ➤ 領域訓練資料之Kappa統計值

Cohen's Kappa係衡量信度之指標，當某一相同目標被不同審核者進行審核時，將可能因審核者認知之差異，有不一致之審核結果，進而影響整體審核結果之可信度。以本研究高品質知識之審核為例進行說明，當審核者須一次性人工大量審視知識之品質時，將可能因認知標準之不同或身心狀況之影響，導致審核得之部分結果出現不一致或錯誤之狀況，亦即審核者A與審核者B於審視同一討論串時，將可能有不一致之審核結果，而如此不一致之結果，將對資料蒐集之可信度呈現負影響，進而導致後續系統績效驗證之準確性，因此，為避免觀測一致性之差異，對整體驗證資料信度之影響，本研究即採用Kappa統計值先行針對驗證資料之信度進行衡量，以維持整體驗證資料之可信度。

為更契合於本研究所著重「群眾智慧」之觀點(已詳述於「系統驗證設計說明」中)，本研究即針對Toba等人(2014) Kappa值應用之方法進行相應之改良。於此，本研究將兩位人工審核者，以及贊同與不贊同數比值(群眾判定之傾向)，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，當中，兩人工審核者同時標記目標討論串為高品質知識，或贊同與不贊同數比值大於80%，方可視該觀測者評斷(標記)目標討論串為「高品質知識」。此部份即「觀測一致性」與「期望一致性」之比例，詳細變數定義如下，而本研究所蒐集兩階段，共680筆驗證資料之Kappa統計值分析之狀況，則彙整如表5.6與表5.7：

Kappa	領域訓練資料品質觀測一致性衡量指標
$O_{er_i}$	第 $i$ 位審核問答討論串之觀測者， $i=1,2$ ， $O_{er_1}$ 為人工審核者、 $O_{er_2}$ 為參與評論之社群成員
$N$	問答討論串之總數
$TF\_O_{er_i}$	人工審核者標記高品質知識，且參與評論社群成員標記非高品質知識之數量
$FT\_O_{er_i}$	人工審核者標記非高品質知識，且參與評論社群成員標記高品質知識之數量
$M_T$	所有審核者一致標記高品質知識之數量

$M_F$  所有審核者一致標記非高品質知識之數量

$$Kappa = \frac{\left( \frac{\text{Sum}(M_T, TF\_Oer_i) \times \text{Sum}(M_F, TF\_Oer_i)}{N} \right) + \left( \frac{\text{Sum}(M_F, FT\_Oer_i) \times \text{Sum}(M_T, FT\_Oer_i)}{N} \right)}{1 - \frac{\left( \frac{\text{Sum}(M_T, TF\_Oer_i) \times \text{Sum}(M_F, TF\_Oer_i)}{N} \right) + \left( \frac{\text{Sum}(M_F, FT\_Oer_i) \times \text{Sum}(M_T, FT\_Oer_i)}{N} \right)}{N}}$$

表 5.6、第一階段驗證資料—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	196	1	197
非高品質	3	0	3
總數	199	1	200
<b>Kappa 值 = 0.9796</b>			

表 5.7、第二階段驗證資料—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	467	7	474
非高品質	6	0	6
總數	473	7	480
<b>Kappa 值 = 0.9722</b>			

➤ 問答契合度解析之召回率

問答契合度解析之召回率乃為一相對比例值，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，其相關變數定義如下：

- W 系統推論問答契合度時，問答討論串之問答契合度推論召回率
- p 問答討論串實際為高契合問答與推論為高契合問答相符之個數
- u 問答討論串實際為高契合問答之個數

$$W = \frac{p}{u}$$

➤ 問答契合度解析之正確率

問答契合度解析之正確率乃為一相對比例值，為「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。相關變數定義如下：

T	系統推論問答契合度時，問答討論串之問答契合度推論正確率
p	問答討論串實際為高契合問答與推論為高契合問答相符之個數
s	系統推論為契合問答之個數

$$T = \frac{p}{s}$$

## (A-2)系統驗證結果分析

本研究將虛擬社群問答契合度解析系統驗證結果分為「第一階段驗證結果分析」與「第二階段驗證結果分析」兩大項目。於執行階段性之驗證前，本研究乃先行由搜搜問問虛擬社群中，隨機挑選 20 筆領域問答內容以作為測試資料（如表 5.5 所示）。於「第一階段驗證結果分析」項目中，乃透過 200 筆領域訓練資料之匯入，以測試系統執行「虛擬社群問答契合度解析」之績效，藉此觀察 200 筆領域訓練文件之匯入，對系統正確性之影響。另一方面，於「第二階段驗證結果分析」項目中，本研究乃劃分六個週期，每期皆分別匯入 80 筆領域訓練文件，共計 480 筆領域訓練文件，以測試不同訓練資料量下系統進行「虛擬社群問答契合度解析」之績效。以下即針對各階段說明系統驗證過程，並分析系統驗證之結果。

### (I)第一階段驗證結果分析

於第一階段系統驗證中，本研究乃由 680 份與「電腦/數碼」、「生活家居」、「奧運體育」、「休閒/愛好」、「藝術/文學」、「社會/人文」、「教育/科學」、「健康/醫療」、「商業/理財」與「娛樂/明星」類別（依搜搜問問社群管理者分類）相關之問答討論串資料中，共挑選 200 份問答討論串資料作為訓練資料，並逐一匯入系統中，以作為第一階段驗證之基礎訓練資料。以下即針對各項指標（召回率與正確率）說明系統驗證過程，並分析系統驗證之結果。

#### ➤ 第一階段問答契合度解析結果（20 筆領域問答內容）

在 200 份網頁知識文件為訓練資料基礎之下，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%，以及問答契合度解析平均正確率為 30%，當中，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，此階段詳細之推論結果如表 5.8 所示。此外，虛擬社群問答契合度解析平均召回率，以及正確率之分佈趨勢如圖 5.64 及圖 5.65 所示。

表 5.8、虛擬社群問答契合度解析—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%

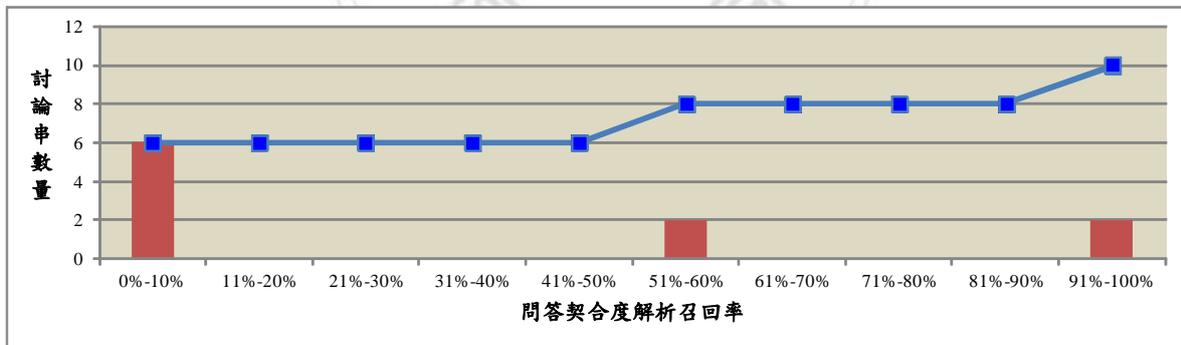


圖 5.64、第一階段問答契合度解析召回率之分佈趨勢

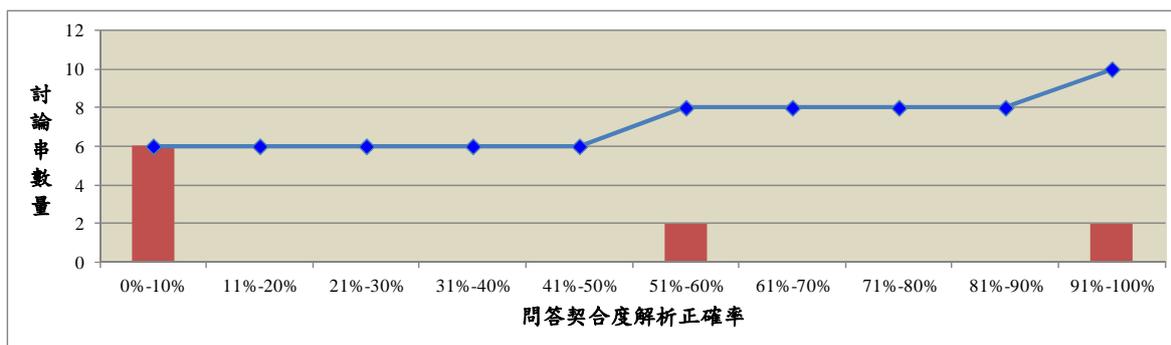


圖 5.65、第一階段問答契合度解析正確率之分佈趨勢

透過表 5.8、圖 5.64 及圖 5.65 統計結果得知，系統於第一階段之問答契合度解析之驗證結果中，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%。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，問答契合度解析之準確率與績效欠佳，尚無法準確判斷提問與回答之契合程度。

## (II) 第二階段驗證結果分析

第二階段系統驗證之作法乃以第一階段驗證中，所匯入系統資料庫之 200 份網頁知識文件為基礎，將剩餘之 480 份訓練用問答討論串資料再分為六個週期（每週期匯入 80 份訓練資料），依序匯入系統後端之資料庫中，並以第一階段驗證時所隨機選取之 20 筆領域問答內容重新進行系統績效測試，以瞭解系統於不同訓練資料數量基礎下進行「問答契合度解析」之績效變化趨勢，進而分析本系統學習成長之能力。以下乃針對「問答契合度解析」之各項指標說明系統第二階段各週期（第二週期至第六週期）驗證過程，並分析系統各週期之驗證結果。

### ➤ 第二階段問答契合度解析結果（20 筆領域問答內容）

本研究將第二階段之驗證分為六個週期，並於各週期中分別匯入 80 份領域訓練文件，以觀察隨訓練文件之增加，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 5.9，而各驗證週期問答契合度解析之績效分佈趨勢如圖 5.66 所示。

表 5.9、虛擬社群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%

藉由對表5.9之觀察可得知，以每週期增加80份領域訓練文件為單位，平均每週期間答契合度解析之召回率及正確率驗證指標，其整體成長率分別為8%與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，召回率及正確率將分別由第一週期之30%提升至80%與85%，綜合上述驗證結果之觀察，本研究所開發之虛擬社群問答契合度解析系統，具備學習能力與相當程度之正確性，亦即透過對虛擬社群問答契合度解析系統之使用，將確實為使用者提供具精確性問答契合程度參考之依據。

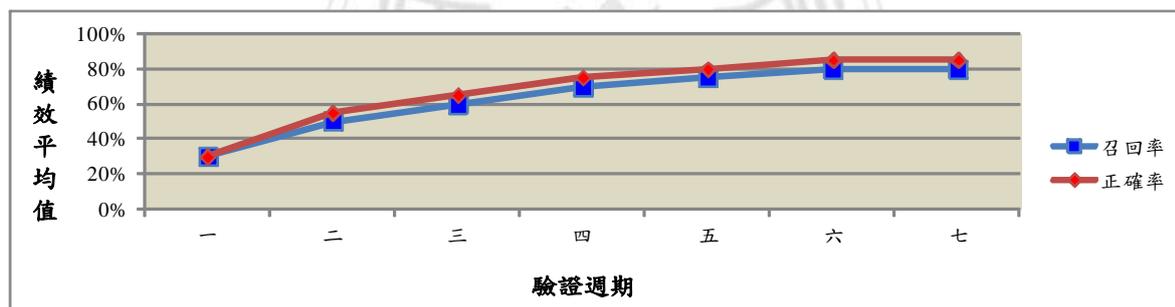


圖 5.66、各驗證週期間答契合度解析之績效分佈趨勢

### (III) 驗證結果整體分析

首先，本系統於問答契合度解析之「第一階段驗證」中，當訓練文件之資料量為 200 份時，其正確率與召回率整體平均值為 30%，由此可知，此階段判定之績效仍欠佳。

然而，於「第二階段驗證」中，系統隨著不同週期及訓練量之增加，其各項數據出現顯著成長之趨勢，尤以第一週期至第二週期之成長狀況最為明顯，於第六週期時，問答契合度解析之正確率與召回率可提升至 80% 及 85%，此外，由驗證結果中亦可得知，平均每週期間答契合度解析之召回率及正確率，其整體成長率分別為 8% 與 9%。承接於上述，本研究於不同訓練資料量下執行「問答契合度解析」核心功能績效之驗證，綜合於各週期驗證結果，可推論本系統之績效將因訓練資料之增加，持續且穩定地成長與提

升，而本研究所建置之「虛擬社群問答契合度解析模組」亦實具其可行性。

最後，綜合兩階段之驗證成效，各項驗證指標之相關結果整理如表 5.10 所示。由表 5.10 整理之結果可得知，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，且各項驗證指標皆於第六週期內呈現收斂狀態，因此，以本研究所隨機選定之驗證資料為例，當系統使用約 680 筆訓練知識文章時，可使系統各項推論指標績效提升至 80% 至 85% 之水準。

表 5.10、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%

整體而言，由第一階段與第二階段之驗證分析中，可得知當系統進行「問答契合度解析」時，訓練資料與週期乃影響系統績效之主要原因，而系統之績效將隨著不同週期及訓練量之增加，呈現持續成長之狀況，並於最終達到穩定且良好之績效水平。故本研究所建置之「問答契合度解析模組」確實能準確判定提問與回答之契合程度，並可依據判定結果，確實為使用者提供具精確性問答契合程度參考之依據。

## 5.2.2—(B) 虛擬社群成員參考度解析之驗證

### ➤ 系統驗證設計說明

本研究「成員參考度解析模組」之目標雖為判定社群成員領域知識可參考程度，但就技術層面廣義言之，Toba 等人 (2014) 透過階層式分類法於目標問答討論串中萃取得之高品質解答（最佳解答），與本研究藉由分群並計算整體群集關聯程度之作法，具備相當之關聯性，而 Liu 等人 (2013) 整合中研院 CKIP 斷詞系統，以及 TF-IDF 詞彙權重衡量法所發展之領域專家尋找模式，與本研究藉 CKIP 系統資料預處理，並透過向量空間模型之作法存在高度相關，此外，就最終結果而言，該研究所取得具專業性排名之領域專家搜尋結果（列表），亦與本研究取得社群成員領域可參考程度具備同質之結果，因此，上述研究所提之驗證方式，可為本研究借鑑之。

Toba 等人 (2014) 係採用「Cohen's Kappa」與「正確率」指標進行績效之驗證，而 Liu 等人 (2013) 則採用「Precision at 5 ; P@5」、「Mean Reciprocal Rank ; MRR」，以及

「Mean Average of Precision ; MAP」等指標進行績效之衡量。針對於上述指標之應用進行討論，雖「Cohen's Kappa」將可應用於對驗證資料蒐集與評估一致性之檢測，但 Toba 等人 (2014) 未考量「群眾智慧」於虛擬社群中所賦予之影響，即系統績效之評估固然關鍵，但針對社群成員所凝聚群眾智慧之層面亦須納入考量，尤以本研究視知識群聚乃虛擬社群形塑具品質知識之關鍵而言，群眾智慧之因子亦須納入系統驗證之考量，是故，為融入群眾智慧之觀點，本研究蒐集與評估驗證資料之標準將以兩面向，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。

另一方面，針對「正確率」與「P@5」系統準確性衡量之指標而言，「P@5」與「正確性」之概念大致相同，差異處為「P@5」可衍伸並應用於對搜尋結果排序正確性之評估，於此，就本模組之輸出結果（即虛擬社群成員領域知識可參考程度）言之，因未涉及「排序」之相關結果，故透過「正確率」指標進行績效之衡量將更為適切，此外，為能使系統績效之驗證更具參考性，本研究亦應用「召回率」指標進行系統績效之驗證。

再者，就「MRR」與「MAP」指標言之，於資料探勘領域中，此兩項指標主要應用於對檢索（搜尋）結果績效之評估，雖 Liu 等人 (2013) 與本研究具備同質之最終分析結果，即最終輸出結果分別為「領域專家列表」及「社群成員領域知識可參考程度」，但兩者最大之差異，即本研究係著重於對「程度」之評估，而該研究則著重於對「檢索結果」之排序，因此，「MRR」與「MAP」兩衡量排序績效之指標，本研究即不予採納。

