

南 華 大 學

資訊管理學系

碩士論文

以論壇文章蘊含情緒為基之文章重組模式

A Forum Article Paraphrasing Model based on Article

Implied Emotions

研 究 生：蔡耀昌

指 導 教 授：楊士霆

中華民國 104 年 4 月 16 日

南 華 大 學

資訊管理學系

碩 士 學 位 論 文

以論壇文章蘊含情緒為基之文章重組模式

A Forum Article Paraphrasing Model based on Article

Implied Emotions

研究生：蔡 耀 昌

經考試合格特此證明

口試委員：

余 豐 榮

楊 心 建

邱 宏 林

指導教授：楊 心 建

系主任(所長)：楊 心 建

口試日期：中華民國

104 年 03 月

23 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：蔡耀昌之碩士畢業論文

中文題目：以論壇文章蘊含情緒為基之文章重組模式

英文題目：A Forum Article Paraphrasing Model based on Article Implied Emotions

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：蔡耀昌 (請親自簽名)

指導老師：楊士霆 (請親自簽名)

中 華 民 國 1 0 4 年 4 月 1 6 日

南華大學碩士班研究生
論文指導教授推薦函

資訊管理系碩士班蔡耀昌君所提之論文
以論壇文章蘊含情緒為基之文章重組模式
係由本人指導撰述，同意提付審查。

指導教授



104年4月16日

誌謝

在這兩年又七個月的研究所生涯，最重要也最感謝的就是指導教授 楊士霆 老師，當初若沒有老師的引導，我可能還是個迷途的羔羊而對未來感到焦慮。感謝老師付出相當多的時間以嚴謹的方式指導我，不論在論文撰寫或處理人事務時，總是要求學生要細心、謹慎並替他人著想。感謝老師每天花時間批改學生的進度，在 Meeting 的過程中培養學生具有邏輯性、創新性與反思的思考能力，使我在未來在職場上能更加適應。同時感謝老師在學生犯錯時直接指責使學生能在錯中學習、不斷成長，再遇到瓶頸時總是給予最合適的建議，在未來的出路上引導學生走向最恰當的道路，在此致上由衷的感謝。

學位口試時，非常感謝大葉大學 余豐榮 老師以及南華大學 邱宏彬 老師。謝謝老師們花時間審視學生的論文，並在口試時提出諸多寶貴的建議與指正，使學生的論文更加完善。而在美國交換學生的一年期間，非常感謝我所遇到的人、事、物，感謝與我一同前往美國的小六，一起完成美國求學的所有手續，感謝 Michelle、Jane、Johnson、Jessica、Daniel、Doris、Sherry 與 Zoe 在生活上或出去玩時總是給予我幫助、帶給我無限的歡樂，感謝 Linda、Stephy 與 Angelina 在美國讀書時總是給予我課業上的幫助，最後，感謝南華大學提供交換學生的機會，讓我得以前往國外求學。

接著，非常感謝系上提供豐富的資源，讓我在處理校務、報帳或有校務問題時得以快速解決，非常感謝伊姐在生活上總是給予我寶貴的經驗與建議，非常感謝資訊志工團隊辦活動時讓我學會團體生活、協調與快速解決問題的能力。謝謝研究室 208 與 204 的夥伴們，感謝丸子、小偉、小紅在我的論文遇到瓶頸時一直給予我寶貴的建議，謝謝學長姐們、宗育每天陪我一起熬夜、一起睡沙發、一起寫進度、相互扶持，謝謝俞佑在我回國後一起分擔研究室的事務，感謝研究室 204 的學弟國凱、獻文、資沂，在回國後一起吃飯、寫進度，再次謝謝你們一路的陪伴。

最後，我要感謝在高雄的家人們與朋友，謝謝家人的支持、鼓勵與包容，非常感謝爸爸媽媽全力支持出國交換學生的事宜，讓我在求學、出國期間能心無旁騖的完成碩士學業。真的非常感謝一路陪伴的所有人，同時再次感謝 楊士霆 老師這四年來包含大學專題期間，對我的指導、鼓勵與建議。

2014/04/16

摘要

現今虛擬論壇自由且便利之發言平台，透過發言規範與論壇管理員之審核，論壇使用者即可輕易地發表各項文章，然而，因使用者逐漸增加下，多數虛擬論壇開始設有管理者用以審核違規之文章，但因大量文章注入虛擬論壇中，導致論壇管理者難以逐一審核並回請修正所有之文章，且對於發文者而言，可能於無法意識中寫入之違規字詞，於此，文章撰寫者發文後將隨即觸犯論壇規範，並需再針對違規文章進行修改，使得發文者再發文之意願逐漸低下。

有鑑於上述問題，本研究乃發展「以論壇文章蘊含情感為基之文章重組模式」，並劃分為「文章表達情緒判定模組」與「發文者文章語句重組模組」兩核心模組，以論壇文章為分析之基礎，發展一套適用於論壇之方法論。於前者中乃以文章代表事件分析、情緒詞彙隸屬係數建立、相似語句分析、情緒機率與穩定值解析等技術，以推論文章之情緒類別；於後者中，本研究則以整合語意相似分析、評閱分數分析、候選填充句建立、以及多重組合句建立等技術，於最終將可重組文章之語句架構。

為確認本研究方法論於實務應用之可行性，本研究乃建構一套以網際網路環境為基礎之以論壇文章蘊含情感為基之文章重組系統。此外，為驗證本系統之績效，本研究乃以「奇摩新聞」與比較研究所採用之「痞客邦」，作為「文章表達情緒判定模組」之驗證資料來源，並以「Mobile01 論壇」作為「發文者文章語句重組模組」之驗證資料來源，並於最終績效驗證指標得知，於「文章表達情緒判定模組」中與其他研究比較後三項指標皆高於近 10% 之水準，於系統自我績效評估部分三項指標皆獲致 80% 之水準；於「發文者文章語句重組模組」中，可獲致整體平均召回率 78.51%、正確率 78.58%、F 值為 77.53% 之績效水平。

綜合言之，本研究乃分析可能具有偏激詞語之文章並針對此類型文重組語句，就論壇管理者而言，可針對特定情緒之文章可快速取得違規文章並將重組違規之語句；就文章撰寫者言之，可自動將違規文章之語句進行修正，以節省修改違規文章之時間與人力，並可立即得知文章違規之可能性。因此，期望透過本研究所發展之模式，協助論壇管理者快速過濾違規之文章，並避免文章撰寫者撰寫之文章觸犯論壇規範。

關鍵字：虛擬論壇、語句相似度分析、情緒解析、文章語句重組

ABSTRACT

Most articles are audited by forum administrator before article posting and sharing in order to improve the quality of the virtual forum articles in virtual forum. However, the users in forum gradually increases, the forum administrator difficultly individually reviews illegal articles and requests users modify their articles. For article posters, the illegal words or sentences may be written unconsciously result in violating community standards; therefore, it may take more time to modify for the illegal articles. This paper constructs a Forum Article Paraphrasing Model based on article implied emotions to analyze the article with extreme words and to rewrite and paraphrase its content to promote the its value. Finally, in order to demonstrate applicability of the proposed methodology, a web-based system is also established based on the proposed model. Furthermore, a real-world case is applied to evaluate the proposed model and system.

Keywords: *Virtual Forum, Sentences Similarity, Emotion Determination, Article Paraphrasing*

目錄

誌謝.....	V
摘要.....	VI
ABSTRACT.....	VII
目錄.....	VIII
圖目錄.....	X
表目錄.....	XV
第一章、研究背景.....	1
1.1 研究動機與目的.....	1
1.2 研究步驟.....	7
1.3 研究範圍與限制.....	9
第二章、文獻回顧.....	11
2.1 研究定位.....	11
2.2 虛擬社群之管理機制探討.....	12
2.2.1 影響社群文章共享因素.....	12
2.2.2 社群文章審核機制.....	16
2.3 社群文章資料探勘.....	20
2.3.1 社群文章語句重組技術.....	20
2.3.2 社群文章評級技術.....	24
2.4 文章撰寫者行為探討.....	29
2.4.1 文章撰寫者情感分類.....	29
2.4.2 文章撰寫者寫作習慣分析.....	34
2.4.2 文章閱讀者閱讀感受分析.....	39
2.5 小結.....	44
第三章、以論壇文章蘊含情緒為基之文章重組模式.....	49
3.1 文章表達情緒判定模組.....	49
3.2 發文者文章語句重組模組.....	59
第四章、系統架構.....	69
4.1 以論壇文章蘊含情緒為基之文章重組系統之核心架構.....	69

4.2 系統功能架構.....	71
4.3 資料模式定義.....	74
4.4 系統流程.....	78
4.4.1 系統功能流程.....	78
4.4.2 系統資料流程.....	83
4.5 系統開發工具.....	84
第五章、系統實作與案例分析.....	87
5.1 系統案例之應用流程.....	87
5.2 系統案例驗證與評估.....	102
5.2.1 文章表達情緒判定之驗證.....	108
5.2.2 發文者文章語句重組之驗證.....	116
5.3 小結.....	129
第六章、結論與未來發展.....	130
6.1 論文總結.....	130
6.2 未來展望.....	133
參考文獻.....	134
附錄(1)、系統功能操作說明.....	143
附錄(2)、模組步驟運算與應用說明.....	184