此外，於社群成員參考性之「實際正確結果」（即實際成員參考性排名）評斷中，為契合於本研究所著重「群眾智慧凝聚」之觀點，本研究評斷實際成員參考性排名之依據，即以群眾意象之整體結果為基礎，進行相關實際正確結果之評估，於此，本研究乃根據搜搜問問於各社群成員資料頁面中，提供之「綜合聲望」作為目標成員實際領域參考度排名之依據，以針對本研究於各領域中隨機挑選之受測者，進行相關成員參考性之排名，藉以由群眾意向之層面，評估本系統判定結果與實際成員參考性排名之符合狀況，當中，「綜合聲望」乃由搜搜問問所提供，其計算方式係「依據目標成員在社區內之回答被採納率、滿意答案數、精華知識數、回答原創度，以及其他用戶對目標成員提供滿意答案的解決度評價，綜合計算出目標成員於社區內的綜合聲望值，代表目標成員於社區內回答問題的專業程度和貢獻水準（引用於搜搜問問）」。

綜合言之，為求系統績效驗證之嚴謹性及合理性，於績效衡量指標採用之層面，本

研究將運用「正確率」與「召回率」進行系統績效量化之衡量；於驗證資料蒐集之層面，本研究將參考 Liu 等人 (2013) 及搜搜問問群眾意象之整體結果，以蒐集相關領域受測者之資料（測試資料），並評估目標受測者實際領域參考性之排名，而於訓練資料之蒐集與評估中，本研究將應用「Cohen's Kappa」進行資料蒐集一致性之檢測。

#### ► 系統驗證實作說明

首先，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）（如表 5.11 所示例）（內容已轉換為繁體中文）。次之，本研究將針對 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。再者，本研究乃於「搜搜問問」中蒐集 680 筆具備相當品質，且具備領域概念之問答內容（如表 5.12 所示例）（內容已轉換為繁體中文），而評斷之標準將以兩面向，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 之間（參考 Toba 等人 (2014)）。爾後，由上述 680 份討論串中隨機挑選出 200 份討論串作為訓練資料，視為第一階段驗證，將其逐一匯入系統中，以獲取「領域性關鍵詞彙集合」，進而作為成員參考度語料庫之分析基準，並透過對「正確率」與「召回率」指標之應用，藉此階段性地評估本系統對於「虛擬社群成員參考度解析」之績效。爾後，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 份網頁知識文件分為 6 週期持續匯入系統中，每週期分別匯入 80 份討論串資料。最後，於各週期中利用前述所選定之 50 位受測者，重新進行「虛擬社群成員參考度解析」之推論，以分析系統於不同訓練資料量下之長期學習趨勢。

表 5.11、虛擬社群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網絡共享的設備，一般的連接和...
				2	有兩種辦法可以實現： 1.改變第二級路由器的連接方式器...
				3	寬帶提示錯誤 769：網卡禁用，一般是計算機網卡沒有成功...
				4	呵呵，WIN7 的 EXCEL 是什麼意思啊，你直接可以安...
				5	斷開你的路由器與上級路由器連接的網線，再用電腦登你...
				6	有可能是你的計算機網卡速度配置與上聯設備的以太網不...
				7	在可以正常上網的計算機上如下操作：右鍵網上鄰居-屬性-...
				8	網絡受限製或無連接有以下可能：1.這說明你的計算機網...
				9	直接恢復路由器出廠設置就可以解決的了，路由器背後有...
				10	1.引起寬帶掉線的原因比較多的，運營商設備的問題、你...

表 5.12、虛擬社群成員參考度解析訓練資料表 (部分資料)

編號	領域類別	提問內容	最佳解答	贊同/不贊同數	審核結果 A	審核結果 B
1	電腦/數碼	安全套真的可以讓普通手機也能水里拍照麼？	看過很多水里拍攝的照片，一打聽才知要特定的相機和相機防水套！一套下來要好...	47/0	高品質知識	非高品質知識
2	教育/科學	可能復活的滅絕動物有哪些？	任何生物的“製作”都離不開基因，基因的保存需要非常小心，如果接觸到陽光或...	1279/0	高品質知識	高品質知識
3	休閒/愛好	人吃狗糧可以麼？	如果你只吃一點點狗糧，可能什麼都不會發生。但狗糧不符合人類食品衛生和安全法...	21/8	非高品質知識	高品質知識
4	奧運體育	為什麼游泳運動員很少有黑人？	很少能看到某個黑人選手一統天下的。這裡面的奧秘到底在什麼地方？ 科學解釋...	2511/1	高品質知識	高品質知識

### (B-1) 評估指標定義

於「虛擬社群成員參考度解析模組」績效之驗證中，測試資料之層面，本研究乃須隨機挑選受測者，並根據其於搜搜問問中所累積之「綜合聲望」，進行成員參考度之排名，將之視為實際狀況中成員領域參考度之排名，以評估本模組於判定成員領域參考

度層面之正確性（相關說明及作法，已分別詳述於「系統驗證設計說明」與「系統驗證實作說明」中）。

領域訓練資料之層面，為蒐集高品質之問答討論串，以作為建立具專業性領域性關鍵詞彙集合之基礎（有效領域訓練資料），於實際蒐集並匯入系統前，針對於虛擬社群之問答討論串，本研究乃先行以人工審核（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）兩面向，進行Cohen's Kappa統計，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之間（參考Toba等人(2014)），方便蒐集得之領域訓練文件具備相當之專業性與可用性（相關說明、作法與結果，已詳述於「(A-1)評估指標定義」中）。

系統績效評估指標層面，為能評估本模組之績效，即成員參考度解析之準確性，本研究亦利用「成員領域可參考性排名召回率」與「成員領域可參考性排名正確率」，此兩項量化指標進行系統績效驗證，藉以檢視本模組所判定之結果，與實際狀況之相符程度。

#### ➤ 成員領域可參考性排名之召回率

成員領域可參考性排名召回率乃為一相對比例值，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，其相關變數定義如下：

W	系統推論成員領域可參考性排名時，成員之領域可參考性排名推論召回率
p	成員領域可參考性排名實際名次與推論名次相符之個數
u	成員領域可參考性實際總名次之個數

$$W = \frac{p}{u}$$

#### ➤ 成員領域可參考性排名之正確率

成員領域可參考性排名正確率乃為一相對比例值，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「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。相關變數定義如下：

T	系統推論成員領域可參考性排名時，成員之領域可參考性排名推論正確率
p	成員領域可參考性排名實際名次與推論名次相符之個數
s	系統推論成員領域可參考性總名次之個數

$$T = \frac{p}{s}$$

## (B-2)系統驗證結果分析

本研究將虛擬社群成員參考度解析系統驗證結果分為「第一階段驗證結果分析」、「第二階段驗證結果分析」、「驗證結果整體量化分析」與「驗證結果整體質化分析」四大項目。首先，於執行階段性之驗證前，本研究乃先行由搜搜問問虛擬社群中，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（如表 5.11 示例）；次之，於「第一階段驗證結果分析」項目中，乃透過 200 筆領域訓練資料之匯入，以測試系統執行「虛擬社群成員參考度解析」之績效，藉此觀察 200 筆領域訓練文件之匯入，對系統正確性之影響；接著，於「第二階段驗證結果分析」項目中，本研究乃劃分六個週期，每期皆分別匯入 80 筆領域訓練文件，共計 480 筆領域訓練文件，以測試不同訓練資料量下系統進行「虛擬社群成員參考度解析」之績效；爾後，於「驗證結果整體量化分析」項目中，本研究乃整合第一階段與第二階段驗證結果分析所得之數據，進行整體性系統量化績效之評估，以探討得本系統於「成員參考性排名」面向中系統所獲致準確性之績效；最後，於「驗證結果質化分析」中，本研究乃針對搜搜問問設計之「綜合聲望」指標，以及本研究最終之成員參考性判定結果，面向「成員參考性百分比」進行兩相質化層面之比較與討論，以完成包含「量化」與「質化」雙面向系統執行績效之驗證與評估。以下即針對各項目說明系統驗證過程，並分析系統驗證之結果。

### (I)第一階段驗證結果分析

於第一階段系統驗證中，本研究乃由 680 份與「電腦/數碼」、「生活家居」、「奧運體育」、「休閒/愛好」、「藝術/文學」、「社會/人文」、「教育/科學」、「健康/醫療」、「商業/理財」與「娛樂/明星」類別（依搜搜問問社群管理者分類）相關之問答討論串資料中，共挑選 200 份問答討論串資料作為訓練資料，並逐一匯入系統中，以作為第一階段驗證之

基礎訓練資料。以下即針對各項指標（召回率與正確率）說明系統驗證過程，並分析系統驗證之結果。

➤ **第一階段成員參考度解析結果（50 位受測者）**

在 200 份網頁知識文件為訓練資料基礎之下，系統針對 20 筆測試資料進行判定，獲得成員參考度解析平均召回率為 26%，以及成員參考度解析平均正確率為 26%，當中，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 5.13 所示。此外，虛擬社群成員參考度解析平均召回率，以及正確率之分佈趨勢如圖 5.67 及圖 5.68 所示。

表 5.13、虛擬社群成員參考度解析—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%

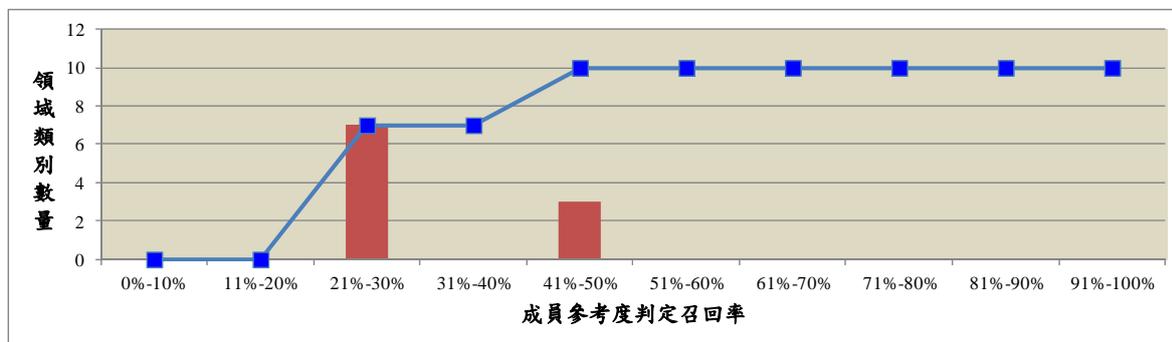


圖 5.67、第一階段成員參考度解析召回率之分佈趨勢

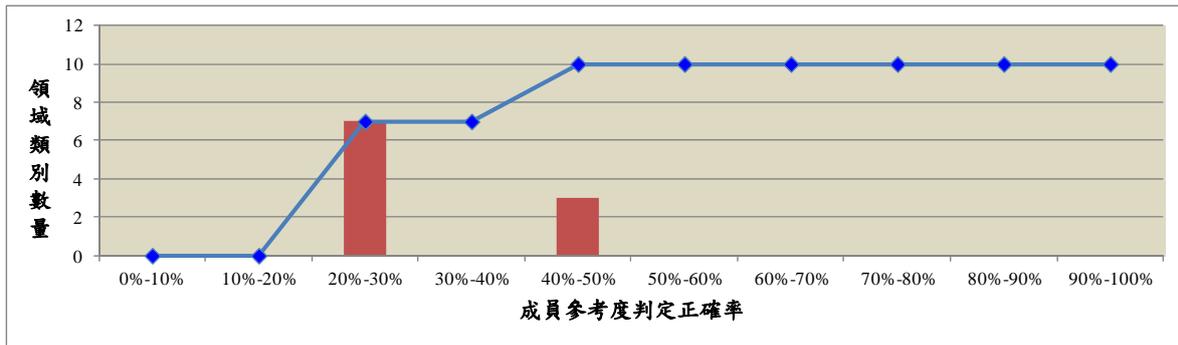


圖 5.68、第一階段成員參考度解析正確率之分佈趨勢

透過表 5.13、圖 5.67 及圖 5.68 統計結果得知，系統於第一階段之成員參考度解析之驗證結果中，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，成員參考度解析之準確率與績效欠佳，尚無法準確判斷成員領域可參考程度。

## (II) 第二階段驗證結果分析

第二階段系統驗證之作法乃以第一階段驗證中，所匯入系統資料庫之 200 份領域訓練文件為基礎，將剩餘之 480 份訓練用問答討論串資料再分為六個週期（每週期匯入 80 份訓練資料），依序匯入系統後端之資料庫中，並以第一階段驗證時所隨機選取之 20 筆領域問答內容重新進行系統績效測試，以瞭解系統於不同訓練資料數量基礎下進行「成員參考度解析」之績效變化趨勢，進而分析本系統學習成長之能力。以下乃針對「成員參考度解析」之各項指標說明系統第二階段各週期（第二週期至第六週期）驗證過程，並分析系統各週期之驗證結果。

### ➤ 第二階段成員參考度解析結果（50 位受測者）

本研究將第二階段之驗證分為六個週期，並於各週期中分別匯入 80 份領域訓練文件，以觀察隨訓練文件之增加，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 5.14，而各驗證週期成員參考度解析之績效分佈趨勢如圖 5.69 所示。

表 5.14、虛擬社群成員參考度解析績效彙整

虛擬社群 成員參考度解析		各週期成員參考度解析驗證—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	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%

藉由對表 5.14 之觀察可得知，以每週期增加 80 份領域訓練文件為單位，平均每週期成員參考度解析之召回率及正確率驗證指標，其整體成長率分別為 8% 與 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，召回率及正確率將分別由第一週期之 26% 提升至 82%，綜合上述驗證結果之觀察，本研究所開發之虛擬社群成員參考度解析系統，具備學習能力與相當程度之正確性，亦即透過對虛擬社群成員參考度解析模組之使用，將確實為使用者提供具精確性成員參考程度參考之依據。

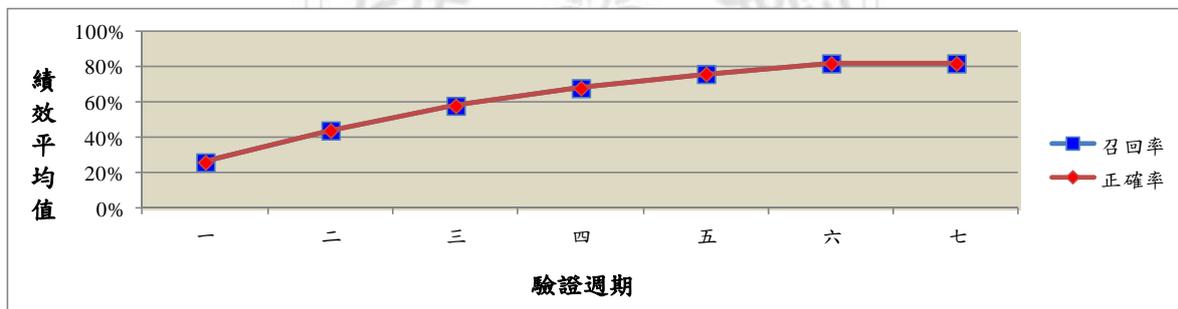


圖 5.69、各驗證週期成員參考度解析之績效分佈趨勢

### (III) 驗證結果整體分析

#### ➤ 驗證結果整體量化分析

首先，本系統於成員參考度解析之「第一階段驗證」中，當訓練文件之資料量為 200 份時，其正確率與召回率整體平均值為 26%，由此可知，此階段判定之績效仍欠佳。

然而，於「第二階段驗證」中，系統隨著不同週期及訓練量之增加，其各項數據出現顯著成長之趨勢，尤以第一週期至第二週期之成長狀況最為明顯，於第六週期時，成員參考度解析之正確率與召回率可提升至 82%，此外，由驗證結果中亦可得知，平均每週期成員參考度解析之召回率及正確率，其整體成長率為 9%。承接於上述，本研究於

不同訓練資料量下執行「成員參考度解析」核心功能績效之驗證，綜合於各週期驗證結果，可推論本系統之績效將因訓練資料之增加，持續且穩定地成長與提升，而本研究所建置之「虛擬社群成員參考度解析模組」亦實具其可行性。

最後，綜合兩階段之驗證成效，各項驗證指標之相關結果整理如表 5.15 所示。由表 5.15 整理之結果可得知，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，且各項驗證指標皆於第七週期內呈現收斂狀態，因此，以本研究所隨機選定之驗證資料為例，當系統使用約 600 筆訓練知識文章時，可使系統各項推論指標績效提升至 82% 之水準。

表 5.15、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
成員參考度解析召回率	59%	第七週期	11.2%	8.7%
成員參考度解析正確率	59%	第七週期	11.2%	8.7%

整體而言，由第一階段與第二階段之驗證分析中，可得知當系統進行「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，而系統之績效將隨著不同週期及訓練量之增加，呈現持續成長之狀況，並於最終達到穩定且良好之績效水平。故本研究所建置之「成員參考度解析模組」確實能準確判定成員領域知識可參考程度，並可依據判定結果，確實為使用者提供具精確性成員參考程度參考之依據。

#### ► 驗證結果整體質化分析

於前述系統績效之量化結果評估中，可得知本研究所開發虛擬社群成員參考度解析模組於「成員領域參考性排名」中，經由若干筆領域訓練資料之匯入，具備相當之可行性與績效。然而，本模組最終輸出之結果係為「成員領域可參考度」，雖於相關績效量化之驗證結果中，針對「成員領域參考性排名」進行驗證，可間接顯示本研究於判定成員領域參考性之績效，但針對成員領域參考性「程度」層面而言，由於成員參考度之「實際正確結果」（此部份乃表示「實際成員可參考性程度」）至今未能有具相當嚴謹性、公正性且標準化之代表性指標進行衡量，故此部份若仍以量化層面進行探討，其所獲致績效衡量結果之合理性及嚴謹性，將相對於以「排名」方式進行績效驗證來得低。有鑑於上述，為針對本模組「成員領域參考性程度」面向，進行相關之驗證，本研究將以「質

化」之層面進行討論，亦即面向「本模組分析得之成員參考度」及「搜搜問問計算得之成員綜合聲望」間之關係，由分析「方法」與「結果」兩層面之差異性，進行質化之深入探討，藉以分析及歸納出相關績效之結論，以使最終之系統驗證結果更具完善性與嚴謹性。

### 1.本系統成員參考度解析方法與結果

本系統係以 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「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，進行相對應群集之命名，藉此取得實具代表性之群集名稱。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，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，本研究由 50 位受測者中，再隨機挑選 10 位受測者（各領域一位受測者），經由本系統進行判定後，所得相關成員領域可參考程度彙整如表 5.16。

表 5.16、本系統判定 10 位受測者所得成員參考度之彙整表

領域類別	受測者	判定時採用歷史發文數量 (隨機挑選測試資料)	領域可參考程度
電腦/數碼	王戰勝	10	76%
生活家居	古順均	10	70%
奧運體育	Stinger	10	67%
休閒/愛好	陳清華	10	65%
藝術/文學	魏婕馨	10	86%
社會/人文	蘆天鵬	10	78%
教育/科學	章炳華	10	75%
健康/醫療	文馳流星	10	71%
商業/理財	杜江	10	61%
娛樂/明星	喵同學	10	61%

## 2. 搜搜問問成員參考度計算方法與結果

於搜搜問問虛擬社群之成員歷史資訊統計中，雖該社群未針對社群成員執行成員領域可參考度之分析，但針對於社群成員之領域專業性而言，使用者可透過對成員歷史資訊統計介面中之「綜合聲望」之參考，間接對目標成員領域專業性作初步之了解，因此，本研究即視「綜合聲望」為搜搜問問計算成員參考性之方式，「綜合聲望」計算方式係「依據目標成員在社區內之回答被採納率、滿意答案數、精華知識數、回答原創度，以及其他用戶對目標成員提供滿意答案的解決度評價，綜合計算出目標成員於社區內的綜合聲望值，代表目標成員於社區內回答問題的專業程度和貢獻水準，而綜合聲望最小值與最大值分別為 0 至 1200(引用於搜搜問問)」，當中，本研究乃根據搜搜問問所提及「綜合聲望最小值與最大值分別為 0 至 1200」之資訊，將各受測者之綜合聲望數據，轉換為標準化之百分比數值，亦即「目標成員綜合聲望值÷1200」，方以於後續進行差異化之對等比較。於此，針對隨機挑選得之 10 位受測者，經由搜搜問問蒐集得綜合聲望值之標準化處理後，所獲致相關成員領域可參考程度彙整如表 5.17。

表 5.17、搜搜問問衡量 10 位受測者所得成員參考度之彙整表

領域類別	受測者	計算時須採用歷史發文數量	領域可參考程度
電腦/數碼	王戰勝	59665	31%
生活家居	古順均	36781	24%
奧運體育	Stinger	15951	22%
休閒/愛好	陳清華	19943	23%
藝術/文學	魏婕馨	30242	37%
社會/人文	蘆天鵬	37106	33%
教育/科學	章炳華	14671	30%
健康/醫療	文馳流星	33232	26%
商業/理財	杜江	27246	18%
娛樂/明星	喵同學	16944	28%

## 3. 整體分析與比較

本系統與搜搜問問計算成員參考性程度之「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」，對知識需求者所提供之實際參考性與貢獻性，首先，本研究將針對「數據(分析結果)」層面進行對等之分析，以找出兩者之相異點，其次，則針對分析數據之差異進行討論，以取得造成不同結果之影響因素，爾後，本研究將根據探討得之影響因素，進行各影響因素質化層面之剖析，最後，藉由質化層面整體差異性之探討，以歸納

並統整出本系統與搜搜問問之整體績效結論。

承接於上述，為了使整體「成員參考度解析」之績效分析能以對等之方式進行，本研究即針對「本系統判定 10 位受測者所得成員參考度之彙整表（如表 5.16 所示）」，以及「搜搜問問衡量 10 位受測者所得成員參考度之彙整表（如表 5.17 所示）」進行對等之比較，相關結果如表 5.18 及圖 5.70 所示。

表 5.18、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%	

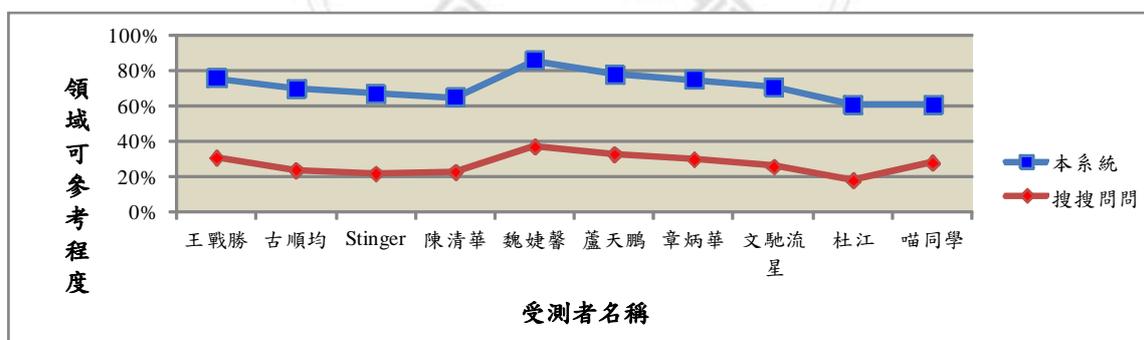


圖 5.70、受測者於本研究與搜搜問問領域可參考程度之分佈趨勢

藉由對表 5.18 及圖 5.70 之觀察，可從中觀測出三項問題點，第一，本系統與搜搜問問「採用發文相差總量」具備相當大幅度之差異，當中，本系統平均採用 10 個歷史發文即可執行判定，而搜搜問問平均須採用 29178 個歷史發文；第二，針對「領域可參考程度平均值」言之，本系統判定之結果平均落於 71%，而搜搜問問計算之結果則約平均落於 27% 間，故亦存在大幅度之差距；第三，反觀「領域可參考程度標準差」，本系

統判定與搜搜問問計算得之結果，標準差之差距約為 2.23%，呈現微幅度之差異。針對上述三大問題點，本研究於下即以質化之層面進行探討，以歸納並統整出本系統與搜搜問問之整體績效結論。

### 大幅度「採用發文相差總量」之質化探討

針對大幅度「採用發文相差總量」而言，主要乃因本系統與搜搜問問判定成員參考度方法之不同，而導致於判定時所須採用之歷史發文總量有所不同，針對於此，本研究成員參考度解析之方法係整合與改良向量空間模型、Zhang 等人 (2010) 所提之分群演算法，以及 Cilibrasi 與 Vitanyi (2007) 所提之 NGD 演算法，設計一套虛擬社群成員參考度解析模組，藉此透過對過去研究之借鑑與本研究之改良，開發成員參考度解析系統，以實具嚴謹與合理性之方式，判定得目標社群成員之領域可參考程度，因此，針對於「分析時採用歷史發文數量」之層面，本系統僅須隨機蒐集少許目標回答者於虛擬社群中之歷史發文，即可執行相關判定之動作，而就判定得結果之正確性而言，本系統亦存在相當程度之績效（詳述於「驗證結果整體量化分析」中）。另一方面，搜搜問問評估成員參考性之方式，主要透過對「綜合聲望」之計算，以衡量目標成員之領域可參考性程度，當中，計算方式係依據目標成員在社區內之回答被採納率、滿意答案數、精華知識數、回答原創度，以及其他用戶對目標成員提供滿意答案的解決度評價，綜合計算出目標成員於社區內的綜合聲望值，是故，針對於「分析時採用歷史發文數量」之層面，搜搜問問除了必須蒐集目標社群成員之全部歷史發文外，亦須針對目標社群成員所獲得之全部評價資訊進行統計，於此藉由對群眾意向之評估，計算得「綜合聲望」並作為成員領域可參考度。綜上而論，「採用發文相差總量」之所以存在大幅之差異，主要原因可歸咎於成員參考度解析方式之不同，而針對於判定方式進行討論，雖搜搜問問以群眾意向評估成員參考性之作法，具備其參考性，但於計算過程中所須採用之巨量資料（全部目標成員歷史發問及評價資訊），將為系統帶來龐大之負擔。

### 大幅度「領域可參考程度平均值」差異之質化探討

針對大幅度「領域可參考程度平均值」進行討論，就成員領域可參考度最終輸出結果言之，搜搜問問以群眾意向計算得之結果，雖可作為同領域中成員專業性排名之鑑別，但因其輸出領域參考度結果偏低下之故（平均落於 27%），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之領域專業性程度，此外，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解析方法論，與搜搜問問自行發展之計算方法相較之，具備更為嚴謹與合理之理論架構，因此，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。

#### 微幅「領域可參考程度標準差」之質化探討

針對微幅差距「領域可參考程度標準差」進行討論，本研究隨機挑選 10 位受測者，分別對其進行本系統及搜搜問問之成員參考度解析，並依序獲致各領域類別之「領域可參考程度標準差」為 7.94%（本系統）與 5.71%（搜搜問問），由上述獲致標準差皆趨近於 0 之結果，可發現本系統與搜搜問問執行成員參考度之判定後，各領域間「領域可參考程度」無太大之差異程度，於此亦進一步探討得本系統與搜搜問問之成員參考度解析之方法，及所獲致之判定結果，可廣泛適用於虛擬社群中之各領域類別，並同時提供知識需求者具備相當參考性之數據。承接於上述之討論，本系統「領域可參考程度標準差」為 7.94%，搜搜問問為 5.71%，當中，兩者標準差呈現微幅之差異（相差 2.23%），就「領域可參考程度標準差」言之，該數值越趨近於 0，將可視為各領域間判定之結果差異越小，亦即具備更高之績效，故僅就標準差數據層面進行比較，搜搜問問之績效將優於本系統，然而，雖搜搜問問取得之標準差更趨近於 0，但針對當中之樣本數而言，由於本研究方法論之設計具備相當之嚴謹性，於樣本數採用總數之層面，將遠低於搜搜問問計算時所採用之樣本總數，而也因樣本數採用數量較少，致使本系統最終「領域可參考程度標準差」與搜搜問問相較之更不趨近於 0。綜上而論，由於「領域可參考程度標準差」趨近於 0 之緣故，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，而就分析樣本採納總數之層面，因本系統運用之方法論具備相當之嚴謹性，故分析樣本採納總數將可大幅低於搜搜問問，於此，就結果面言之，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，即可完成實據績效（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。

#### 整體績效質化探討之總結

綜合於上述三大項問題點之質化探討結果。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（全部目標成員歷史發問及評價資訊），將為系統帶來龐大之負擔，而本研究僅須隨機蒐集少許目標回答者於虛擬社群中之歷史

發文，即可執行相關判定之動作，而就判定得結果之正確性而言，本系統亦存在相當程度之績效。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之領域專業性程度，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解析方法論，具備更為嚴謹與合理之理論架構，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，於成員參考度解析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，即可完成實據績效（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。

### 5.2.3—(C) 虛擬社群知識群聚度判定管理意涵之探討

本研究所開發之「虛擬社群知識群聚度判定系統」乃包含「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」，藉由上述兩模組績效實質之獨立驗證，可得知本系統於「問答契合度解析」及「成員參考度解析」層面具備相當之可用性與績效。然而，本研究係以探討「虛擬社群知識群聚度」為目標，亦即提供「問答契合度」及「成員參考度」解析結果予使用者，令使用者藉以評估知識群聚度，故本研究開發之模式與系統實存相當之關聯性，若僅以獨立方式進行個別模組織績效驗證，實難評估本研究發展整合性模式與系統於實務應用面之績效，有鑑於此，本研究即整合「問答契合度解析」及「成員參考度解析」兩核心模組，以整體「虛擬社群知識群聚度判定」層面為主軸，以案例為導向進行實務應用之探討，藉以由整合性層面，討論並評估本研究發展「虛擬社群問答契合度判定」模式與系統之管理意涵。

#### ► 整體系統驗證之設計與實作說明

本系統係整合「問答契合度解析模組」與「成員參考度解析模組」以判定虛擬社群知識群聚度，使用者可藉問答契合度解析模組之應用，取得目標討論串中提問與回答之契合程度，使用者亦可應用成員參考度解析模組，分析得目標成員之領域可參考程度，而透過對本系統分析得成員參考度與問答契合度之參考，使用者可以此為依據衡量虛擬社群知識群聚（凝聚）之狀況，因此，雖本系統兩核心模組乃獨立運行之，但最終分析得之問答契合度，以及成員參考度結果具備相當程度之關聯性，故針對於「問答契合度」

與「成員參考度」兩層面分析結果之數據，實須進行整合性之深入探討。

有鑑於上述，本研究以使用者實際應用層面為切入點進行考量，「問答契合度」與「成員參考度」分析結果數據之關係大致可劃分為「1.問答契合度高、成員參考度高」、「2.問答契合度高、成員參考度低」、「3.問答契合度低、成員參考度高」，以及「4.問答契合度低、成員參考度低」等狀況，是故，本研究於「搜搜問問」中針對上述四種狀況，以隨機方式進行共 8 筆測試資料之蒐集（各狀況 2 筆資料），並透過對「問答契合度解析模組」與「成員參考度解析模組」之運用（已訓練完成之系統），分析與討論各狀況中問答契合與成員參考度之關聯性，以及對知識群聚度之影響，以於最終進行整體化質化系統管理意涵層面之探討。相關管理意涵探討之示意圖如圖 5.71 所示。



圖 5.71、本研究整體系統管理意涵探討之示意圖

➤ 整體系統驗證之實作

首先，本研究由「搜搜問問」虛擬社群中，根據「1.問答契合度高、成員參考度高」、「2.問答契合度高、成員參考度低」、「3.問答契合度低、成員參考度高」，以及「4.問答契合度低、成員參考度低」等狀況，隨機蒐集共 8 筆測試資料，當中，各筆目標測試資料必須包含「回答者」、「提問內容」、「回答內容」、「討論串領域類別」，以及隨機 10 筆「回答者於目標領域類別之其他回答內容」，以作為系統判定問答契合度與成員參考度之依據，如表 5.19 與表 5.20 所示。

表 5.19、虛擬社群知識群聚度判定之測試資料(1) (部分內容)