圖目錄

圖 1.1、範例文章.....	2
圖 1.2、文章撰寫者所提之觀點.....	2
圖 1.3、文章所違反之規範.....	3
圖 1.4、MOBILE01 中具有互動之違規文章.....	4
圖 1.5、卡提諾論壇中具有互動之違規文章.....	4
圖 1.6、論壇管理機制審核文章之既有模式.....	5
圖 1.7、論壇管理機制審核文章之期望模式.....	6
圖 1.8、研究架構.....	9
圖 2.1、研究定位圖.....	12
圖 3.1、以論壇文章蘊含情緒為基之文章重組模式之流程架構.....	49
圖 3.2、文章表達情緒判定示意圖.....	51
圖 3.3、發文者文章語句重組示意圖.....	60
圖 4.1、以論壇文章蘊含情緒為基之文章重組系統之流程架構.....	70
圖 4.2、以論壇文章蘊含情緒為基之文章重組系統之功能架構.....	72
圖 4.4、以論壇文章蘊含情緒為基之文章重組系統之資料關聯.....	77
圖 4.5、論壇文章維護模組之功能流程.....	79
圖 4.6、情緒詞彙維護模組之功能流程.....	80
圖 4.7、文章表達情緒判定模組之功能流程.....	81
圖 4.8、發文者文章語句重組模組之功能流程.....	83
圖 4.9、系統資料流程.....	84
圖 5.1、以論壇文章蘊含情感為基之文章重組系統之應用流程.....	88
圖 5.2、MOBILE01 論壇之樣本資料(1).....	89
圖 5.3、MOBILE01 論壇之樣本資料(2).....	89
圖 5.4、MOBILE01 論壇之樣本資料(3).....	89
圖 5.5、門檻值修改介面.....	90
圖 5.6、門檻值修改成功提示.....	90
圖 5.7、論壇文章上傳輸入介面.....	91
圖 5.8、論壇文章上傳成功介面.....	91

圖 5.9、各情緒詞彙之關係係數解析結果.....	92
圖 5.10、各情緒詞彙與類別隸屬係數解析結果.....	93
圖 5.11、候選事件標題加權分數解析介面.....	94
圖 5.12、候選事件前後文配對分數解析介面.....	94
圖 5.13、候選事件詞性相關性分數解析介面.....	94
圖 5.14、文章代表事件解析介面.....	95
圖 5.15、取得文章語句中具有代表事件之語句.....	96
圖 5.16、取得文章代表語句之相似語句.....	96
圖 5.17、取得相似語句之情緒機率值.....	96
圖 5.18、取得文章表達之情緒類別.....	97
圖 5.19、文章初始分數計算介面.....	98
圖 5.20、文章評閱分數計算介面.....	98
圖 5.21、取得文章各語句之關聯程度介面.....	99
圖 5.22、去除語意相似之語句介面.....	99
圖 5.23、取得代表及重要語句之候選填充句介面.....	100
圖 5.24、取得候選填充句之多重組合句.....	100
圖 5.25、取得文章多重語句架構-第 1 組.....	101
圖 5.26、取得文章多重語句架構-第 2 組.....	101
圖 5.27、取得評閱分數最高之文章組合.....	102
圖 5.28、評閱分數最高之文章資訊與內容.....	102
圖 5.29、系統驗證與評估之架構.....	103
圖 5.30、痞客邦部落格之文章資料.....	107
圖 5.31、奇摩新聞之新聞文章.....	107
圖 5.32、MOBILE01 論壇之文章資料.....	108
圖 5.33、本研究與 TUNG 與 LU (2012) 推論數據之比較.....	113
圖 5.34、第一階段情緒類別推論三項指標推論之分佈趨勢.....	114
圖 5.35、系統自我績效評估-各驗證週期推論績效之分佈趨勢.....	115
圖 5.36、發文者文章語句重組推論召回率.....	125
圖 5.37、發文者文章語句重組推論正確率.....	126
圖 5.38、發文者文章語句重組推論 F 值.....	126

圖 5.39、發文者文章語句重組推論-各驗證週期推論績效之分佈趨勢.....	128
圖 A.1、論壇文章資訊上傳.....	144
圖 A.2、論壇文章資訊擷取.....	145
圖 A.3、論壇文章新增成功.....	145
圖 A.4、論壇文章查詢條件.....	146
圖 A.5、論壇文章查詢結果.....	147
圖 A.6、論壇文章詳細資訊.....	147
圖 A.7、論壇文章資料查詢.....	149
圖 A.8、選取欲修改之論壇文章.....	149
圖 A.9、修改目標文章內容.....	150
圖 A.10、論壇文章修改成功介面.....	150
圖 A.11、輸入論壇文章查詢條件.....	151
圖 A.12、論壇文章查詢結果.....	152
圖 A.13、論壇文章確認刪除介面.....	152
圖 A.14、論壇文章資料刪除成功介面.....	153
圖 B.1、情緒詞彙新增介面.....	154
圖 B.2、情緒詞彙新增狀況.....	154
圖 B.3、情緒詞彙查詢介面.....	155
圖 B.4、情緒詞彙查詢結果.....	156
圖 B.5、情緒詞彙之類別隸屬係數結果.....	156
圖 B.6、情緒詞彙查詢介面.....	157
圖 B.7、情緒詞彙修改介面.....	158
圖 B.8、情緒詞彙修改結果介面.....	158
圖 B.9、情緒詞彙查詢介面.....	159
圖 B.10、情緒詞彙查詢結果及詳細資料.....	160
圖 B.11、情緒詞彙刪除結果.....	160
圖 C.1、文章表達情緒判定模組各功能步驟運作圖.....	162
圖 C.2、文章查詢介面.....	164
圖 C.3、文章斷詞與語句分割介面.....	164
圖 C.4、候選事件標題加權分數解析介面.....	165

圖 C.5、候選事件前後文配對分數解析介面	165
圖 C.6、候選事件詞性相關性分數解析介面	166
圖 C.7、文章代表事件解析介面	166
圖 C.8、分析訓練文章所含之情緒語句	167
圖 C.9、各情緒詞彙之關係係數解析結果	168
圖 C.10、各情緒詞彙與類別隸屬係數解析結果	168
圖 C.11、選擇已判定代表事件之文章	170
圖 C.12、取得文章語句中具有代表事件之語句	170
圖 C.13、取得文章代表語句之相似語句	171
圖 C.14、取得相似語句之情緒機率值	171
圖 C.15、取得文章表達之情緒類別	172
圖 D.1、發文者文章語句重組模組各功能步驟運作圖	173
圖 D.2、文章查詢介面	174
圖 D.3、文章初始分數計算介面	175
圖 D.4、文章評閱分數計算介面	175
圖 D.5、取得重要語句之英文詞彙介面	177
圖 D.6、取得重要語句之英文延伸定義詞彙介面	178
圖 D.7、取得文章各語句之關聯程度介面	178
圖 D.8、去除語意相似之語句介面	179
圖 D.9、取得代表及重要語句之候選填充句介面	179
圖 D.10、取得候選填充句之多重組合句	180
圖 D.11、取得文章多重語句架構-第 1 組	181
圖 D.12、取得文章多重語句架構-第 2 組	182
圖 D.13、取得評閱分數最高之文章組合	182
圖 D.14、評閱分數最高之文章資訊與內容	183
圖 E.1、文章表達情緒判定模組之流程步驟	184
圖 E.2、標題「台灣與 FBI 犯罪心理官」之實際文章	185
圖 E.3、標題「唉~無奈」之實際文章	186
圖 E.4、標題「女友吵架...」之實際文章	187
圖 E.5、發文者文章語句重組模組之流程步驟	190

圖 E.6、標題「台灣這些球員...」之實際文章 190
圖 E.7、標題「下雪天開車，可怕的經驗」之實際文章 193



表目錄

表 2.1、影響社群文章共享因素文獻彙整表.....	16
表 2.2、社群文章審核機制文獻彙整表.....	20
表 2.3、社群文章語句重組技術文獻彙整表.....	24
表 2.4、社群文章評級分析文獻彙整表.....	29
表 2.5、文章撰寫者情感分類文獻彙整表.....	34
表 2.6、文章撰寫者寫作習慣分析文獻彙整表.....	39
表 2.7、文章閱讀者閱讀感受分析文獻彙整表.....	43
表 2.8、本研究與過去文獻差異彙整表.....	45
表 3.1、方言情緒字詞之比較.....	50
表 3.2、訓練文章所有候選事件詞性種類之統計.....	55
表 3.3、情緒類別所對應之情緒詞彙.....	56
表 3.4、情緒詞彙與各情緒類別之隸屬係數.....	57
表 3.5、相似語句與情緒類別之機率值.....	58
表 3.6、代表語句與情緒類別之穩定值.....	59
表 3.7、論壇文章架構與評閱分數彙整表.....	67
表 3.8、參考文獻延伸與本研究發展彙整表.....	68
表 5.1、各大部落格之質化比較（數據為 2010 年 7 與 8 月之統計結果）.....	104
表 5.2、各大新聞網站之質化比較（統計數據為 2013 年 5 月）.....	105
表 5.3、各大科技與 3C 產品討論論壇之質化比較.....	106
表 5.4、痞客邦實際測試文章（以 5 份為例）.....	109
表 5.5、奇摩新聞之 20 份實際測試文章.....	110
表 5.6、系統自我績效評估之情緒類別與情緒代號.....	111
表 5.7、系統自我績效評估之第一階段驗證結果（共 200 筆訓練文章）.....	114
表 5.8、系統自我績效評估之推論結果彙整.....	115
表 5.9、三項驗證指標成長率之彙整表.....	116
表 5.10、測試文章編號 1 之偏激語句判斷問卷.....	118
表 5.11、測試文章編號 1 之偏激語句判斷問卷(續).....	118
表 5.12、20 份測試文章之實際內容與偏激語句數.....	119

表 5.13、20 份測試文章之實際內容與偏激語句數(續).....	120
表 5.14、30 位受測者對於 20 份測試文章之偏激語句數調查結果.....	121
表 5.15、30 位受測者對於 20 份測試文章之偏激語句數調查結果(續).....	122
表 5.16、30 位受測者主觀差異性調查結果.....	123
表 5.17、系統自我績效評估之第一階段驗證結果（共 200 筆訓練文章）.....	127
表 5.18、發文者文章語句重組之推論結果彙整.....	127
表 5.19、三項驗證指標成長率之彙整表.....	129
表 E.1、「治安史」事件之論壇文章資訊.....	185
表 E.2、訓練文章資訊.....	188
表 E.3、目標語句與相似語句之詞彙頻率表.....	189
表 E.4、論壇之文章資訊.....	191
表 E.5、訓練文章資訊.....	191
表 E.6、論壇文章資訊.....	193



第一章、研究背景

本章乃說明本研究之研究背景，以下即依序針對「研究動機與目的」、「研究步驟」與「研究定位」等主題進行說明。

1.1 研究動機與目的

過去論壇尚未風行之時，使用者多數僅能與相關熟悉人士討論事物之看法，但現今論壇自由且便利之發言平台，已成為人們發表自身觀點與看法之主要媒介 (Hsu 與 Lin, 2007; Li 等人, 2012)，如「Mobile01」、「巴哈姆特」及「伊利討論區」等，透過發言規範與論壇管理員之審核，即可輕易地發表各項文章 (Guido 與 Leendert, 2006)。然而，因論壇使用者逐漸增加下，文章撰寫者所發表之文章篇幅與內容大多不盡相同，論壇管理者必須依據發言規範與內容之詞語用法審視所有之文章，方可從中過濾違規之文章，以維持論壇之文章品質 (Alavi 與 Leidner, 2001)。

針對上述，雖多數論壇設有管理者用以審核違規之文章，但因大量文章注入論壇中，導致論壇管理者難以逐一審核所有之文章 (Gu 與 Grossman, 2010)。此外，對於特定事物具有特別看法與觀點之文章撰寫者而言，可能因個人疏忽之關係，而於無意識中將帶有偏激或批評之詞語寫入文章內，因而違規而遭致論壇平台或管理者移除，如以圖 1.1 文章為例，文章標題為「HTC Desire HD VS SAMSUNG GALAXY S I9000 選擇性」，文章內容有「此次楊 X 君事件，讓我看到媒體炒作及操控的功力不凡：閃躲、不敢去碰飛彈瞄準對著台灣，...，可以任意捏、踹？這個好用的韓國出氣包，雖然前科不少，但好像也只限於運動賽事...等」，該篇文章撰寫者原先主要乃比較 HTC 與 Samsung 兩手機廠牌之選擇性 (如圖 1.1 色框處)，但因 Samsung 廠商所屬之國家為韓國，該篇文章撰寫者乃將此篇內容引導至韓國與大陸對台灣國際上行為之看法，其中包含「一直想者終極統一的是誰」、「對我們台灣始終充滿著敵意」等數個大陸欲統一台灣之獨特看法，且文章亦有提及「雖然前科不少，但好像也只限於運動賽事」等韓國於國際上對待台灣行為之觀點 (如圖 1.2 色框處所示)，然因文章中具有數個惡意稱呼代替或影射詞以及用字未正名等規定 (如圖 1.3 色框處，包含「大流氓」、「大榴槿」乃影射中國、「軟柿子」影射南韓、「陳 x 扁」未正名等規定)，因而違反論壇之發言規範而遭刪除。



圖 1.1、範例文章

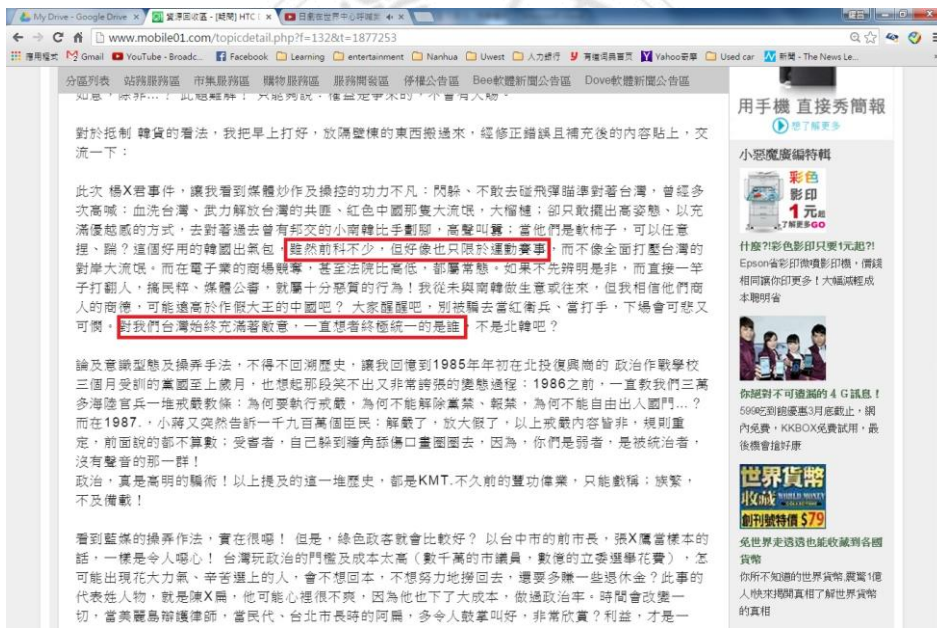


圖 1.2、文章撰寫者所提之觀點

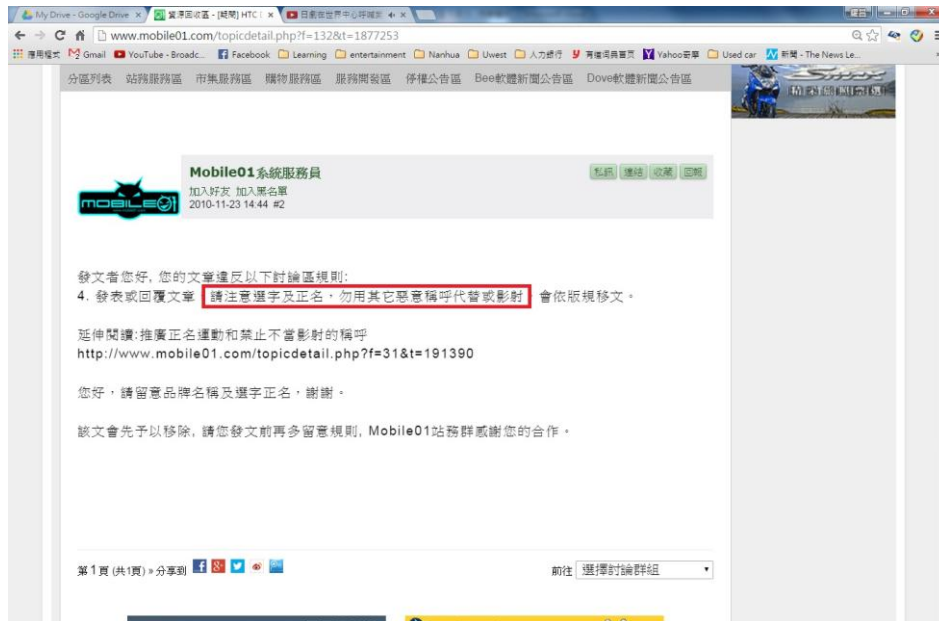


圖 1.3、文章所違反之規範

此外，論壇為能維持社群文章之品質，以現有文字比對機制或委請管理者以嚴謹角度審核違規文章，於此文章撰寫者發文後將較大可能性觸犯論壇規範，亦即具有討論價值與獨特見解之文章將遭致移除。如以 Mobile01、飛格寵物及卡提諾等論壇之資源回收區及待移帖區為例，Mobile01 論壇於圖 1.4 之違規文章列表頁面中違規文章有 30 筆，但於 30 筆違規文章中尚有 15 筆文章具有 2 筆以上之回覆（即圖 1.4 色框處），亦即顯示約有 50% 之違規文章尚有互動且具有討論之價值存在，而卡提諾論壇之違規文章列表頁面中約有 55%，如圖 1.5 色框處所示，且 Mobile01、卡提諾論壇為討論頻繁，流動性快速（Lai 與 Chen，2014；高郁倫，2010）、成員人口數、熱門度皆為台灣熱門網站中排名前 20 之網站（數位時代雜誌，2014），是故，藉由上述兩個台灣代表性論壇可得知文章撰寫者寫入違規字詞時將隨即移至資源回收與待移帖區，且論壇中尚存有高於 50% 比例之違規文章實具實質之互動及討論。



圖 1.4、MOBILE01 中具有互動之違規文章



圖 1.5、卡提諾論壇中具有互動之違規文章

根據上述內容，論壇管理者僅能遵循論壇之規範將所有違規之文章移除，甚者，論壇之文章審核機制可以現有之文字比對方法，將出現違規字詞之文章直接移除 (Ichifuji, 2010)，雖可藉由管理者以自身觀點逐一閱讀所有違規文章，從中獲取具有獨特觀點之文章，並自行或再回請文章撰寫者修正，但此舉將花費大量人力與時間，以致於執行之效率不佳。綜合上述，其既有之運作模式如圖 1.6 之 AS-IS Model 所示。