領域類別	回答者	提問內容	回答內容
電腦/數碼	楊威濤	ps 裡 Lab 通道的 a 和 b 是什麼意思?L 是指明度，那麼 a 和 b 呢？	Lab 通道中 a 表示從紅色至綠色的範圍顏色，b 表示從黃色至藍色的範圍顏色...
健康/醫療	張清熠	為什麼人會控制不住很多事？	有因就有果，事物之間也是矛盾的存在。第一種：改變處理信息的方法，類同...
生活家居	東流水	吃菠蘿為什麼要用鹽水泡？	菠蘿的果肉除富含維生素 C 和糖分以外還含有不少有機酸，如蘋果酸、檸檬酸等，另外...
奧運體育	吃貨	艾弗森在熱火復出的可能性大嗎???	我覺得應該不會，如果是打算讓 AI 打首發的 PG，那熱火就一個球，本來就不夠...
社會/人文	陳明仁	張居正的父親叫什麼？	叫張文明,字治卿。希望答案您還滿意
教育/科學	信子飛翔	我是體育院校畢業的，報考人民警察公務員，應報考什麼職務？	你好！就報考警察職位就好了啊！
商業/理財	QQ_740588793	我是新手什麼都不懂，你可以教下我嗎	可以啊，加我名字好友，有什麼不明白的就問我吧
休閒/愛好	破鏡難圓	我們家的狗兩個月大，不吃東西，餵葡萄糖水一喝就吐。...	煮點豬骨粥給它喝

表 5.20、虛擬社群知識群聚度判定之測試資料(2) (部分內容)

領域類別	回答者	回答編號	回答內容 (隨機挑選)
電腦/數碼	楊威濤	1	一款常用的視頻編輯軟件，由 Adobe 公司推出。這款軟件廣泛應用於廣告製作和電視節目製作中。其最新版本為 Adobe Premiere...
		2	這個要是簡單純背景的，你用魔術棒點擊然後刪除填充藍色即可。要是複制背景按照下面方法處理...
		3	這個別直接用 PHOTOSHOP 去打開閃圖，而是用 PHOTOSHOP 附帶的一個閃圖製作和編輯軟件 ImageReady 打開...
		4	下面給你歸納幾點： 1.減少本透明物體的使用率 2.減少大型圖片的移動，這個是最最耗資源的，罪魁禍首！ 3.減少大型元件的...
		5	在 PHOTOSHOP 裡，羽化是針對選區的一項編輯，羽化原理是令選區內外銜接的部分虛化。起到漸變的作用從而達到自然銜接的效果。...
		6	不是你這個理解。其實要是你重裝系統，其實系統本身自帶了所有設備的初始驅動，所以上安裝後即可使用。...
		7	這個你可以利用 IE 緩存來下載這個視頻電影。首先在線播放，完成後別關閉或刷新，然後在桌面 IE 圖標右擊，選擇屬性...
		8	這個與破解版和使用版沒關係，這是軟件本身的限制，要是錄製文件過大相應佔用的系統資源就更大，...
		9	FLV 是視頻文件，不可用 FLASH 編輯，它只可以調用。要是想編輯你必須用格式轉化器把 FLV 轉換為 AVI 格式，用視頻編輯軟件...
		10	這個那你插入圖片後然後在圖片上單擊，即可看到變化框，你可進行調節縮放進行壓縮或者利用系統畫筆工具...

次之，待完成測試資料之蒐集，本研究即運用已訓練完成之「虛擬社群問答契合度解析」與「虛擬社群成員參考度」模組，分別解析各筆測試資料之問答契合度與成員參考度，並將數據分析之結果彙整如表 5.21，當中，前述已完成訓練之模組，其系統資料庫乃由本研究所蒐集 680 筆領域訓練資料建置而得，相關訓練過程已詳述於各模組之獨立驗證中。

表 5.21、虛擬社群知識群聚度系統判定結果

回答者	目標問答契合度	成員參考度	分析結果隸屬狀況
楊威濤	88%	83%	1.問答契合度高、成員參考度高
張清熠	81%	85%	1.問答契合度高、成員參考度高
東流水	86%	14%	2.問答契合度高、成員參考度低
吃貨	84%	21%	2.問答契合度高、成員參考度低
陳明仁	43%	86%	3.問答契合度低、成員參考度高
信予飛翔	29%	88%	3.問答契合度低、成員參考度高
QQ_740588793	16%	26%	4.問答契合度低、成員參考度低
破鏡難圓	23%	15%	4.問答契合度低、成員參考度低

接著，針對於問答契合度與成員參考度之分析結果，本研究乃以測試資料為例，並以使用者實務應用之觀點，進行「虛擬社群知識群聚度」層面之討論，以評估各類狀況對使用者評估知識群聚度所造成之影響，並於最終面向本研究虛擬社群知識群聚度分析系統層面，進行整體之探討，以歸納得本系統於實務應用層面之管理意涵。

### 1.問答契合度高、成員參考度高狀況之討論

- 參考價值評估：具備高品質問答內容、未離題回答內容，且回答者具備領域專業性，故皆可視為正確解答。
- 知識需求者之實務應用：於知識內容及知識發佈者層面，皆為使用者提供實具參考之價值。
- 社群管理者之實務應用：獎勵目標成員，並增加目標回答曝光度，以為社群累積群眾智慧。

### 2.問答契合度高、成員參考度低狀況之討論

- 參考價值評估：具備高品質問答內容與未離題回答內容，雖成員參考度較低，但仍可作為候選解答之參考。
- 知識需求者之實務應用：應將之視為「參考資料」，並於蒐集多筆相似參考資料後，評估最終參考價值。
- 社群管理者之實務應用：獎勵目標成員，提供相應社群地位，藉以鼓勵其分享具品質之知識。

### 3.問答契合度低、成員參考度高狀況之討論

- 參考價值評估：由於具備低品質及離題之回答內容，故雖成員參考度高，但仍不具

備參考價值。

- 知識需求者之實務應用：可參考成員其他之領域回答，但不應參考目標回答內容。
- 社群管理者之實務應用：懲戒目標成員，降低社群地位，以進行實質公正性社群地位之管理。

#### 4.問答契合度低、成員參考度低狀況之討論

- 參考價值評估：具備低品質與離題之回答內容，且成員參考度低下，故不具備參考價值。
- 知識需求者之實務應用：應直接排除目標回答與成員相關之發布內容，以避免擷取錯誤知識。
- 社群管理者之實務應用：不應目標成員分享大量知識，而給予社群地位，以避免誤導知識需求者，並維持社群之正常運作。

#### ➤ 虛擬社群知識群聚度系統判定結果之整體探討

本研究開發之虛擬社群知識群聚度判定模式，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，藉由對此兩模組之運行，將可分別解析得「問答契合度」與「成員參考度」，而面向解析得之兩結果言之，將衍生四種不盡全然相同之狀況，分別為「1.問答契合度高、成員參考度高」、「2.問答契合度高、成員參考度低」、「3.問答契合度低、成員參考度高」，以及「4.問答契合度低、成員參考度低」等狀況，經由相關之深入探討後，針對狀況1中所取得之系統判定結果，於知識內容層面及知識發佈者層面，皆為使用者提供實具參考之價值；於狀況2之實務應用情況中，使用者可將目標回答作為參考資料，並於蒐集多筆相似參考資料後，以評估最終參考價值；於狀況3中，因此情況於最低知識採納限度條件中即不符合，故雖另一層面中成員具備高度之參考性，該筆目標問答仍不具相應之參考價值；狀況4於實際虛擬社群運作情況中，目標回答大多不具參考之價值，故使用者可藉本系統過濾之。

整體而論，藉由對上述四種分析狀況之討論，可得知本研究所開發之模式與系統，雖劃分為兩模組獨立執行之，但綜合於兩模組執行後各項組合結果之運用，亦可以整體化知識群聚之層面，供使用者衡量與評估目標問答之參考價值，於此，針對社群管理者而言，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，就知識需求者言之，則降低其篩選知識時所需花費之成本，且同時提升其獲取

實具契合性與品質知識之機率，綜合觀之，將因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。



## 第六章、結論與未來發展

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，進而提升虛擬社群之整體發展與造就群眾智慧之累積。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，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群眾智慧運作機制中面臨之問題，藉由對近期相關文獻之借鑑與改良，本研究發展一套「虛擬社群知識群聚度判定」模式，當中，核心模組主要分為「虛擬社群問答契合度解析模組」與「虛擬社群問答契合度解析模組」，於前者中，本研究以群眾智慧之影響性為切入點，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，本研究乃整合上述兩模組，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，提供相關數據予知識擷取者，以減少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。另一方面，本研究除發展模式與方法論外，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。以下將於第 6.1 小節總結本研究所完成之工作與任務，並於第 6.2 小節歸納本研究未來發展之議題與方向。

### 6.1 論文總結

依據第 1.2 小節之「研究步驟」所述，本論文可將完成工作分為三項任務，以下乃分別說明各項任務之成果。

#### 1. 發展「虛擬社群知識群聚度判定」方法論

本研究發展之模式主要針對「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，於「虛擬社群問答契合度解析模組」中，本研究乃結合「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」，並於藉數據之整合，分析並計算得「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模組」中，本研究透過對虛擬社群成員歷史發文所屬領域，及其與該領域中其他討論串所對應之分群結果，當中群集間相似程度之解析，以取得相關領域性群集間之內文與語意關係距離，並藉「正規化」方式取得相互比較之基準點，以於最終獲致實具參考價值目標虛擬成員領域知識可參考程度。

## 2. 建置「虛擬社群知識群聚度判定系統」

本研究乃以「虛擬社群知識群聚度判定模式」方法論為依據，運用 JSP (Java Server Pages) 語法進行開發，並利用 SQL Server 2008 來存取資料庫，以建置一套「虛擬社群知識群聚度判定系統」，令使用者能進行相關欲分析資料與訓練文件之上傳與維護，並透過對系統兩核心模組之運用，藉由自動化方式分析得問答契合度及成員參考度，以提供使用者相關之參考數據。

## 3. 方法論與系統績效驗證

為驗證本研究所提出方法論與系統之正確性與有效性，本研究以「搜搜問問」虛擬社群之真實問答資料作為驗證(測試)資料之樣本，並以該虛擬社群中之精華知識文章，作為領域訓練資料之樣本，分別針對「虛擬社群問答契合度解析」與「虛擬社群成員參考度解析」兩項子議題，進行獨立之驗證，確保兩相議題間之驗證結果不相互影響，並於最終績效驗證指標中，分別獲致問答契合度解析召回率 80%、正確率 85%，以及成員參考度解析召回率 82%、正確率 82% 等結果，此外，本研究亦針對「虛擬社群知識群聚度判定」主題，以案例為導向探討整合「問答契合度解析」及「成員參考度解析」兩層面所發展之「虛擬社群知識群聚度判定系統」，於應用層面之實質效用，以及相關管理意涵，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。

綜合於各小節之系統驗證結果，包含「(A)虛擬社群問答契合度解析」、「(B)虛擬社群成員參考度解析」，以及「(C)虛擬社群知識群聚度判定」相關量化之分析與質化層面之探討，本研究所提出之「虛擬社群知識群聚度判定」方法論與系統，於實務應用層面中具備相當之績效與可行性。以下即以「理論方法」、「技術開發」與「實務應用」等三種層面說明本研究之具體貢獻與成效。

## 理論方法層面

本研究乃以虛擬社群討論串資料為基礎，發展一套「虛擬社群知識群聚度判定」模式，以發問者之「發問標題」、「發問內容」與「討論串所屬領域」，回答者之「回答內容」，以及虛擬社群成員之「歷史發文」等資料作為分析之基礎，解析得「問答契合度」與「成員參考度」，藉以由虛擬社群知識群聚之觀點，供使用者進行整合性之參考。本方法論相關重點成效乃歸納如下：

- 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。
- 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。

## 技術開發層面

本研究乃以「虛擬社群知識群聚度判定」方法論為依據，並以JSP(Java Server Pages)以及SQL Server 2008等系統開發工具建置「虛擬社群知識群聚度判定」系統，其系統具體成效乃歸納如下：

- 本系統所開發之「虛擬社群知識群聚度判定」乃建置於網際網路中，並可執行「問答契合度解析」與「成員參考度解析」，藉此取得目標問答中之提問品質、回答品質與問答契合程度，以及目標社群成員各領域知識之可參考程度。
- 由驗證結果得知，於量化層面中，本系統之各項驗證指標皆具學習能力，當使用一定訓練資料量時，系統即可有效判定問答契合度及成員參考度；於質化層面中，藉由多方性觀點之切入，亦可探討得本系統於實務應用層面中之管理意涵，且具備相當程度之可行性。

## 實務應用層面

本研究可準確分析虛擬社群討論串中，問答品質與問答契合之程度，並透過對虛擬社群成員領域歷史發文之解析，計算得實具參考價值之成員領域知識可參考程度，以提

升使用者於資訊爆炸的現今，擷取與管理相關知識內容之效能與效率，因此，本研究於實務應用層面，可應用於線上虛擬社群中，具體貢獻與成效乃歸納如下：

- 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。
- 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。
- 於整體虛擬社群之層面，將因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。

## 6.2 未來展望

依據第 6.1 小節所述，本研究乃完成研究步驟所規劃之各項任務，並提出「虛擬社群知識群聚度」方法論與「虛擬社群知識群聚度」系統理論層面、技術層面以及實務應用層面之成效與貢獻。而後續發展方面，綜合本論文之研究成果與既有文獻再結合未來資訊發展之拓展，發現本論文尚有若干研究主題具有深入研究之價值，歸納未來發展方向如下：

1. 發展整合社群成員發文情緒解析之虛擬社群知識群聚度判定模式：本研究兩核心模組分析結果中存在如「問答契合度低；成員參考度高」等狀況，故期望未來由成員發文情緒之觀點進行切入與探討，進一步以使用者情緒為導向分析其對知識品質之影響性，以發展更具精確性與多方性之判定模式。
2. 發展一套跨語言虛擬社群知識群聚度判定模式：現階段本研究主要解析中文虛擬社群中之討論內容，然而中文文字之詞彙用法與各國語言不同，故期望發展跨語言問答契合解析技術，以協助系統進行跨國語言虛擬社群資料之解析。
3. 發展一套整合圖片資訊解析之虛擬社群知識群聚度判定模式：本研究方法論以分析討論串中之文字內容為主，而欠缺對討論內容中所包含圖片隱含資訊之討論，故期望未來可發展整合圖片資訊解析之模式。

## 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. Liang, T.-P., Yang, Y.-F., Chen, D.-N. and Ku, Y.-C., 2008, "A semantic-expansion approach to personalized knowledge recommendation," *Decision Support Systems*, Vol. 45, No. 3, pp. 401-412.
41. 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.
42. 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.
43. 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.
  44. 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.
  45. 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.
  46. 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.
  47. 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.
  48. 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.
  49. 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.
  50. Ni, X., Lu, Y., Quan, X., Wenyan, 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.
  51. 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.
  52. 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.
  53. Schall, D., 2012, "Expertise ranking using activity and contextual link measures," *Data*

- & *Knowledge Engineering*, Vol. 71, No. 1, pp. 92-113.
54. Sorg, P. and Cimiano, P., 2012, "Exploiting Wikipedia for cross-lingual and multilingual information retrieval," *Data & Knowledge Engineering*, Vol. 74, pp. 26-45.
  55. 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.
  56. Teplovs, C., 2008, "The Knowledge Space Visualizer: A tool for visualizing online discourse," In *Proceedings of the International Conference of the Learning Sciences*.
  57. 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.
  58. 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.
  59. 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.
  60. 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*.
  61. 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.
  62. 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.
  63. 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.
  64. 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.

65. 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.
66. 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.
67. 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.
68. Yang, S. J. H. and Chen, I. Y. L., 2008, "A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network," *International Journal of Human-Computer Studies*, Vol. 66, No. 1, pp. 36-50.
69. Zhang, W., Yoshida, T., Tang, X. and Wang, Q., 2010, "Text clustering using frequent itemsets," *Knowledge-Based Systems*, Vol. 23, No. 5, pp. 379-388.
70. 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.
71. 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.
72. 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.
73. 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.
74. 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.