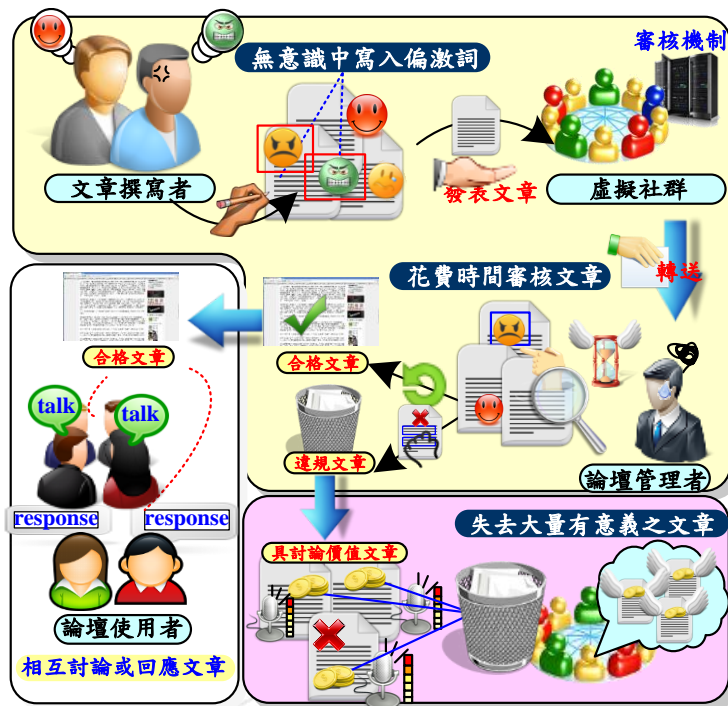


圖 1.6、論壇管理機制審核文章之既有模式

如圖 1.6 所示，目前文章撰寫者可自由地於論壇中發表文章，但論壇對於文章撰寫者發表文章之審核仍有部分缺失，本研究之研究動機與目的可歸納為以下兩點：

1. 文章撰寫者欲表達自身想法時，可能因個人之疏忽而於無意識中寫入不當或抨擊之詞語，因而違反論壇之發言規範。
2. 論壇管理者較難以針對文章內容逐一修改不當之詞語用法，僅能遵循論壇之規範移除違規之文章，以致於論壇失去大量有討論意義之文章。

有鑑於此，為能協助論壇管理者從違規之文章中，獲取具有討論價值之文章，以保留有意義之文章，本研究乃建構「以論壇文章蘊含情緒為基之文章重組」模式，從中分析帶有偏激詞語之文章，並針對此類型之文章重組語句內容，以避免違反論壇之發言規範而遭移除。因此，本研究之期望運作模式如圖 1.7 所示，並將本研究之重點歸納為以下兩點：

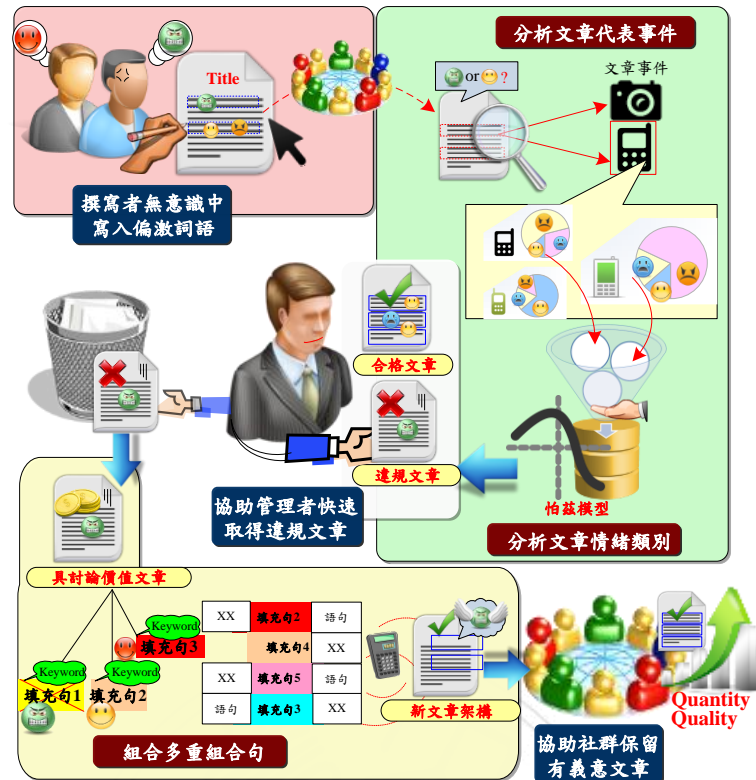


圖 1.7、論壇管理機制審核文章之期望模式

1. 分析文章撰寫者之情緒類別

文章撰寫者對於特定事物欲表達看法時，常會以不同之情緒字詞襯托以加強語句之語氣，但當文章撰寫者用以偏激之情緒字詞時，即容易觸犯論壇之規範，是故，為了取得文章撰寫者之情緒類別，本研究乃藉由文章之代表語句並分析語句之相似程度，再藉由情緒分析技術分析所有相似語句，以取得文章之情緒類別。

2. 重組具偏激情緒與字詞語句之文章

當得知文章撰寫者帶有惡意或偏激之情緒時，本研究針對此類型之文章分析內容之評級分數，對於評級分數較低之文章，即表示內容之順暢性與連結性不佳，因此，本研究乃分析文章內較為重要之語句，透過語句之多重組合句重組文章之內容，以避免文章內之用詞過於偏激而違反論壇規範，進而協助論壇管理者保留有意義且未違規之文章。

整體而言，為了協助論壇管理者保留具有討論意義且未違反論壇規範之文章，本研究乃分析文章之情緒類別，藉此分析文章帶有惡意或偏激情緒之可能性，再分析文章內容之評閱分數，以了解語句連貫程度，並以重要語句之組合句重組文章之內容，因此，本研究之目的即協助論壇管理者快速取得違規之文章，並針對違規但具有獨特看法之文章（可增進論壇使用者持續討論之文章）重組其內容，進而保留有意義之文章。

1.2 研究步驟

本研究之目的乃協助論壇管理者快速取得違規之文章，並針對違規但可增進論壇使用者持續討論之文章重組其內容。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，本研究所涉及主題包含「虛擬群之管理機制探討」、「論壇文章資料探勘」與「文章撰寫者行為探討」等議題，透文獻蒐集與研讀，以了解論壇管理方式、文章語句重組技術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

透過文獻蒐集與探討即可得知，文章過量問題容易導致論壇管理者需花費大量人力與時間審核文章，故需透過系統審核機制過濾文章，但現行對於違規文章之審核多數僅能以文字比對技術，將出現違規字詞之文章移除，以維持論壇整體之品質；透過「文章撰寫者行為探討」之相關研究得知各論壇之文章格式與寫作習慣皆不相同，以致於過去研究大多乃針對特定論壇分析文章之內容以得知文章撰寫者之寫作風格與習慣；藉由「論壇文章資料探勘」之相關研究以了解文章重組之分析方式，並了解相關研究主要乃藉由語意規則及語句相似特徵重組語句之架構，以重組文章整體之架構。由於既有論壇管理機制大多僅能透過管理者或文字比對技術取得違規文章，且違規文章中可能尚存有討論價值之文章，因此，本研究乃發展一套「以論壇文章蘊含情緒為基之文章重組模式」，從中解析帶有偏激情緒與字詞之文章並視為違規文章，以提升論壇過濾文章之效率，同時針對此類型文章重組內容，以避免違反論壇之發言規範，進而協助論壇保留具有討論價值之文章。

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「論壇文章資料解析與探討」、「論壇文章情緒類別與語句重組技術之探討」、「以論壇文章蘊含情緒為基之文章重組方法論發展」與「系

統功能開發」，以下為各項主題之細述說明：

主題一、論壇文章資料解析與探討

- 蒐集並回顧論壇文章內容解析等議題之相關文獻
- 瞭解論壇文章內容解析之相關技術，以建立論壇文章語意法則
- 利用斷詞工具分析文章之斷詞結構與詞性種類
- 利用 WordNet 工具建立重要詞彙之延伸定義詞彙

主題二、論壇文章情緒類別與語句重組技術之探討

- 分析情緒字詞與情緒類別之關聯性，並建立情緒詞彙庫
- 分析文章語句之評級標準
- 建立文章歧義句解析法則
- 建立文章多重組合語句結構之判定法則

主題三、以論壇文章蘊含情緒為基之文章重組方法論發展

- 建立文章表達情緒判定方法論
- 建立發文者文章語句重組方法論

主題四、系統功能開發

- 開發情緒詞與情緒類別關聯自動建立功能
- 開發文章表達情緒分析功能
- 開發文章語句結構重組功能

步驟四、案例驗證

於此步驟乃將本研究所建構之「以論壇文章蘊含情緒為基之文章重組模式」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說明可知，本論文首先乃根據研究動機與目的蒐集論壇管理方

式、論壇文章解析方式、文章撰寫者撰寫習慣等相關文獻，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，並以此為基礎開發以論壇文章蘊含情緒為基之文章重組系統。最後，再以案例驗證本研究所發展之方法論與系統模組，確認本研究之實用價值與發展性。本論文之研究架構如圖 1.8 所示。