75. 中研院 CKIP 中文斷詞系統，<http://ckipsvr.iis.sinica.edu.tw/>。



## 附錄、系統功能操作說明

本研究所發展之虛擬社群知識群聚度判定系統乃以本文中第 4.4.1 節所提出系統功能流程為依據，並開發「(A)虛擬社群資料維護模組」、「(B)虛擬社群問答契合度解析模組」與「(C) 虛擬社群成員參考度解析模組」等三大模組，以下分別針對此三大模組中各功能進行詳細之說明。

### (A) 虛擬社群資料維護模組

為使權限內使用者方便維護領域文件、討論串，以及成員歷史發文等相關資料，本研究乃開發「虛擬社群資料維護模組」，本模組乃提供使用者於線上新增、查詢、刪除及修改各資料，而虛擬社群資料維護模組包含「領域文件」、「討論串」，以及「成員歷史發文」三層面資料之「新增」、「查詢」、「修改」與「刪除」等功能。「新增」乃提供使用者基本資料新增與維護於資料庫之功能；「查詢」乃提供使用者查詢所有資料之內容，以方便使用者瞭解系統內各項資料之維護結果；「修改」與「刪除」乃提供使用者進行修改與維護錯誤或缺乏時效性之資料，進而保持資料之正確性與即時性。

#### (A-1) 領域文件資料新增功能—程式說明

當權限內使用者進入領域文件資料新增功能時，系統即呈現領域文件基本資料之欄位，以提供使用者輸入相關之基本資料（如圖 A.1 所示），而當使用者送出基本資料時，系統將針對使用者提供之資料，進行領域文件內文之資料預處理，包含領域文件內文之斷詞，以及有效詞之篩選，於此同時，系統亦將使用者提供未經預處理之原始資料，以及經系統預處理後之結果，儲存並維護於資料庫中（如圖 A.2 所示）。

舉例而言，權限內使用者進入此功能時，系統即呈現領域文件標題與領域文件內容欄位，供使用者輸入如標題「電腦老師自己斷網...」及內容「有沒有什麼檢測...」等基本資料，此外，系統亦提供領域文件所屬類別與上傳時間之下拉式選單，供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013 年 12 月 20 日」，使用者亦可點擊「瀏覽...」，以上傳於「D:\data」中「\*mht」領域文件之原始檔案，於此，當使用者完成相關資料維護並點擊「確認」後（如圖 A.1 所示），系統將針對使用者上傳原始檔案，進行斷詞與有效詞篩選資料預處理作業，分別取得如「有沒有(ADV) 什麼(DET) 檢測(Nv)...」，以及「方法；家；路由器...」等內文斷詞與有效詞篩選結果，於此，系統亦同時將未經預處理原始資料，以及經預處理後結果儲存並維護於資料庫中（如圖 A.2 所示）。



圖 A.1、領域訓練文件基本資料之輸入



圖 A.2、領域訓練文件資料之預處理與新增

### (A-2)領域文件資料查詢功能—程式說明

當權限內使用者進入領域文件資料查詢功能時，系統即呈現條件式查詢之介面，以提供使用者進行領域文件資料之查詢，而當使用者執行查詢功能後，系統亦將呈現相關領域文件查詢之結果予使用者參考（如圖 A.3 所示）。

舉例而言，權限內使用者進入此功能時，系統將提供條件式搜尋功能予使用者，以進行資料庫內討論串之搜尋，當使用者未輸入資訊而直接點擊「查詢」按鈕時，系統將呈現目前存在於資料庫中之全部領域文件，如文件編號「1271150632」、標題「MLB／波沙達...」、內容「只有 5%...」，以及領域類別「運動」等查詢結果供使用者參考（如圖 A.3 所示）。



圖 A.3、領域文件資料之查詢

### (A-3)領域文件資料修改功能—程式說明

當權限內使用者進入領域文件資料修改功能時，系統將呈現條件式查詢之介面，以提供使用者進行欲修改領域文件資料之查詢，當使用者查詢得相關領域文件資料後，可選擇所需之領域文件以進行資料之修改（如圖 A.4 所示），使用者可修改標題、內容，以及領域類別等資訊，於此，若使用者僅修改標題或領域類別資訊，系統將直接執行資料修改之動作（如圖 A.5 與圖 A.6 所示）；若使用者修改領域文件之內容，因內文資訊將變更（如圖 A.7 與圖 A.8 所示），故系統將再次執行斷詞與有效詞篩選等資料預處理動作，待完成資料之預處理後，系統即儲存修改後之相關內容（包含資料預處理結果）於資料庫中，以完成領域文件之修改，並呈現修改後之資訊予使用者參考（如圖 A.9 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲修改之領域文件，其編號為「1271150632」時，於點擊「資料修改」後（如圖 A.4 所示），系統即以開新視窗之方式，呈現修改領域文件資料之介面，並自動將原始之標題與內容資訊填入文字方塊中，於此，使用者點擊領域類別之下拉式選單，於選擇「生活娛樂」並點擊「確認修改」後（如圖 A.5 所示），系統即執行領域文件資料之修改，將原先領域類別為「運動」之文件，修改為「生活娛樂」（如圖 A.6 所示）；另一方面，若使用者將領域文件之內容，由「只有 5%...」修改為「MLB／波沙達...」並點擊「確認修改」後（如圖 A.7 所示），因文件內容將被修改，故系統必須再次執行斷詞與有效詞篩選之動作，於此，於使用者點擊「執行資料預處理」後（如圖 A.8 所示），系統即完成修改後內容資料預處理，並同時將修改後領域文件「1271150632」相關資訊，更新並維護於資料庫中（如圖 A.9 所示）。



圖 A.4、欲修改領域文件資料之查詢



圖 A.5、領域文件資料之修改(1)



圖 A.6、完成領域文件資料之修改(1)



圖 A.7、領域文件資料之修改(2)



圖 A.8、領域文件資料預處理確認介面



圖 A.9、完成領域文件資料之修改(2)

#### (A-4)領域文件資料刪除功能—程式說明

當權限內使用者進入領域文件資料刪除功能時，系統將呈現條件式查詢之介面，以提供使用者進行欲刪除領域文件資料之查詢，當使用者查詢得相關領域文件資料後，可選擇所需之領域文件以進行資料之刪除（如圖 A.10 所示），於此，系統將再次與使用者確認其欲刪除之領域文件（如圖 A.11 所示），當使用者確認並執行刪除功能後，系統即完成領域文件之刪除（如圖 A.12 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲刪除之領域文件，其編號為「1271150632」時，於點擊「刪除」後（如圖 A.10 所示），系統即以開新視窗之方式，再次與使用者確認欲刪除之領域文件，其編號為「1271150632」，標題為「MLB／波沙達...」，於此，當使用者點擊「確認刪除」後（如圖 A.11 所示），系統即完成領域文件「1271150632」之刪除（如圖 A.12 所示）。



圖 A.10、欲刪除領域文件資料之查詢



圖 A.11、確認欲刪除領域文件之資料



圖 A.12、完成領域文件資料之刪除

#### (A-5) 討論串資料新增功能—程式說明

當權限內使用者進入討論串資料新增功能時，系統即呈現討論串基本資料之欄位，以提供使用者輸入相關之基本資料（如圖 A.13 所示），而當使用者送出基本資料後，系統將呈現使用者所輸入提問之基本資料，並根據使用者輸入之回答筆數，提供相應數量回答內容之輸入欄位（如圖 A.14 所示），當使用者完成所有資料之輸入後，首先，系統將針對討論串之提問內容，以及各筆回答內容，執行包含內文斷詞與有效詞篩選之資料預處理動作，於此同時，系統亦將使用者提供未經預處理之原始資料，以及經系統預處理後之結果，儲存並維護於資料庫中（如圖 A.15 與圖 A.16 所示）。

舉例而言，權限內使用者進入此功能時，系統即呈現提問標題與內容，以及回答筆

數等欄位，供使用者輸入如提問標題「各路大仙...」、提問內容「各路大仙 看看我的電腦網速...」，以及回答筆數「5」等基本資料，此外，系統亦提供討論串所屬類別與上傳時間之下拉式選單，供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013 年 12 月 22 日」，於此，當使用者完成相關資料之輸入並點擊「確認」後（如圖 A.13 所示），系統將呈現使用者所輸入提問之基本資料，並根據使用者方才輸入之回答筆數「5」，創建 5 個回答內容輸入之欄位，使用者根據討論串中回答之順序，依序輸入「網速和網絡帶寬...」、「看看有多少進程」、「看看有沒有偷網用的...」、「網速是和運營商...」與「下個魯大師」等回答內容，並於點擊「確認」（如圖 A.14 所示），系統將針對領域討論串「1034646889」執行斷詞與有效詞篩選等資料欲處理動作，以分別取得如「各(DET) 路(M) 大仙(N)...」，以及「大仙；我；電腦...」等內文斷詞與有效詞篩選結果，並同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中，接著，當使用者點擊「下一步」後（如圖 A.15 所示），系統即針對領域討論串「1034646889」之各筆回答執行資料預處理動作，以分別取得如「網速(N) 和(C) 網絡(N)...」，以及「網速；網絡；帶寬...」等資料預處理結果，亦同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中（如圖 A.16 所示）。

The screenshot shows a web browser window with the EDS logo in the top left. The main content area is titled "討論串資料新增" (Add Discussion Thread Data). Below the title, there is a form with the following fields:

- 提問標題** (Question Title): 各路大仙 看看我的電腦網速...
- 提問內容** (Question Content): 各路大仙 看看我的電腦網速為什麼越來越慢 本人小區300M寬頻換了 軟件新裝一天後 還是沒效果
- 回答筆數** (Number of Answers): 5
- 討論串所屬類別** (Discussion Thread Category): 電腦 數碼
- 上傳時間** (Upload Time): 2013年 12月 22日

At the bottom of the form, there are two buttons: "確認" (Confirm) and "清除" (Clear).

圖 A.13、討論串基本資料之輸入



圖 A.14、討論串各筆回答資料之輸入



圖 A.15、討論串提問資料之預處理與新增



圖 A.16、討論串各筆回答資料之預處理與新增

## (A-6)討論串資料查詢功能—程式說明

當權限內使用者進入討論串資料查詢功能時，系統即呈現條件式查詢之介面，以提供使用者進行討論串資料之查詢，而當使用者執行查詢功能後，系統亦將呈現相關討論串查詢之結果予使用者參考（如圖 A.17 所示），此外使用者亦可檢視討論串之詳細資訊（如圖 A.18 所示）。

舉例而言，權限內使用者進入此功能時，系統將提供條件式搜尋功能予使用者，以進行資料庫內討論串之搜尋，當使用者未輸入資訊而直接點擊「查詢」按鈕時，系統將呈現目前存在於資料庫中之全部討論串，如文件編號「1」、標題「新手慢跑...」、提問內容「小弟因為心肺...」，以及領域類別「運動」等查詢結果供使用者參考，當使用者點擊「詳細資訊」時，系統將以開新視窗之方式，呈現包含討論串基本資料，以及回答基本資料等詳細資訊，供使用者參考（如圖 A.18 所示）。



圖 A.17、討論串資料之查詢



圖 A.18、檢視詳細討論串資訊

#### (A-7) 討論串資料修改功能—程式說明

當權限內使用者進入討論串資料修改功能時，系統將呈現條件式查詢之介面，以提供使用者進行欲修改討論串資料之查詢，當使用者查詢得相關討論串資料後，可選擇所需之討論串以進行資料之修改（如圖 A.19 所示），於執行修改動作前，系統將呈現討論串之詳細資料予使用者參考（如圖 A.20 所示），而使用者可針對討論串之標題、提問內容與領域類別，以及討論串中之回答內容進行修改，接著，將因使用者之修改情況而有不同之執行方式：

1. 若使用者僅修改討論串標題或領域類別資訊，系統將直接執行資料修改之動作（如圖 A.21 與圖 A.22 所示）。
2. 若使用者僅修改討論串之提問內容，因內文資訊將變更（如圖 A.23 與圖 A.24 所示），故系統將再次執行斷詞與有效詞篩選等資料預處理動作，待完成資料之預處理後，系統即儲存修改後之相關內容（包含資料預處理結果）於資料庫中，以完成提問內容之修改，並呈現修改後之資訊予使用者參考（如圖 A.25 所示）。
3. 若使用者僅修改討論串之回答內容，因內文資訊將變更（如圖 A.26 與圖 A.27 所示），同於第二種情況，系統亦須針對修改後之內容執行資料預處理動作，待完成資料之預處理後，系統即儲存修改後之回答內容於資料庫中，以完成回答內容之修改，並呈現修改後之資訊予使用者參考（如圖 A.28 所示）。
4. 若使用者同時修改討論串之提問與回答內容，因提問與回答內文資訊將同時變更（如圖 A.29 與圖 A.30 所示），故系統必須同時針對修改後之提問與回答內容，再次執行資料預處理動作，待完成資料之預處理後，系統即儲存修改後之相關內容於

資料庫中，以完成提問與回答內容之修改，並呈現修改後之資訊予使用者參考（如圖 A.31 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲修改之討論串，其編號為「1783227754」時，於點擊「修改」後（如圖 A.19 所示），系統即以開新視窗之方式，呈現包含「討論串基本資料」，以及「回答基本資料」等詳細資訊，當使用者點擊「修改資料」後（如圖 A.20 所示），系統將呈現修改討論串資料之介面，並自動將原始之資訊，包含討論串之標題「請教進入...」、提問內容「各位前輩...」，以及回答內容如「唯一便宜...」等資訊填入文字方塊中，於此，使用者點擊領域類別之下拉式選單，於選擇「科技」並點擊「確認修改」後（如圖 A.21 所示），系統即將原先領域類別為「相機與攝影」之討論串，修改為「科技」（如圖 A.22 所示）；另一方面，若使用者將討論串之提問內容，由「各位前輩...」修改為「小弟想開始...」並點擊「確認修改」後（如圖 A.23 所示），因提問內容將被修改，故系統必須再次執行斷詞與有效詞篩選之動作，於此，於使用者點擊「執行資料預處理」後（如圖 A.24 所示），系統即完成修改後提問內容資料之預處理，呈現相關結果予使用者，並同時將修改後討論串「1783227754」之相關資訊，更新並維護於資料庫中（如圖 A.25 所示）；此外，若使用者將討論串之回答內容，由「唯一便宜...」、「衷心建議...」，以及「F4s 或 F4E...」修改為「值得買的...」、「Nikon 當代」與「不會後悔...」並點擊「確認修改」後（如圖 A.26 所示），因回答內容將被修改，故系統必須再次執行斷詞與有效詞篩選之動作，於此，於使用者點擊「執行資料預處理」後（如圖 A.27 所示），系統即完成修改後回答內容資料之預處理，呈現相關結果予使用者，並同時將修改後討論串「1783227754」之回答資訊，更新並維護於資料庫中（如圖 A.28 所示）；再者，若使用者同時將討論串之提問與回答內容，分別由「小弟想開始...」與「值得買的...」，修改為「目前手上有...」及「買的 nikon...」並點擊「確認修改」後（如圖 A.29 所示），因提問與回答內容將被修改，故系統必須再次執行斷詞與有效詞篩選之動作，於此，於使用者點擊「執行資料預處理」後（如圖 A.30 所示），系統即完成修改後提問與回答內容資料之預處理，呈現相關結果予使用者，並同時將修改後討論串「1783227754」之相關資訊，更新並維護於資料庫中（如圖 A.31 所示）。



圖 A.19、欲修改討論串資料之查詢



圖 A.20、討論串詳細資料之檢視



圖 A.21、討論串資料之修改(1)



圖 A.22、完成討論串資料之修改(1)



圖 A.23、討論串資料之修改(2)



圖 A.24、提問內容資料預處理確認介面



圖 A.25、完成討論串資料之修改(2)



圖 A.26、討論串資料之修改(3)



圖 A.27、回答內容資料預處理確認介面



圖 A.28、完成討論串資料之修改(3)



圖 A.29、討論串資料之修改(4)



圖 A.30、提問與回答內容資料預處理確認介面



圖 A.31、完成討論串資料之修改(4)

#### (A-8)討論串資料刪除功能—程式說明

當權限內使用者進入討論串資料刪除功能時，系統將呈現條件式查詢之介面，以提供使用者進行欲刪除討論串資料之查詢，當使用者查詢得相關討論串資料後，可選擇所需之討論串以進行資料之刪除（如圖 A.32 所示），系統亦將呈現討論串之詳細資料（如圖 A.33 所示），而使用者可針對討論串或回答進行刪除，當中，若使用者刪除討論串時，所有回答將一併刪除，此外，系統將再次與使用者確認其欲刪除之討論串（如圖 A.34 與圖 A.36 所示），當使用者確認並執行刪除功能後，系統即完成討論串之刪除（如圖 A.35 與圖 A.37 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲刪除之領域文件，其編號為「1783227754」時，於點擊「刪除」後（如圖 A.32 所示），系統即以開新視窗之方式，呈現討論串之詳細資訊（如圖 A.33 所示），若使用者欲針對回答標號為「1711034330」之回答進行刪除，於此根據相關之欄位點擊「刪除」，系統將再次與使用者確認欲刪除之回答資料，其編號為「1711034330」，內容為「買的...」，接著，當使用者點擊「確認刪除」後（如圖 A.34 所示），系統即完成回答資料「1711034330」之刪除（如圖 A.35 所示）；若使用者欲針對討論串為資料進行刪除，於此根據相關之欄位點擊「刪除」，系統將再次與使用者確認欲刪除之討論串資料，其編號為「1783227754」，標題為「請教進入...」，接著，當使用者點擊「確認刪除」後（如圖 A.36 所示），系統即完成討論串資料「1783227754」之刪除（包含該筆討論串之全部回答）（如圖 A.37 所示）。



圖 A.32、欲刪除討論串資料之查詢



圖 A.33、檢視詳細討論串資訊



圖 A.34、確認欲刪除之回答資料



圖 A.35、完成討論串回答資料之刪除



圖 A.36、確認欲刪除之討論串資料



圖 A.37、完成討論串資料 (包含該筆討論串所有回答) 之刪除

## (A-9)成員歷史發文資料新增功能—程式說明

當權限內使用者進入成員歷史發文資料新增功能時，系統即呈現成員名稱與歷史發文筆數欄位，以提供使用者輸入相關之基本資料（如圖 A.38 所示），而當使用者送出基本資料後，系統將根據使用者輸入之歷史發文筆數，提供相應數量歷史發文內容之輸入欄位，當中，系統將判定目標成員之資料是否已存在於資料庫中，若目標成員資料已存在，將不重複新增該筆成員之資料（如圖 A.39 所示），反之，則建置新成員資料，並給定其成員編號（如圖 A.40 所示），而當使用者完成所有資料之輸入後，系統將針對目標成員各筆歷史發文內容，執行包含內文斷詞與有效詞篩選之資料預處理動作，於此同時，系統亦將使用者提供未經預處理之原始資料，以及經系統預處理後之結果，儲存並維護於資料庫中（如圖 A.41 所示）。

舉例而言，權限內使用者進入此功能時，系統即呈現成員名稱與歷史發文筆數欄位，供使用者輸入如成員名稱「galant7072」，以及歷史發文筆數「8」等基本資料，此外，系統亦提供討論串所屬類別與上傳時間之下拉式選單，供使用者選擇如類別「電腦/數碼」，以及上傳時間「2013 年 12 月 22 日」，於此，當使用者完成相關資料之輸入並點擊「確認」後（如圖 A.38 所示），系統將比對資料庫中是否已存在該筆成員資料，若目標成員資料（以王小明為例）已存在將不重複新增該筆成員之資料（如圖 A.39 所示），反之，則建置新成員資料（以 galant7072 為例），並給定其成員編號「957828877」（如圖 A.40 所示），接著，系統將根據使用者方才輸入之歷史發文筆數「8」，創建 8 個歷史發文內容輸入之欄位，使用者根據目標成員之領域歷史發文，依序輸入標題「東京景點分享」，以及「這次是第一次...」、「為什麼都顯示網址...」、「感謝樓上大大...」...等歷史發文內容，並於點擊「確認」後（如圖 A.40 所示），系統將針對目標成員「galant7072」之所有領域發文資料，執行斷詞與有效詞篩選等資料預處理動作，以分別取得如「這(DET)次(M) 是(Vt)...」，以及「航空；酷航；空姐...」等資料預處理結果，亦同時將未經預處理前之原始資料，以及經預處理後之結果儲存並維護於資料庫中（如圖 A.41 所示）。





圖 A.41、目標成員各筆歷史發文資料之預處理與新增

### (A-10) 成員歷史發文資料查詢功能—程式說明

當權限內使用者進入成員歷史發文資料查詢功能時，系統即呈現條件式查詢之介面，以提供使用者進行成員歷史發文資料之查詢，而當使用者執行查詢功能後，系統亦將呈現相關成員歷史發文資料查詢之結果予使用者參考（如圖 A.42 所示），此外使用者亦可檢視成員歷史發文之詳細資訊，以及目標成員領域知識可參考程度（如圖 A.43 所示）。

舉例而言，權限內使用者進入此功能時，系統將提供條件式搜尋功能予使用者，以進行資料庫內討論串之搜尋，當使用者未輸入資訊而直接點擊「查詢」按鈕時，系統將呈現目前存在於資料庫中之全部討論串，如成員名稱「galant7072」、領域類別「電腦/數碼」，以及領域歷史發文總數「8」等查詢結果供使用者參考，當使用者點擊「詳細資訊」時，系統將以開新視窗之方式，呈現目標成員歷史發文之詳細資訊，以及目標成員領域知識可參考程度，如成員名稱「galant7072」於「電腦/數碼」領域之領域知識可參考程度為「0.21268」，供使用者參考（如圖 A.42 所示）。



2. 若使用者修改歷史發文之內容，因內文資訊將變更（如圖 A.48 與圖 A.49 所示），故系統將再次執行斷詞與有效詞篩選等資料預處理動作，待完成資料之預處理後，系統即儲存修改後之相關內容（包含資料預處理結果）於資料庫中，以完成歷史發文內容之修改，並呈現修改後之資訊予使用者參考（如圖 A.50 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲修改之成員歷史發文資料，其名為「王小明」且領域類別為「奧運體育」時，於點擊「修改」後（如圖 A.44 所示），系統即以開新視窗之方式，呈現包含「發文編號」、「發文標題」、「發文內容」，以及「領域類別」等詳細資訊，當使用者點擊「修改資料」後（如圖 A.45 所示），系統將呈現修改歷史發文資料之介面，並自動將原始之資訊，包含歷史發文之標題「NBA 是啥」，以及發文內容「請問 NBA 到底...」等資訊填入文字方塊中，於此，使用者修改發文標題為「FBI 是啥」，並點擊領域類別之下拉式選單，於選擇「財經」並點擊「確認修改」後（如圖 A.46 所示），系統即將原先發文標題為「NBA 是啥」，以及領域類別為「奧運體育」之歷史發文，修改標題為「FBI 是啥」與領域類別為「財經」（如圖 A.47 所示）；此外，若使用者將歷史發文之內容，由「請問 NBA 到底...」修改為「請問 FBI 到底...」，並點擊「確認修改」後（如圖 A.48 所示），因發文內容將被修改，故系統必須再次執行斷詞與有效詞篩選之動作，於此，於使用者點擊「執行資料預處理」後（如圖 A.49 所示），系統即完成修改後發文內容資料之預處理，呈現相關結果予使用者，並同時將修改後發文編號「1331998246」等發文資訊，更新並維護於資料庫中（如圖 A.50 所示）。



圖 A.44、欲修改成員歷史發文資料之查詢



圖 A.45、成員歷史發文詳細資料之檢視



圖 A.46、成員歷史發文資料之修改(1)

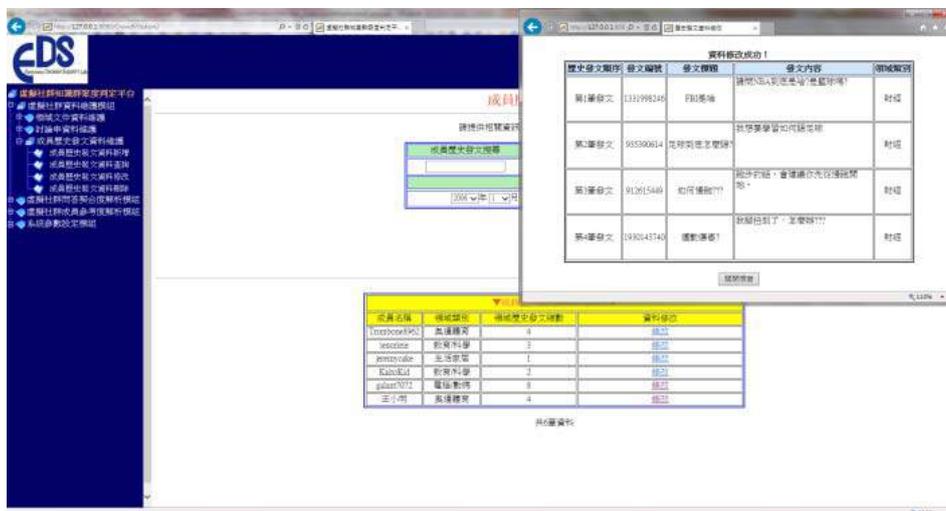


圖 A.47、完成成員歷史發文資料之修改(1)

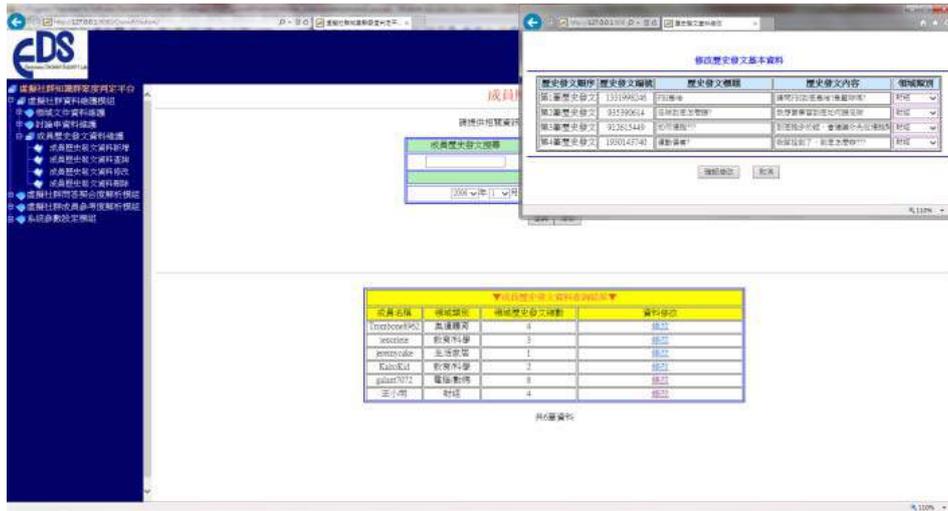


圖 A.48、成員歷史發文資料之修改(2)



圖 A.49、發文內容資料預處理確認介面



圖 A.50、完成成員歷史發文資料之修改(2)

## (A-12)成員歷史發文資料刪除功能—程式說明

當權限內使用者進入成員歷史發文資料刪除功能時，系統將呈現條件式查詢之介面，以提供使用者進行欲刪除成員資料之查詢，當使用者查詢得相關成員資料後，可選擇所需之討論串以進行資料之刪除（如圖 A.51 所示），系統亦將呈現討論串之詳細資料（如圖 A.52 所示），而使用者可針對成員資料或其發佈之歷史發文進行刪除，當中，若使用者刪除成員資料時，目標成員之所有歷史發文將一併刪除，此外，系統將再次與使用者確認其欲刪除之討論串（如圖 A.53 與圖 A.55 所示），當使用者確認並執行刪除功能後，系統即完成成員資料與其歷史發文資料之刪除（如圖 A.54 與圖 A.56 所示）。

舉例而言，當權限內使用者進入此功能，並查詢得欲修改之成員歷史發文資料，其名稱為「王小明」且領域類別為「奧運體育」時，於點擊「刪除」後（如圖 A.51 所示），系統即以開新視窗之方式，呈現成員之詳細資訊（如圖 A.52 所示），若使用者欲針對發文標號為「1930143740」之歷史發文進行刪除，於此根據相關之欄位點擊「刪除」，系統將再次與使用者確認欲刪除之發文資料，其編號為「1930143740」、標題為「運動傷害?」，以及內容為「我腳扭到了...」，接著，當使用者點擊「確認刪除」後（如圖 A.53 所示），系統即完成發文資料「1930143740」之刪除（如圖 A.54 所示）；若使用者欲針對成員資料進行刪除，於此根據相關之欄位點擊「刪除」，系統將再次與使用者確認欲刪除之目標成員，其編號為「1750491756」，名稱為「王小明」，接著，當使用者點擊「確認刪除」後（如圖 A.55 所示），系統即完成成員資料「1750491756」之刪除（包含目標成員所有歷史發文）（如圖 A.56 所示）。



圖 A.51、欲刪除成員資料之查詢

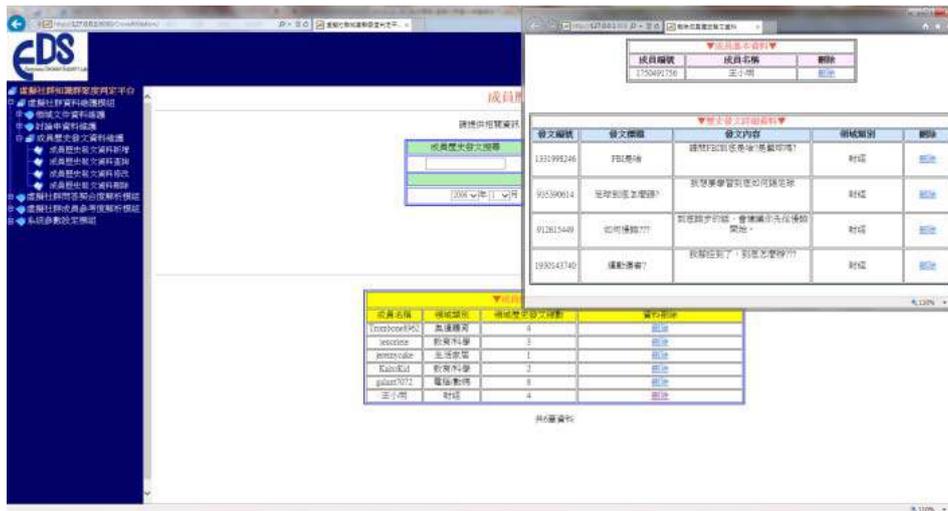


圖 A.52、檢視詳細成員資料資訊



圖 A.53、確認欲刪除之歷史發文資料



圖 A.54、完成發文資料之刪除



圖 A.55、確認欲刪除之成員資料



圖 A.56、完成成員資料（包含目標成員所有歷史發文）之刪除

## (B) 虛擬社群問答契合度解析模組

本研究開發之「虛擬社群問答契合度解析模組」乃結合「領域性關鍵詞彙集合」之概念，並透過對虛擬社群中發問者之「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」等資料之分析，以解析得虛擬社群討論串中各筆回答與該討論串提問之契合程度，本研究將此模組劃分為「(B-1)領域性關鍵詞彙集合建立功能」與「(B-2)問答契合度解析功能」兩大功能。

於「(B-1)領域性關鍵詞彙集合建立功能」中，首先，本功能根據蒐集之領域文件資料，進行內文斷詞及領域有效詞篩選之自動化資料預處理作業，次之，本功能將以領域有效詞之篩選結果為依據，計算各領域有效詞組間之關係距離，接著，即根據上述計算

得之關係距離數據，分別以四分位數、中位數、平均值與使用者自訂等門檻標準，分析並審視領域有效詞組達成之狀況，並於最終將符合門檻之領域有效詞彙組，各自分離成獨立之詞彙，以用於建構領域性關鍵詞彙集合。

於「(B-2)問答契合度解析功能」中，主要係解析已維護於資料庫之討論串，首先，本功能針對討論串進行內文斷詞、有效詞篩選，以及關鍵詞彙篩選（以領域性關鍵詞彙集合為依據）等資料預處理作業，次之，本功能以關鍵詞彙為中心進行前後文語句之分割與斷詞，並計算相關之前後文配對及標題加權整體分數，接著，即依據詞類距離採納值建置前推算集合，分析當中詞類相關性平均分數，並根據設置之比重值計算得關鍵詞彙之代表性得分，且同時以出現頻率為基礎解析得提問與回答之品質，爾後，本功能除分析問答內文之契合度外，亦針對問答間之語意關係距離進行解析，最終，整合於提問品質、問答品質、問答內文契合度，以及問答語意關係距離等數據，本功能即針對目標提問與回答計算得整體問答契合度得分，並於正規化後獲致整體問答之契合程度。