圖 1.8、研究架構

1.3 研究範圍與限制

本研究主要乃探討論壇管理者需花費大量時間過濾違規文章、文章撰寫者常無意識中將帶有偏激之詞語寫入文章等問題，當中所分析之對象為中文論壇文章，然而，文章可以多樣化方式發表於論壇中，而本研究乃針對特定類型之文章進行探討，並將研究之範圍與限制訂定如下：

- 台灣各論壇所含違規文章內容大多為「涉及人身攻擊、商業廣告、蓄意攻訐或公開他人隱私之文章..等」，而本研究所探討之違規文章乃針對「涉及人身攻擊」之文章類型進行探討。
- 目前可取得之情緒詞彙庫包含臺灣地區華人情緒與相關心理生理資料庫、華語文網路言論負向情緒用詞檢核軟體研發、知網 (Hownet)，本研究乃選定「分析來源為網路文章」且「具有一定之完整性與公信力」之資料庫「知網」(2090 個情緒詞彙)，作為本研究之情緒詞彙庫。

- 本研究主要分析之語言為正體(繁體)中文，尚無法針對他國語言之文章進行分析，但若基礎文章庫之資料為簡體中文，亦可適用於簡體中文之論壇。
- 本研究尚無法針對文章內容包含、「圖片」、「影片」或其他多媒體方式呈現進行全面性解析，本研究僅針對文章之「文字」資料進行分析。



第二章、文獻回顧

本研究之目的乃協助論壇管理者快速取得違規之文章，並針對違規但可增進社群使用者持續討論之文章重組語句內容，以協助社群保留有意義之文章，並節省論壇管理者過濾違規文章之時間。因此，於探討相關文獻前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

本研究所涉及之研究主題乃包括「虛擬社群之管理機制探討」、「社群文章資料探勘」與「文章撰寫者行為探討」三大研究方向，以下即針對此三大主題之相關研究進行文獻回顧及探討。

於虛擬社群之管理機制探討之議題中，根據過去研究可分為「影響社群文章共享因素」及「社群文章審核機制」兩方面進行探討，於影響社群文章共享因素中，相關研究大多乃探討系統層面與社群提供之服務對於社群成員知識共享意願之影響；對於社群文章審核機制議題而言，可分為「以管理者進行審核」及「以審核系統進行審核」兩子議題進行探討，以瞭解相關研究利用管理者及系統審核與管理社群文章之方法。

於社群文章資料探勘議題中，過去相關研究可歸納為「社群文章語句重組技術」、「社群文章評級分析」等兩方面進行探討。於社群文章語句重組技術議題中，大多研究利用語句之語意關係及相似程度，重組文章語句之架構以取得文章之重點；而社群文章分析及分析課題中，過去相關研究之成果大多利用前後文關係、相異詞語數及語言文法特性，分析文章語句之連貫程度，以藉此取得評級分數。

最後，於文章撰寫者行為探討之議題中，過去研究可分為「文章撰寫者情感分類」、「文章撰寫者寫作習慣分析」以及「文章閱讀者閱讀感受分析」等三方面探討相關文獻。於文章撰寫者情感分類議題中，過去研究大多以資料探勘、詞語相似度及語意關係分析文章之內容，藉此取得文章撰寫者欲表達之情感傾向；於文章撰寫者寫作習慣分析課題中，其相關研究大多乃針對文章撰寫者寫作習慣或寫作風格進行探討；而於文章閱讀者閱讀感受分析課題中，相關研究之成果可分為文章篇幅及圖表對於閱讀者閱讀之感受兩方向進行相關文獻之探討。

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，

藉由過去相關之研究成果，本研究乃依不同研究議題搜尋相關文獻資料，並針對不同主題進行細節說明。本章文獻回顧即針對「虛擬社群之管理機制探討」、「社群文章資料探勘」與「文章撰寫者行為探討」三大議題進行說明。

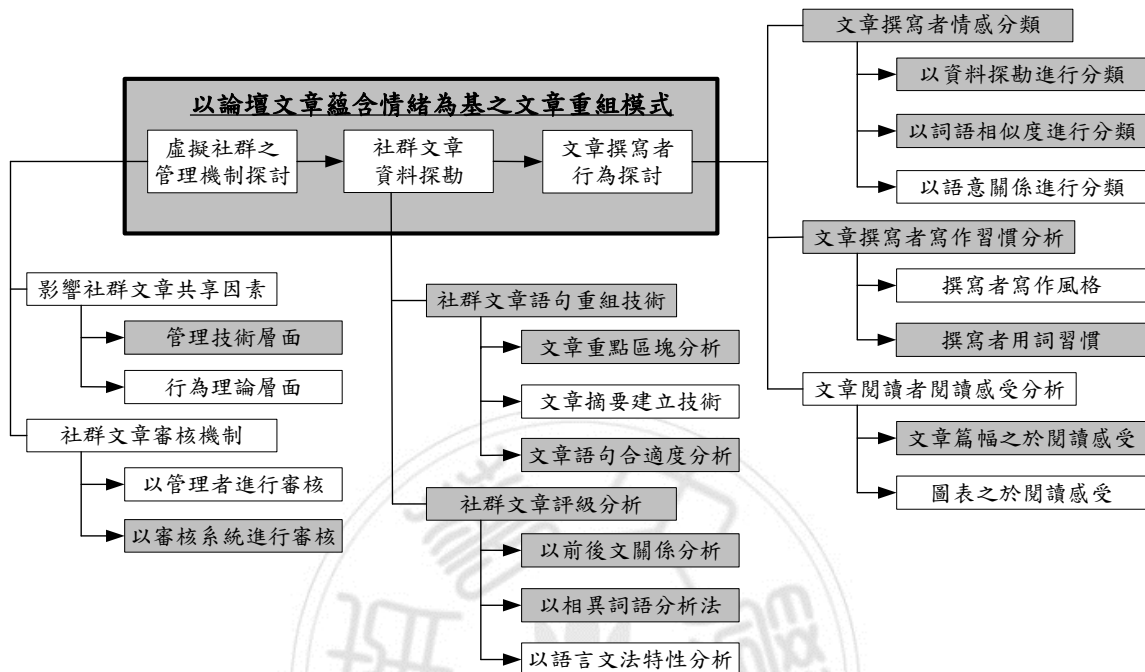


圖 2.1、研究定位圖

2.2 虛擬社群之管理機制探討

對於社群之管理機制探討議題而言，本研究乃針對「影響社群文章共享因素」及「社群文章審核機制」進行相關文獻探討，期望於其中觀察此議題應用於不同類型之不同角度與層面解析，以更深層探討社群之管理機制之特性。

2.2.1 影響社群文章共享因素

於影響社群文章共享因素中，本研究乃針對「管理技術層面」及「行為理論層面」等兩主題進行文獻探討，期望從中探討影響社群文章共享因素所涉及之範圍與領域。

(A) 管理技術層面

針對管理技術層面部分，使用者於虛擬社群中搜尋知識或服務時，其搜尋字串常與專業詞彙之語意無關聯，導致搜尋結果不彰，是故，Peng 等人 (2008) 乃提出一個以相似本體概念為基礎之語句解析方法，以分析搜尋語句與領域知識之關聯性。該研究乃結

合 Web 服務之本體語言 (Ontology Web Language for Services ; OWL-S) 與統一描述、發現語與集成 (Universal Description, Discovery, and Integration ; UDDI) 之概念建立一個匹配分析法，首先將單詞 (即包含詞彙、片語等) 於各領域中定義單詞間之匹配路徑 (即兩單詞之關聯性)，並依據匹配程度標記單詞之候選取代、候選不相關及候選反義三個路徑，以了解單詞之同義與反義詞，進而建立一個特定領域之本體樹，最後，利用匹配演算法分析三個候選路徑，以精確取得單詞間之 同義匹配值、可取代匹配值及不匹配值，以幫助使用者串聯搜尋語句與各領域專業詞彙之關聯性；而 Li 等人 (2006) 透過 WordNet 詞彙庫中詞彙之分層架構，依據詞彙之層級、延伸詞之距離向量分析詞彙間之相似權重值，再藉由詞彙於訓練語句內之位置權重，以取得詞彙於前後文之語意關係，進而建構一個相似語意詞彙庫，其次，該研究利用目標語句之詞彙向量分析所有詞彙之語意關係，同時根據詞彙於語句之位置分析語意權重值，以將兩相似語意權重值之語句視為相似語句，以幫助使用者以簡短之語句，取得不同語法架構但語意相似之語句，進而藉此應用於搜尋引擎中延伸搜尋範圍及條件。

此外，過去知識擷取系統大多需透過人工標註與訓練，以確認知識於特定領域之適用性，然而，當知識轉移至其他領域或知識之解釋有變化時，其系統將可能擷取錯誤之知識，因此，Jung 等人 (2005) 發展一套 POSIE (Postech Information Extraction System) 資訊擷取系統，該研究乃擴充 Eikvil (1999) 中資訊擷取系統之功能。並以一個知識問與答之架構，將知識擷取者所提出之問題轉為結構文件以進行分類，並利用 Lexico-semantic 模式處理已分類文件中文字語意關係及語意特徵模糊之問題，以串聯問題與知識之語意關係，進而提供使用者對應之知識，此外，該系統亦可協助使用者得知該問題領域之專家，進而聯繫與知識探討，並回饋知識擷取者正確之解答，亦可藉由討論與聯繫之行為，提升系統之使用率與熱絡度。甚者，隨著網路之發展使得虛擬社群知識庫之知識量逐漸增加，但大部分之知識乃無經鑑定與審核，導致知識庫之知識容易出現不真實、矛盾之情況，因此，為了解決上述之問題，Kaila 等人 (2006) 提供含三項功能之知識分享平台以過濾知識庫之知識，其中包含：(1)該平台可自動擷取外部知識庫之知識，將其轉化為內部知識庫真正所需之知識、(2)並利用內外知識語意之關係設計一個邏輯平台，將被錯誤知識所誤導之正確知識尋回、以及(3)提供知識擷取者正確之知識，亦可確保正確之知識不被淘汰，是故，藉由研究所提之方法，可幫助管理者辨識知識之正確性，以維持虛擬社群之知識品質。

另一方面，知識共享之行為乃虛擬社群成功發展關鍵之一，為了提升使用者共享知

識之意願，Wang 等人 (2008) 乃以 Gnutella 點對點網路為架構設計一套知識分享系統，該系統透過相同知識分享者與相似上下文訊息，將相似之知識彙集於同一知識庫，再將知識庫以不同格式及功能級別加以分類，在依據同級別與格式之知識庫中所有之知識分享者分於一類，以利用即時通訊工具協助知識擷取者與分享者直接溝通，是故，藉由該系統之運作，將可協助知識擷取者取得所需之知識，並可同時藉由即時通訊功能與該知識之分享者進行直接溝通，亦即將同一領域之知識分享者歸於一類以促進知識之共享。

(B) 行為理論層面

針對行為理論層面之課題，虛擬社群之存在價值在於成員之知識共享，因此，為了瞭解社群成員共享知識意願之因素，Zhang (2009) 乃將過去影響知識共享因素之相關研究進行彙整，以協助學者與社群管理者了解影響社群共享知識之因素。該研究乃將知識分享因素之研究分為三個研究導向，首先為管理技術類別，此類別主要強調社群管理制度將影響成員共享知識之意願，其影響因素主要有社群完整性及訊息品質等，接著為行為理論類別，此類別學者認為社群成員於共享知識前，會自我評估共享後之結果與意義，因此，影響因素主要有自我效能、知覺感受、社會認同、等會間接影響成員心理之過程，並影響成員共享知識之意願，最後為激勵理論類別，該類別認為知識共享之需要、動機、目標與行為間之關係，將會影響成員知識共享之意願，主要影響因子有激勵制度、自我肯定等。透過該研究之彙整，將可幫助社群管理者了解建設或經營虛擬社群時，必須先分析社群之功能完整性及激勵制度與否，且對於社群使用者而言，須分析使用者參與社群活動之過程並解析使用者之感受，將可針對使用者之需求改進虛擬社群之管理制度。甚者，Fang 等人 (2010) 乃透過 143 位 IT 論壇之成員進行問卷調查，並於問卷中針對社群回饋成員之公平性（如激勵機制）、成員共享知識之方式（如知識分享平台之運作方式）及成員間、成員對社群之信任、知識可信度之分析，以瞭解成員使用社群之意圖及回饋機制對成員持續使用社群之成效，其分析結果可得知社群之回饋機制與社群成員對論壇之信任兩因素，主要乃影響成員持續與參與共享知識之因素。

而使用者欲於論壇中找尋所需之熱門主題時，須花費大量時間搜尋，此外，論壇文章所表達之情緒常反映使用者對於特定對象之評價，因此，若能即時於論壇之熱門主題中分析文章之情緒類別，將可提供決策者改變產品或服務之策略，有鑑於此，Li 及 Wu (2010) 以新浪(Sina)論壇作為分析對象，先以人工標記主題之名稱、文章及網頁 URL，以形成主題之樹狀分類結構圖並作為訓練資料，之後將各主題之文章進行分詞處理，以

利用文章之關鍵詞與知網 (HowNet) 所建構之情緒詞進行字詞比對，以取得文章關鍵詞之情緒值 (正、負向情緒)，再將所有關鍵詞之情緒值予以加總藉以得知文章之情緒類別，其次，透過論壇特徵、主題數及子主題等，以 K-means 群集演算法 (K-means Clustering) 依序分為 31 個熱絡程度值，並利用主題之回應數、使用時間及存在數等數據，以支援向量機 (Support Vector Machine; SVM) 將主題分類於 31 個熱絡程度中，進而得知最熱門之主題。藉由該研究之方法，使用者可得知交流頻繁且符合所需之主題，亦可藉由文章表達之情緒了解產品或服務之評價，以適時提出應對策略。

除此之外，Teo 等人 (2003) 認為使用者對知識之可擷取性，以及使用者對虛擬社群之適應性乃是影響虛擬學習社群 (Virtual Learning Communities; VLC) 可持續發展與經營之關鍵，因此，該研究利用科技接受模型 (Technology Acceptance Model; TAM) 之三項結構因素 (知覺有用性、知覺易用性及使用意圖等)，並加入社群適應性影響知覺易用性之構面、無障礙信息 (如利用語音功能傳達訊息) 影響知覺有用性之構面、歸屬感影響使用意圖之構面，該研究針對 84 位使用者 (皆為學生) 進行兩階段之測驗並同時接受問卷調查，接著透過偏最小平方法 (Partial Least Squares; PLS) 分析得知，使用者對虛擬社群之適應性愈高，相對提高使用者對虛擬社群之歸屬感；而當虛擬社群之知識訊息完善時，可使使用者隨時擷取所需之知識，進而提升使用者對虛擬社群之知覺有用性及易用性。透過該研究之方法可得知，社群所提供之完整訊息與使用者之適應程度乃影響社群持續經營之關鍵。最後，過去研究對於知識分享行為之探討議題主要可分為「定性」與「定量」，然而，甚少研究整合定性與定量之相關研究，並同時探討知識分享之個體層面，因此，Wang 與 Noe (2010) 整合知識分享相關之文獻，統整並建立一個知識分享模式，該模式將知識分享分為五個主要重點領域，分別為組織環境、人際關係與團隊特徵、文化特徵、個人特徵以及激勵因素。該研究結果顯示，知識分享之數量及質量與組織行為及環境因素具有極高之相連性，而知識分享之品質將對組織中之知識管理成功與否有顯著之影響，因此，藉由該研究之整合探討，將可協助使用者了解行為理論與系統層面對於知識管理之影響，亦可幫助管理者針對使用者給予合適之服務。

綜合上述，針對影響社群文章共享因素課題皆有許多文獻研究相繼而出，並得知無論根據社群之管理技術或使用者行為層面，皆可透過系統改善或提升服務以增進成員共享知識之意願，本研究即針對社群文章之品質、數量及功能完整性部分加以改善，期望保留違規但具有討論價值文章之數量，以供成員搜尋所需之知識更為廣泛，並藉此提升成員共享知識之意願，於此議題中各相關文獻影響因素及分析過程彙整於表 2.1。

表 2.1、影響社群文章共享因素文獻彙整表

文獻議題	過去研究	使用之技術或方法	探討與分析對象	影響因素
管理技術 層面	Peng 等人 (2008)	結合 Web 服務之本體語言與統一描述、發現語與集成之概念建立語句匹配分析法。	使用者搜尋語句	搜尋成效不彰
	Li 等人 (2006)	依據詞彙之層級、延伸詞之距離向量分析詞彙前後文之語意關係	使用者查詢語句	搜尋成效不彰
	Jung 等人 (2005)	利用 Lexico-semantic 處理文字語意特徵，以串聯問題與知識之關係	知識文件內容	系統擷取錯誤之知識
	Kaila 等人 (2006)	提供含三項功能之知識分享平台以過濾知識庫之知識	知識文章內容	知識容易出現不真實、矛盾之情況
	Wang 等人 (2008)	以 Gnutella 點對點網路為架構設計一套知識分享系統	知識文章內容	同領域之使用者無法相互討論
行為理論 層面	Zhang (2009)	將過去影響知識共享因素之相關研究進行彙整	知識分享 相關文獻	虛擬社群之功能完整性及激勵制度
	Fang 等人 (2010)	透過 143 位 IT 論壇之成員進行問卷調查	論壇成員	回饋機制與使用者對社群之信任
	Li 及 Wu (2010)	以支援向量機將主題進行分類，以得知最熱門之主題	社群文章	搜尋熱門主題之時間
	Teo 等人 (2003)	針對 84 位使用者進行問卷調查，以分析六大構面之影響	一般大學生	知識之可擷取性及使用者對虛擬社群之適應性
	Wang 與 Noe (2010)	整合知識分享相關之文獻，統整並建立一個知識分享模式	知識分享 相關文獻	知識之數量及質量、組織行為及環境因素

2.2.2 社群文章審核機制

於社群文章審核機制中，過去研究乃針對「以管理者進行審核」、「以審核系統進行審核」等兩部分中，探討文章審核機制所涉及之範圍與領域。

(A) 以管理者進行審核

針對以管理者進行審核之課題，目前虛擬社群已逐漸有知識過量之問題，導致社群管理者難以監管與分類知識，是故，Gu 與 Grossman (2010) 針對知識交流量較高之區域網路，架設一個知識分享平台，以協助管理者管理過量之知識，亦可協助知識擷取者於過量之知識中取得正確之知識量，該平台採用 P2P 模式使知識擷取及分享者能隨時共享知識，並利用分佈式知識分析法將知識進行分類，管理者亦可針對高知識量（即較使用者分享或討論）之領域，額外開設一個知識分享平台，以便於同領域之知識擷取及分享者更集中地共享，是故，透過該平台可協助管理者於大量之知識流中彙整與過濾知識之專有領域，並可方便使用者擷取及分享所需知識。此外，Matsubara 等人 (2006) 提