### **(B-1)領域性關鍵詞彙集合建立功能—程式說明**

#### ➤ 程式運作邏輯：

當權限內使用者進入並使用領域性關鍵詞彙集合建立功能時，於開始進行相關運算前，使用者可參考系統呈現之「目前領域文件於資料庫內分布狀況」統計資訊，以選定欲建構之領域性關鍵詞彙集合（如圖 B.1 所示），接著系統即根據使用者選擇之領域類別，進行領域文件斷詞與有效詞篩選之動作，並呈現相關「資料預處理結果」予使用者（如圖 B.2 所示），爾後，系統針對資料預處理結果中有效詞集之部份，執行重複詞彙過濾之動作，對領域文件之有效詞集進行彙整，以取得選定領域中所有文件之有效詞集（如圖 B.3 所示），之後根據領域有效詞集中之詞彙進行「有效詞彙組之劃分」，並提供文字方塊供使用者填入有效詞彙之搜尋結果數（如圖 B.4 所示），接著系統即針對領域中之各有效詞彙組，以主要有效詞、次要有效詞，以及同時包含主要與次要有效詞之搜尋結果數為依據，進行語意關係距離之計算，並以語意關係程度由小至大方式排列（如圖 B.5 與圖 B.6 所示），隨後系統乃根據語意關係距離之數據，進行整體關係距離之計算，並提供使用者門檻篩選方式之選擇，而使用者亦可以複選之方式選擇（如圖 B.7 所示），接著系統即藉使用者選擇之門檻標準，以選定領域中所有詞彙組合之整體關係距離為基礎，進行門檻值之計算，並根據計算得之結果判定詞彙組合門檻之符合狀況（如圖 B.8 與圖 B.9 所示），待完成初步詞彙組門檻之篩選後，系統將根據使用者選定之項

目，進行進一步詞彙組合之分割與篩選，藉以劃分並分析得獨立詞彙之重要性程度（如圖 B.10 所示），於最終使用者選定欲建構之詞彙網路後，系統即建構選定領域及項目之領域性關鍵詞彙集合，並將相關之結果儲存於資料庫中（如圖 B.11 所示）。

► 程式運作流程：

舉例而言，權限內使用者進入此功能時，系統即呈現「目前領域文件於資料庫內分布狀況」統計資訊予使用者參考，而使用者可根據目前資料量較多（領域文件總數為「4」）之「電腦/數碼」領域，進行關鍵詞彙集合之建構，使用者於選定電腦/數碼領域核選鈕後，按下「確認建置」之按鈕後（如圖 B.1 所示），系統將針對選定領域之所有文件斷詞及篩選有效詞，並呈現相關之資料預處理結果，包含「斷詞內容」與「有效詞集」予使用者參考，接著當使用者按下「彙整【電腦/數碼】之領域有效詞集」按鈕後（如圖 B.2 所示），系統即針對領域中所有文件，進行重複詞彙之過濾，以彙整得實具精確之領域有效詞資訊，爾後當使用者點擊「劃分領域有效詞彙」按鈕後（如圖 B.3 所示），系統即針對領域有效詞集中之所有詞彙，以一個主要有效詞如「筆記本」，次要有效詞如「頻」等詞彙為一組，進行有效詞彙組之劃分，並提供搜尋結果數輸入之介面，當相關數據輸入完畢後，點擊「解析語意關係距離」按鈕（如圖 B.4 所示），系統即針對各有效詞彙組，進行語意關係距離之計算，並以小數點後五位進行四捨五入後，呈現相關計算結果之數據如「方法」與「遊戲」之語意關係距離為「1.0」予使用者參考，接著於使用者按下「計算整體關係距離」按鈕後（如圖 B.5 與圖 B.6 所示），將透過系統計算得有效詞彙組如「病毒」與「QQ」之整體關係距離為「0.19999...」，於此亦提供篩選門檻方式之選定，而當使用者以複選方式選定門檻篩選方式「四分位數」與「中位數」，並點擊「進行門檻之篩選與過濾」按鈕後（如圖 B.7 所示），系統即根據使用者選定之門檻項目，計算得相應之門檻值，並呈現門檻符合狀況予使用者，而使用者將可視篩選狀況與己身需求，選定欲進一步分析之項目，於此以「四分位數」與「中位數」為例，當使用者勾選並按下「分割為獨立詞彙，並進行領域詞彙之篩選」後（如圖 B.8 與圖 B.9 所示），系統將對符合門檻之詞彙組合進行獨立詞彙之劃分，同時以獨立詞彙「達成門檻之次數」為基礎，衡量詞彙之重要性、進行篩選，以及呈現相關資訊予使用者，當使用者選定欲建構項目之核選按鈕，並按下「建立領域性關鍵詞彙集合」後（如圖 B.10 所示），於最終系統將儲存相關符合條件之詞彙於資料庫中，以完成電腦/數碼類別領域性關鍵詞彙集合之建構（如圖 B.11 所示）。





圖 B.4、有效詞劃分與搜尋結果輸入畫面



圖 B.5、語意關係距離計算畫面(1)

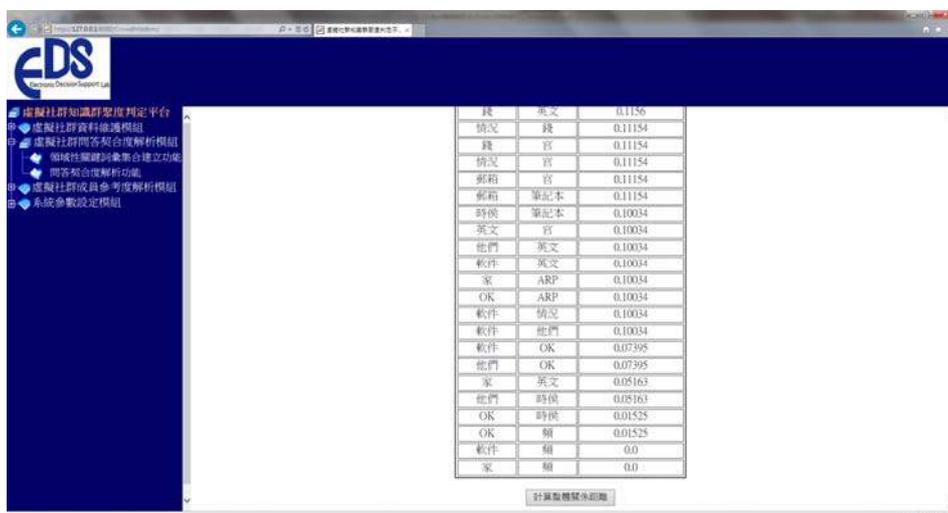


圖 B.6、語意關係距離計算畫面(2)



圖 B.7、整體關係距離計算與門檻選擇畫面



圖 B.8、門檻值計算與詞彙組門檻符合狀況篩選畫面



圖 B.10、獨立詞彙劃分與門檻符合狀況篩選畫面



圖 B.11、領域性詞彙關係網路建立完成畫面

## (B-2)問答契合度解析功能—程式說明

### ➤ 程式運作邏輯：

當權限內使用者進入並使用問答契合度解析功能時，於開始進行相關運算前，使用者可於系統介面中查詢得欲分析之討論串，並於選定討論串後，對該討論串進行「資料預處理」之動作（如圖 B.12 所示），其次，系統將呈現選定討論串中「目標提問」與「各筆回答」之資料預處理結果，使用者可依需求選定回答以進行「語句分割」（如圖 B.13 所示），接著系統即根據目標提問與選定之回答，進行「前文語句」及「後文語句」之分割，而於執行「語句斷詞」功能後（如圖 B.14 所示），系統將針對相關之前後文語句進行斷詞，由語句劃分為獨立詞彙，方以計算「前後文配對及標題加權整體分數」（如圖 B.15 所示），爾後，系統即同時計算關鍵提問，以及回答詞彙之「前後文配對及標題加權整體分數」，並呈現數據予使用者參考（如圖 B.16 所示），於此同時，系統將依使用者輸入之「詞類採納距離採納值」，建置詞類距離採納集合（如圖 B.17 所示），並針對集合中之詞彙進行「詞類相關性平均分數」之計算（如圖 B.18 所示），其次，系統即彙整先前計算得之「前後文配對及標題加權整體分數」與「詞類相關性平均分數」，並以使用者輸入之解析比重值為依據（如圖 B.19 所示），計算得關鍵提問與回答詞彙之「代表性得分」（如圖 B.20 所示），接著，系統將針對提問與回答品質之資訊進行彙整（如圖 B.21 所示），以根據代表性得分及詞彙出現頻率，整合所有關鍵提問及回答詞彙，以分析並計算得提問與回答品質得分（如圖 B.22 所示），此外，系統亦彙整「問答內文契合度之分析資訊」，（如圖 B.23 所示），以運用提問與回答斷詞詞彙發生頻率，分析並計算得「問答內文契合度得分」（如圖 B.26 所示），接著，系統即彙整「問答語意關係距

離分析之資訊」，以取得提問與回答中有效詞彙集合（如圖 B.27 所示），並以排列組合方式進行有效詞組合之劃分（如圖 B.28 所示），以計算針對各問答詞彙組合進行語意關係距離之計算（如圖 B.29 與圖 B.30 所示），當取得所有詞彙組之語意關係距離後，系統隨即以有效詞彙組合總數為基礎，整合並計算得「問答語意關係距離」（如圖 B.31 所示），最終，藉由對「提問品質」、「回答品質」、「問答內文契合度得分」，以及「問答語意關係距離」等數據之整合，系統將計算得「整體問答契合度得分」，並於正規化後獲致「整體問答契合程度」（如圖 B.32 所示）。

➤ 程式運作流程：

舉例而言，權限內使用者進入此功能時，系統將提供條件式搜尋功能予使用者，以進行資料庫內討論串之搜尋，當使用者未輸入資訊而直接點擊「查詢」按鈕時，系統將呈現目前存在於資料庫中之全部討論串，如標題為「電腦藍屏...」與「各路大仙...」等討論串，於此，使用者選擇欲解析之編號為「1034646889」之討論串並點擊「資料預處理」（如圖 B.12 所示），系統即針對使用者選取之討論串，執行資料預處理之功能，並提供如「斷詞」、「有效詞篩選」與「關鍵提問詞彙」等資訊予使用者，而使用者可由目標提問所屬之全部回答中，選定欲進行問答契合度解析之回答，於此勾選編號「1763146692」之回答並點擊「語句分割」（如圖 B.13 所示），系統隨即針對關鍵提問詞彙如「電腦」，以及回答詞彙如「軟件」，進行前文語句與後文語句之分割，並同時呈現相關之結果，於此使用者進一步點擊「語句斷詞」後（如圖 B.14 所示），系統將進行前後文語句斷詞，以劃分得獨立之詞彙，如「各」、「路」與「大仙」...等，並於使用者點擊「計算前後文配對及標題加權整體分數」後（如圖 B.15 所示），計算得關鍵提問詞彙如「電腦」，以及關鍵回答詞彙如「軟件」之前後文配對及標題加權整體分數，分別為「0.2437」與「0.6315」，而當使用者點擊「建置詞類距離採納集合」後（如圖 B.16 所示），系統即以系統管理者已設定之詞類距離採納值「6」為基礎，建置如關鍵提問詞彙「電腦」之前推算集合「的；我；看看；大仙；路；各」，接著當使用者點擊「計算詞類相關性平均分數」後（如圖 B.17 所示），系統將根據關鍵詞彙與其前推算集合，計算得如「0.5」之詞類相關性平均分數，於此，使用者點擊「彙整前後文配對與標題加權整體，以及詞類相關性平均數據」後（如圖 B.18 所示），系統即彙整已計算得之前後文及詞類相關性數據，並呈現予使用者參考，且系統將依系統參數值，分別為前後文「0.6」與詞類相關性「0.4」，進行計算，於此，使用者點擊「取得關鍵提問與回答詞彙代表性

得分」(如圖 B.19 所示)，系統將解析得如關鍵提問詞彙「電腦」之代表性得分為「0.34622」，爾後，當使用者點擊「彙整提問與回答品質分析資訊」後(如圖 B.20 所示)，系統即針對提問與回答品質之分析數據，如代表性得分「0.34622」、於提問中出現頻率「1」，以及關鍵提問總數「2」進行彙整，並於使用者點擊「計算提問與回答品質得分」後(如圖 B.21 所示)，分別以所有關鍵提問及回答詞彙為基礎，分析並計算得問答品質得分，分別為「0.73966」與「1.49634」，接著，當使用者點擊「彙整問答內文契合度分析資訊」後(如圖 B.22 所示)，系統即彙整目標提問與回答之斷詞資訊，並於使用者點擊「彙整斷詞詞彙之詞頻資訊」後(如圖 B.23 所示)，分別針對斷詞後之詞彙於提問及回答語句中統計其出現頻率，並呈現相關如斷詞詞彙「各」，於提問與回答語句中分別出現「1次」及「0次」等結果予使用者，而當使用者點擊「計算問答內文契合度得分」時(如圖 B.24 與圖 B.25 所示)，系統即以彙整得之詞頻為基礎，分析並計算得問答內文契合度得分「0.52」，其次，當使用者點擊「彙整問答語意關係距離分析資訊」後(如圖 B.26 所示)，系統即針對目標提問與回答之有效詞集合進行彙整，並於使用者點擊「劃分提問與回答之有效詞組合」後(如圖 B.27 所示)，呈現獨立詞彙如提問詞彙「大仙」與回答詞彙「網速」，以及詞彙組合如「大仙+網速」等數值輸入區塊予使用者，而當使用者輸入相關之搜尋結果數，並點擊「計算各詞彙組合之語意關係距離」(如圖 B.28 所示)，系統將針對使用者之輸入資訊進行彙整，並同時運用此些數據計算得如提問詞彙「大仙」與「網速」之語意關係距離為「0.18182」等，各提問及回答詞彙組合之語意關係距離，爾後，當使用者點擊「各詞彙組合語意關係距離」(如圖 B.29 與圖 B.30 所示)，系統即根據有效詞彙組合總數「81」，進行相關語意關係距離之彙整及運算，以取得問答語意關係距離「0.20681」，最後，當使用者點擊「計算整體問答契合度得分」後(如圖 B.31 所示)，系統即以提問品質得分「0.73966」、回答品質得分「1.49634」、問答內文契合度得分「0.52」，以及問答語意關係距離「0.20681」為依據，進行相關數據之彙整與整合，以分析並計算得整體問答契合度得分「3.62117」，並同時於正規化後獲致「整體問答契合程度」分數為「0.42771」(如圖 B.32 所示)。



圖 B.12、討論串查詢與分析選擇介面



圖 B.13、目標提問與各筆回答資料之預處理畫面



圖 B.14、關鍵詞彙前後文語句分割畫面



圖 B.15、關鍵詞彙前後文語句斷詞畫面



圖 B.16、前後文配對及標題加權整體分數計算畫面



圖 B.17、詞彙距離採納集合建置畫面



圖 B.18、詞類相關性平均分數計算畫面



圖 B.19、前後文與詞類數據彙整畫面



圖 B.20、關鍵詞彙代表性得分解析畫面



圖 B.21、提問與回答品質分析資訊彙整畫面



圖 B.22、提問與回答品質得分計算畫面



圖 B.23、問答內文契合度分析資訊彙整畫面



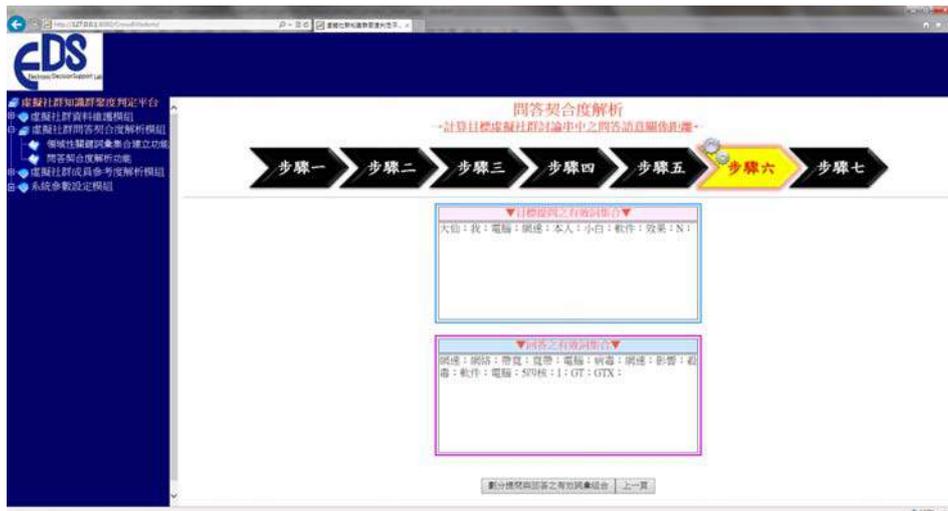


圖 B.27、問答語意關係距離分析資訊彙整畫面



圖 B.28、提問與回答之有效詞組合劃分畫面



圖 B.29、各詞彙組合語意關係距離之計算畫面(1)