出一套知識分享系統，該系統採用點對點技術模式將知識分享者所共享之知識檔案彙集於中央管理器之虛擬目錄中並加以分類，再由中央管理器控制虛擬目錄中知識檔案之安全性（如非正版之音樂檔案）及內容，藉由個人知識之數據流量、副檔名、內部編碼據分析知識檔案之安全性，以方便管理者有效管理知識檔案安全性及過量之問題。透過該研究之方法，增設虛擬目錄並適當地控制其虛擬目錄之知識，可有效降低擷取知識檔案時所隱藏之危險性（盜版之檔案），亦可方便管理者審核知識安全性之問題。

另一方面，過去關鍵字擷取技術中，多數須以訓練資料之方式建構關鍵字庫，且較少研究以詞彙鏈之方式提取文章關鍵詞，故 **Ercan 及 Cicekli (2007)** 建構一個以基線系統 (Baseline System) 為基之關鍵詞擷取模式，首先藉由語法資料庫 (WordNet) 所提供之多層次語意關係 (即詞彙之多種解釋)，將文章中所有名詞 (不包含複合名詞) 之上下義詞進行串聯，以建構名詞語意關係之詞彙鏈，再依據文章基線圖中首次與最後出現之詞彙特徵，與詞彙鏈進行相關性比對以取得各詞彙鏈中得分最高之詞彙，以自動推論文章之核心關鍵詞。該研究以基線圖為基並結合詞彙鏈之關鍵字擷取方法，可於無監督學習之方式下自動取得文章之關鍵詞，亦即無須以人工方式建立關鍵字語意之關係及關鍵字庫，即可自動比對關鍵字語意關係以取得文章關鍵詞。甚之，使用者於論壇中常需花費大量時間尋找知識，且當取得與需求相關之知識時，其延伸之話題常與原主題不符 (即其他使用者對於主題之回覆訊息)，導致使用者難以從中取得所需之知識，有鑑於上述問題，**Li 等人 (2012)** 提出一個基於馬爾可夫鏈之社交網路 (Social Network-Based Markov Chain; SNMC) 系統，以推薦合適之知識專家並透過交流取得相關之延伸話題。該系統首先去除重複之問題 (即語意相同但語法不同之問題)，並根據使用者欲尋求之知識語法以自然語言方法自動搜尋相關之主題，從中取得所有相關主題之回答者，最後，該系統根據搜尋路徑分析符合使用者需求知識之專家，同時分析所有知識專家之社交親密度、知名度，以利用上述分析數據從中獲取最合適使用者之知識專家，以提供使用者一個交流之管道，是故，藉由該研究所提之系統，可協助使用者容易取得符合需求之知識，且透過使用者與知識專家之交流，將可協助論壇提升知識共享之意願。

(B) 以審核系統進行審核

於以審核系統進行審核之課題中，甚者，為了解決問答系統推論答案時，常因詞彙具有同義或多義之情況而導致擷取錯誤答案或忽略正確答案，因此，**Yu 等人 (2006)** 提出一個以潛在語意分析 (Latent Semantic Analysis; LSA) 為基之問答相似句演算法，

該研究先將問題內容去除停用詞（即會影響關鍵詞顯現度之詞彙），以擷取名詞、副詞及形容詞之關鍵詞，並將此些關鍵詞透過 Google 搜尋引擎（Google Search Engine）查詢前數個搜尋結果作為摘要，並將摘要依長度、標點符號等規則劃分成語句，以形成候選答案語句，接著利用潛在語意分析法分析所有候選語句與問題之相關性，同時將語句中名詞、時間、人名等九種類型之詞彙納入相關性權重之考量，以更精確取得與問題最相關之語句，最後即於候選答案中取前五個與問題最相關之語句作為答案以供使用者查詢。是故，該研究先行藉由去除停用詞、擷取重要關鍵詞及劃分候選語句等前置作業，再利用潛在語意分析法及過濾九種類型詞彙之作業，即可有效處理語句同義或多義詞彙之問題以回饋更精確之答案供使用者參考。

除此之外，「屬性詞」乃近代所衍伸之新興詞類，大多可用以描述事件之屬性或層級，甚者，可修飾動詞之表現方式與性質，因此，網路評論中所存在之屬性詞可視為評論者對於產品之觀點，為了從特定領域之網路評論中取得屬性詞，Li 等人 (2011) 乃以手機評論作為訓練資料，利用分詞系統針對評論進行分詞處理，並藉由 N 連詞 (N-Gram) 以詞彙、詞性與詞義為單位，再將語句加以分詞以形成數個組合語句，其次，該研究利用詞性序列模板，將評論內文之名詞、動詞及動名詞短句進行序列統計，以高出現頻率之詞句作為候選屬性詞，最後，該研究乃以兩步驟過濾不屬於屬性詞之詞句，以精確取得該領域之屬性詞，包含：(1) 先行定義該領域之六個屬性特徵（手機、電池等），再以 Google 搜尋引擎查詢所有屬性詞後，將屬性詞與屬性特徵之匹配出現頁數，透過點式交互訊息法 (Point-Wise Mutual Information; PMI) 量化方法計算屬性機率值，以過濾機率值較低之詞句、(2) 利用知網之詞彙庫，將候選屬性詞中屬於情感詞、單字動詞與特殊動詞之詞句去除，以保留真正用以描述產品觀點之屬性詞。此外，由於網路文件格式及語句結構多數不相同，使用者常須以人工方式閱讀文件後，才可取得語法結構不同但相似之文件（相似之事件或主題），以取得所需之知識，有鑑於上述問題，Jonathan 及 Yiu (2006) 提出一個模糊集檢索 (Fuzzy Set Information Retrieval) 方法，利用維基百科數據庫 (Wikipedia Database) 將單一文件作為一組詞句集合，並針對所有文件去除無意義之標點符號、介詞、連接詞，以取得與文件主題較為相關之詞句組合，接著，該研究利用三種聚類分析方法分析所有文件中詞彙、語句之相關性，首先，透過對比聚類 (Correlation Cluster) 分析法統計兩詞彙之交錯頻率（如單一文件中詞彙「貓」出現次數，而詞彙「狗」未出現），藉由交錯後之匹配結果分析兩詞彙之相關程度，其次，利用關聯聚類 (Association Cluster) 分析法將兩詞彙之共生頻率加以統計，以取得兩詞彙

之相關性分數，最後，利用公制群集（Metric Cluster）分析法將兩詞彙向量化，以解析兩詞彙之向量距離，進而取得兩詞彙之相似值。該研究同時比較三種分析法之搜尋成效，其中以公制群集分析法成效最佳，亦即使用者透過模糊集之處理，再結合公制群集分析法即可有效取得所有詞彙之相關性，進而比對所有詞彙後以取得相似之文件。

在最終方面，知識管理乃一個能使知識更有效地被捕捉、儲存、使用與傳播之技術，但知識之變化、更新快速，甚者，陳述性知識與結構型知識等類型之不同，以致於知識管理者難以控管不同型態之知識，是故，為了能更進一步地提升知識管理之效率與效能，Lai (2007) 提出一套以知識管理為主之知識工程理論模式（Knowledge Management through Knowledge Engineering; KMKE），該模式透過四種階層機制，分別為知識建模、知識核實機制、知識分層儲存與知識更新機制進行知識管理。透過四階層機制之運作，將使不同類型之知識能以相似之表達方法呈現，並透過核實機制針對以形式化語意呈現之知識進行審核，以將所有知識轉換為相似型態並將知識分類與分層儲存於系統中，以協助知識管理者管理龐大且變動頻繁之知識。其中，使用者擷取資料之來源數眾多且常涉及龐大之資料紀錄，以致於尋找並轉換符合使用者需求之資料格式時須花費大量時間，因此，Pérez 等人(2007)以資料探勘網格結構(Data Mining Grid Architecture; DMGA)為基礎設計一套系統，並利用 Apriori 關聯規則演算法為基礎發展一個網格運算法。該系統首先針對大規模之資料分析屬性，從中將資料分類至同屬性之網格中，以將資料過濾及複製再放置合適之資料庫中，該系統亦可針對資料庫整合資料存取之過程，於擷取資料之過程中分散計算再進行彙整以給予使用者完整之資料，因此，藉由該系統之運作將可提升使用者對資料擷取之流暢度，且藉由分散計算過程將可降低資料擷取之時間。

綜合上述，社群文章審核機制議題中多數研究乃針對文章內文或知識流加以分析，以能協助社群更準確過濾或轉換，使得文章審核技術更加多元且應用範圍更為廣泛，而本研究於此議題之探討得知知識過量而不易管理之問題，是故，本研究即協助社群管理者審核大量之文章，從中快速取得違規之文章，以使社群之功能性更加完善。於此議題相關文獻之應用技術與分析過程彙整如表 2.2。