詞彙	詞性	詞性	詞性	詞性	詞性
N	900	關係	700	30	0.2374
大仙	100	影響	800	40	0.23107
我	200	影響	800	50	0.23444
電腦	300	影響	800	60	0.11953
關係	400	影響	800	70	0.14898
本人	500	影響	800	80	0.16621
小白	600	影響	800	90	0.17843
軟件	700	影響	800	100	0.18791
效果	800	影響	800	10	0.19566
N	900	影響	800	20	0.20221
大仙	100	殺毒	900	30	0.20788
我	200	殺毒	900	40	0.21289
電腦	300	殺毒	900	50	0.21736
關係	400	殺毒	900	60	0.22141
本人	500	殺毒	900	70	0.22511
小白	600	殺毒	900	80	0.22851
軟件	700	殺毒	900	90	0.11373
效果	800	殺毒	900	100	0.14337
N	900	殺毒	900	10	0.16071

圖 B.30、各詞彙組合語意關係距離之計算畫面(2)

問答契合度解析  
→ 計算目標虛擬社群討論串中之問答語意關係距離 →

步驟一 步驟二 步驟三 步驟四 步驟五 步驟六 步驟七

※數值越趨近於0，詞彙間語意關係越趨緊密

討論串標題	提問內容	回答內容	有效詞彙組合總數	問答語意關係距離
各路大仙，看看我的電腦因為什麼越來越慢，本人小白360清理N次了，軟件卸載一大堆，電腦沒效果	各路大仙，看看我的電腦因為什麼越來越慢，本人小白360清理N次了，軟件卸載一大堆，電腦沒效果	關係和關係帶寬有關，可以考慮升級寬帶，如果電腦中毒了關係也會受影響，用殺毒軟件全盤掃描一次吧。這與電腦配置無關，15/20核電行，GT630有點弱，建議升級GTX660。	81	0.20681

圖 B.31、提問與回答有效詞組合之劃分畫面

問答契合度解析  
→ 計算目標虛擬社群討論串中之整體問答契合度得分 →

步驟一 步驟二 步驟三 步驟四 步驟五 步驟六 步驟七

提問品質得分	回答品質得分	問答內文契合度得分	問答語意關係距離
0.75966	1.49634	0.52	0.20681

↓ 解析整體問答之契合程度

討論串標題	提問內容	回答內容	整體問答契合度得分	整體問答契合度(修正後)
各路大仙，看看我的電腦因為什麼越來越慢，本人小白360清理N次了，軟件卸載一大堆，電腦沒效果	各路大仙，看看我的電腦因為什麼越來越慢，本人小白360清理N次了，軟件卸載一大堆，電腦沒效果	關係和關係帶寬有關，可以考慮升級寬帶，如果電腦中毒了關係也會受影響，用殺毒軟件全盤掃描一次吧。這與電腦配置無關，15/20核電行，GT630有點弱，建議升級GTX660。	3.62117	0.42771

圖 B.32、整體問答契合度得分計算畫面

### (C) 虛擬社群成員參考度解析模組

本研究開發之「虛擬社群成員參考度解析模組」乃以目標虛擬社群成員「歷史發文」資料，以及「歷史發文所屬領域」為基礎，解析目標成員領域發文間群集之相似度，並同時藉其他成員間相似數據之比較，以於最終透過正規化之方式，取得目標虛擬社群成員領域知識之可參考程度，本研究將此模組劃分為「(C-1)成員資料分群功能」與「(C-2)成員參考度解析功能」兩大功能。

#### (C-1) 成員資料分群功能—程式說明

##### ➤ 程式運作邏輯：

當權限內使用者進入並使用成員資料分群功能時，於開始執行分群前，使用者可透過系統查詢得所有社群成員之基本資料，亦可檢視成員詳細之歷史發文（如圖 C.1 所示），其次，系統將針對使用者選擇之社群成員，以其歷史發文中之領域分佈為基礎，統整並呈現相關之歷史發文領域分佈數據（如圖 C.2 所示），接著，系統將以使用者所選領域為基礎，針對當中之所有領域歷史發文執行資料預處理之動作（如圖 C.3 與圖 C.4 所示），並於排除重複詞彙後，以取得目標成員之有效詞集彙整資訊（如圖 C.5 所示），其次，系統將根據最小支持度之設定，於統計各詞彙項目集出現於歷史發文中之數據後，以獲致符合條件之頻繁項目集（如圖 C.6 所示），爾後，系統即以頻繁項目集之交集結果為基礎，於分析各發文間之相似程度後，呈現相似度數據予使用者，並於彙整最大及不為 0 之最小相似度數據後，判定得目前資料所屬之分群類型（如圖 C.7 所示），接著，系統根據初步判定得之分群類型劃分得領域資料之群集，並將最大相似度數據取代為 0，以判定進一步群集劃分之分群類型，於此，若成員之資料未全部完成劃分，系統將重複上述步驟至全部資料皆完成分群為止（如圖 C.8 與圖 C.9 所示），最後，當目標成員之領域資料皆完成分群時，系統即針對相關之分群結果進行彙整，並呈現予使用者參考（如圖 C.10 所示）。

##### ➤ 程式運作流程：

舉例而言，權限內使用者進入此功能，並於選定「發文總篇數由少至多」與點擊「排列」後，系統將呈現所有社群成員之名稱、發文總篇數如「galant7072」與「11」等資訊，使用者亦可點擊「發文內容查詢」之超連結，以開新視窗方式檢視其選定成員詳細之歷史發文資訊，當中，若無發文標題即表示該筆發文屬於「回答內容」（如圖 C.1 所

示)，接著，當使用者按下「顯示成員歷史發文領域分佈狀況」後，系統將根據選定之成員「galant7072」，統計其歷史發文之領域分佈狀況，如該成員於「相機與攝影」領域中發佈了「8」篇文章，接著，當使用者選取「相機與攝影」並點擊「確認並進行資料之預處理」後（如圖 C.2 所示），系統即對目標使用者於「相機與攝影」領域中之所有歷史發文，進行「斷詞」與「有效詞篩選」等資料預處理動作，其次，於使用者點擊「彙整【相機與攝影領域】之有效詞集」後（如圖 C.3 與圖 C.4 所示），系統即於排除重複詞彙後，針對目標成員進行「相機與攝影領域」有效詞集之彙整，並呈現領域歷史發文中全「8」筆之有效詞彙資訊予使用者，而當使用者輸入「25」之最小支持度，並點擊「篩選頻繁項目集」後（如圖 C.5 所示），系統將於統計各項目集於歷史發文中之出現頻率後，根據最小支持度「25%」篩選得如「{淺草寺}」之頻繁項目集，爾後，當使用者點擊「建置相似矩陣並判定分群類型」（如圖 C.6 所示），系統將針對目標成員與領域中之所有發文，以頻繁項目集之交集結果為基礎，分析得各發文間之相似程度，如發文編號「1」與「2」之相似度為 0，並同時彙整得最大相似度為「1」，以及不為 0 之最小相似度為「1」，藉此判定得目前所屬之分群類型（CType）為「2」，接著，當使用者點擊「劃分領域資料之群集」時（如圖 C.7 所示），系統即根據前一步驟中判定得之分群類型「CType=2」，將發文編號「4」與「10」劃分為同一群集，並將相似度數據由原先之「 $\text{Sim}(4,10)=1$ 」取代為「 $\text{Sim}(4,10)=0$ 」，於此同時，系統亦針對經取代相似度數據後之新相似矩陣，彙整得最大相似度為「1」，以及不為 0 之最小相似度為「null（無符合條件之數值）」，並藉此判定得分群類型（CType）為「3」，此外，由於部分成員資料仍未完成分群，故使用者於此點擊「繼續執行分群」後（如圖 C.8 所示），系統將根據前步驟中之分群類型，再次執行分群之動作，以將「1」、「3」、「5」、「6」、「7」與「11」等發文編號劃分為同一群集，於此，由於全部之資料皆已完成群集之劃分，故使用者於點擊「彙整最終分群結果」後（如圖 C.9 所示），於最終，系統將彙整目標社群成員「galant7072」於「相機與攝影領域」中，歷史發文之分群結果，並呈現如發文編號「1」之發文內容、隸屬於「群集 2」，以及目標社群成員所有發文之分群狀況等資訊，以提供使用者參考，此外，當使用者點擊「顯示群集分佈圖」時，系統將以彈跳新視窗方式，呈現使用者領域歷史發文群集分佈結果，如分群結果中包含「Cluster1」與「Cluster2」兩群集，而「Cluster1」中乃包含編號「4」及「10」等歷史發文（如圖 C.10 所示）。



圖 C.1、查詢成員歷史發文資料



圖 C.2、統計成員歷史發文領域分佈



圖 C.3、成員領域歷史發文之資料預處理(1)



圖 C.4、成員領域歷史發文之資料預處理(2)



圖 C.5、彙整領域有效詞與輸入最小支持度

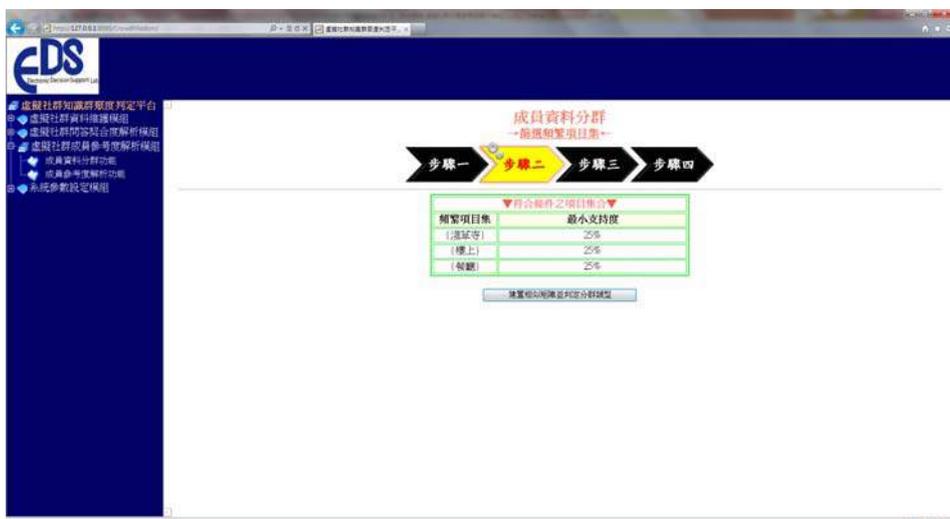


圖 C.6、篩選頻繁項目集



圖 C.7、建置相似矩陣與判定分群類型



圖 C.8、劃分領域資料之群集(1)



圖 C.9、劃分領域資料之群集(2)



圖 C.10、彙整最終分群結果

### (C-2) 成員參考度解析功能—程式說明

#### ➤ 程式運作邏輯：

當權限內使用者進入並使用成員資料分群功能時，於開始執行分群前，使用者可透過系統查詢得所有社群成員之基本資料，亦可檢視成員詳細之歷史發文（如圖 C.11 所示），其次，系統將針對使用者選擇之社群成員，以其歷史發文中之領域分佈為基礎，統整並呈現相關之歷史發文領域分佈數據（如圖 C.12 所示），接著，系統將以使用者所選領域為基礎，以彙整於「(B-1) 成員資料分群功能」中分群所得目標社群成員之領域歷史發文群集（如圖 C.13 所示），爾後，系統即根據頻繁項目集與群集間之關係，定義並獲致各群集之名稱（如圖 C.14 所示），其次，為了於後續分析領域討論串與群集間之關聯程度，以判定成員領域可參考之程度，系統將於判定成員參考度前，先行彙整所有領域討論串，並針對討論串進行斷詞與有效詞篩選等資料預處理動作（如圖 C.15 所示），待完成資料之預處理後，系統即以向量空間模型計算所有群集與討論串間之內文相似度，以及整體內文相似度，以獲致目標成員與領域討論串間之內文契合程度（如圖 C.16 所示），另一方面，系統將以群集為單位進行群集與討論串詞彙之劃分（如圖 C.17 所示），並根據使用者輸入之搜尋結果數，計算得群集與討論串間之語意關係距離（如圖 C.18 所示），最終，系統即根據整體內文相似度與語意關係距離，解析得領域知識可參考性分數，並於正規化後，取得目標社群成員之領域知識可參考程度（如圖 C.19 所示）。

#### ➤ 程式運作流程：

舉例而言，權限內使用者進入此功能，並於選定「發文總篇數由少至多」與點擊「排

列」後，系統將呈現所有社群成員之名稱、發文總篇數如「galant7072」與「11」等資訊，使用者亦可點擊「發文內容查詢」之超連結，以開新視窗方式檢視其選定成員詳細之歷史發文資訊，當中，若無發文標題即表示該筆發文屬於「回答內容」（如圖 C.11 所示），接著，當使用者按下「顯示成員歷史發文領域分佈狀況」後，系統將根據選定之成員「galant7072」，統計其歷史發文之領域分佈狀況，如該成員於「相機與攝影」領域中發佈了「8」篇文章，接著，當使用者選取「相機與攝影」並點擊「彙整分群資訊」後（如圖 C.12 所示），系統將彙整目標社群成員經「(B-1)成員資料分群功能」分群後，所得「相機與攝影」領域之分群結果，並統計得群集內頻繁項目集出現之頻率，如「淺草寺」與「餐廳」皆出現「2」次，而當使用者點擊「定義群集名稱」時（如圖 C.13 所示），系統即針對頻繁項目集與群集間之關係，定義並取得如群集編號「1」之群集名稱為「淺草寺、餐廳」，於此，當使用者點擊「彙整領域知識可參考程度討論串資訊」後（如圖 C.14 所示），系統將根據使用者選定之領域類別，進行目標領域中所有討論串之彙整，並同時執行斷詞與篩選有效詞等資料預處理動作，接著，於使用者點擊「計算討論串與領域歷史發文群集間之內文相似度」後（如圖 C.15 所示），系統即以向量空間模型之方式，計算得如群集編號「1」與討論串編號「1」之內文相似度為「0.81626」，並於整合所有群集與討論串間之內文相似度後，計算得目標成員領域歷史發文及領域討論串間之整體內文相似度為「0.35933」，另一方面，當使用者點擊「劃分有效詞彙組合」後（如圖 C.16 所示），系統將以群集為單位提供有效詞彙組合輸入之介面，於此，當使用者點擊「輸入搜尋結果數」並於填寫完成後，點擊「儲存並維護於系統中」，相關搜尋結果之數據將儲存於系統，待使用者完成全數群集搜尋結果數據之維護後，將於點擊「計算討論串與領域歷史發文群集間之語意關係距離」後（如圖 C.17 所示），系統將計算得「0.28336」與「0.3641」等語意關係距離，以及整體語意關係距離「0.32373」（如圖 C.18 所示），最終，當使用者點擊「計算目標社群成員領域知識可參考程度」後，將透過系統解析得目標社群成員「galant7072」於「運動」領域類別中之領域知識可參考性分數為「1.10997」，以及領域知識可參考程度為「0.21268」（如圖 C.19 所示）。



圖 C.11、查詢成員歷史發文資料



圖 C.12、統計成員歷史發文領域分佈



圖 C.13、彙整成員領域歷史發文之分群資訊



圖 C.14、定義群集名稱



圖 C.15、彙整並預處理領域討論串之資訊



圖 C.16、計算領域討論串與群集間之內文相似度



圖 C.17、劃分群集與討論串之有效詞組合



圖 C.18、計算領域討論串與群集間之語意關係距離



圖 C.19、計算目標社群成員領域知識可參考程度