表 2.2、社群文章審核機制文獻彙整表

文獻議題	過去研究	採用技術	審核過程	解決之問題
以管理者進行審核	Gu 與 Grossman (2010)	知識分享平台	採用 P2P 模式使知識擷取及分享者能隨時共享知識，並利用分佈式知識分析法將知識進行分類	協助管理者管理過量之知識
	Matsubara 等人 (2006)	點對點技術	藉由中央管理器控制虛擬目錄中知識檔案之安全性及內容	幫助管理者審核知識安全性之問題
	Ercan 及 Cicekli (2007)	基線系統為基之關鍵詞擷取模式	依據文章基線圖中詞彙特徵，自動推論文章之核心關鍵詞	比對關鍵字語意關係以取得文章關鍵詞
	Li 等人 (2012)	馬爾可夫鏈之社交網路系統	分析所有知識專家之社交親密度、知名度，從中獲取最合適使用者之知識專家	協助使用者容易取得符合需求之知識
以審核系統進行審核	Yu 等人 (2006)	潛在語意分析、相似句演算法	利用潛在語意分析法分析所有候選語句與問題之相關性	回饋更精確之答案供使用者參考
	Li 等人 (2011)	N 連詞、點式交互訊息法	將候選屬性詞中屬於情感詞、單字動詞與特殊動詞之詞句去除，以保留真正之屬性詞	從特定領域之網路評論中取得屬性詞
	Jonathan 及 Yiu (2006)	對比聚類與關聯聚類法、公制群集分析法	統計兩詞彙之交錯頻率、共生頻率並向量化，以解析兩詞彙之向量距離	比對所有詞彙後以取得相似之文件
	Lai (2007)	知識工程理論模式	透過四大核查機制針對語意呈現之知識進行審核，以轉換為相似型態並分類與分層儲存於系統中	協助知識管理者管理龐大且變動頻繁之知識
	Pérez 等人 (2007)	Apriori 演算法、網格運算法	將資料分類至同屬性之網格中，以將資料過濾及複製再放置合適之資料庫中	提升使用者對資料擷取之流暢度

2.3 社群文章資料探勘

針對社群文章資料探勘之課題，本研究乃針對「社群文章語句重組技術」以及「社群文章評級分析」等兩議題進行相關文獻探討，以更深層瞭解文章語句重組之方法，並從中探討評級文章之分類方法與技術。

2.3.1 社群文章語句重組技術

於社群文章語句重組技術議題中，本研究乃針對「文章重點區塊分析」、「文章摘要建立技術」以及「文章語句合適度分析」等三主題進行相關文獻探討，期望從中探討語句重組與改寫之方法。

(A) 文章重點區塊分析

由於網路商品之評價更新快速且大多皆有正反兩面之評價，為協助網路使用者可快

速得知商品最新之評價及情感傾向性，Pang (2010) 提出一個基於主題規則之評價分析方法，先行將商品文件標記詞性並分詞，並依據商品之品牌、型號等數據建立候選主題詞彙，同時透過情感詞彙之頻率建立評價正反情感之傾向性，其次，依據主題詞彙、中英品牌、詞性關係等過濾之規則篩選與商品較無關之主題，最後，藉由動態共詞法 (Co-word Analysis) 針對主題中語句位置、詞頻等計算主題與商品之範圍程度 (即評價所評論之範圍)，並以向量空間模型 (Vector Space Model; VSM) 計算主題與評價之相似性，以得知與商品文件最相關之評價。藉由該研究之技術可有效快速取得商品之最新評價，並同時得知商品之正反評價以供一個網路使用者購買商品之參考依據。此外，使用者對網路商品評論之情感傾向及特徵大多可視為商品及服務品質之依據，為協助網路服務商取得使用者對產品之評論重點，Li 等人 (2008) 乃先行標記評論之語句詞性，並以語句為單位利用 LingPipe 自然語言分析語句之情感傾向，再以 Aprori 關聯規則演算法建構商品之特徵 (如外型、功能) 屬性並排序商品特徵詞彙之頻率，以利用具有商品特徵之語句依頻率權重比例推論評論之情感傾向性，進而依據特徵語句之頻率權重作為商品之主要特徵。藉由該研究之方法，可幫助網路賣家取得使用者對商品之評價及關注特徵，亦可提供網路賣家改進產品或提高服務之決策依據。

另一方面，為了能從文字數較長之網頁文件中取得重要語句，並將語句組合為該文件之主旨，Chan (2006) 提出一個擷取最具代表性語句之量化模式，首先透過潛在語意分析技術解析文件之內容以建立各語句之語意關係，並將具有代表性或連結性之詞彙串連成詞彙網路圖，透過詞彙之連結給予語句間之關聯程度及語句順序編號，同時以語句之詞彙向量加強相似語句之關聯權重，以藉此分析最相關之語句並視為代表性語句，進而依據語句之順序及數個代表性語句生成文件之主旨。透過該研究之方法，使用者可有效於大量且文字數過長之特定領域中得知所有網頁文件之重點，以協助使用者有效率地閱讀並節省過濾所有文件之時間。甚之，為了能從非結構化之產品評論文章中取得評論之情感類別，以協助使用者了解產品之特性，Fan 等人 (2009) 乃先定義主題之概念，利用詞語之延伸詞彙集合出現頻率分析該集合之重要程度，同時分析詞彙集合於評論各段落之頻率，以得知詞彙集合之分佈範圍，並將詞彙集合之重要程度與分佈範圍整合，以得知該產品評論之主題概念 (即詞語之延伸詞彙集合)，其次，將評論中包含主體概念之語句視為候選語句並與主題概念比對相似程度，並藉由條件隨機域 (Conditional Random Field; CRF) 分別針對候選語句標記情感詞彙、轉折詞及連結詞等權重值之整合，篩選與主題最為相關之語句，並透過情感詞彙之詞頻數分析語句之情感類別，以作

為評論之情感傾向。透過該研究所提之方法，使用者可有效於非結構化之評論中取得情感傾向，亦可得知評論之重點主題，以幫助使用者更精確了解產品之優缺點。

(B) 文章摘要建立技術

虛擬社群常限制使用者須以制定格式發表文章，但考量非所有文章皆以特定格式(如標題與語句位置)撰寫於此影響多數文件摘要技術之形成，因此，Ko 與 Seo (2008) 提出一個混合式統計連續虛擬語句之摘要形成模式，先行透過雙連搜尋 (Bi-gram) 技術將文件之兩個連續語句結合成一個虛擬語句，並以實驗法方式加入使用者滑動視窗於兩語句間之時間權重，再以詞頻計算方式 (TF-based Query Method) 篩選關鍵字詞，將關鍵字詞作為語句相似度衡量依據，以取得最重要、最具代表性之語句以形成文章摘要。藉由該研究之方法，可有效於多文章與非結構化文章中取得代表性語句以形成摘要。而針對自動摘要技術常忽略文章次主題之問題，其形成之摘要將可能不完整。是故，Xu 等人 (2008) 提出一個局部主題 (段落) 為基之關鍵句擷取方法以形成較完整之摘要，該研究透過向量空間模型將文章中各段落之語句詞彙向量化，以給予各語句詞彙位置權重，並藉由階層式聚合分群法 (Hierarchical Agglomerative Clustering; HAC) 將文章劃分數個局部主題並計算各語句間之相似度，以使相似之局部主題合併為一個局部主題，再透過各語句詞彙之位置權重分析各局部主題之關鍵句，擷取數個重要關鍵句形成各局部主題之摘要。該研究藉由局部主題之關鍵句所形成之局部摘要，可有效擷取次主題之重要語句，進而形成較完整之文章摘要供使用者查閱。

此外，有鑑於產品評論數過多，導致使用者須花費大量時間閱讀方可得知評論之重點，是故，Zhang 等人 (2009) 利用詞類標記器 (Part-of-Speech Tagger) 先標註訓練評論之詞性，同時標註具有產品特徵之名詞組合並區分為顯性 (直接描述產品之詞句) 及隱性 (須延伸與詞彙組合之詞句) 特徵，並透過特徵詞之前後順序與詞頻門檻值分類語句之顯性及隱性，該研究將提取顯性內之語句作為語句之重點，利用顯性內所有語句之特徵向量訂定特徵標準值，再計算各語句之相似向量、長度及首尾權重三種分數並加總，即可依據分數高低區分語句重要順序，並依據段落數給予使用者對應之語句，以幫助使用者於大量之評論中，取得各評論之重點語句與主旨，進而節省閱讀評論之時間。

(C) 文章語句合適度分析

醫療領域中具有大量之專有名詞、歧異詞及分歧句，導致知識需求者於網路中尋求

資訊時，因搜尋字詞無法與該領域之特殊字詞串聯，以致搜尋之效果不彰，有鑑於上述問題，Win 及 Zhang (2006) 提出一個以本體語義為基之知識檢索方法並將其方法分為三步驟，首先，利用 UMLS 之醫療資料將特殊字詞以標籤方式定義語意間之關係，以建構特殊字詞之本體知識，進而串聯醫療領域之本體知識及語義規則，其次，分析知識需求者查詢語句與本體知識之語義相關性分數，以延伸查詢語句之查詢詞彙，進而擴大搜尋相關之醫療知識文件，最後，比對查詢語句與醫療知識之標題、關鍵詞之相似程度，以藉由文件相似度之排名給予知識尋求者相對應之醫療知識。透過該研究之方法可協助知識尋求者於不熟悉之領域中，以簡短之搜尋字串中取得欲得知之領域知識。而維吾爾語具有延展性、時態變化及複雜之語法結構等特性，以致於過去研究甚少針對維吾爾語分析語句之情感傾向，因此，Huang 等人 (2012) 首先利用情感詞多數分佈於形容詞、名詞及動詞前後之特性，以人工方式標記情感詞並依據情感詞之詞頻給予情感係數以建構情感語料庫，之後透過語句各詞彙與情緒詞之序列位置，以條件隨機域模型分析語句局部特徵與各情感類別之條件機率，以藉由各局部特徵中最大之條件機率作為該語句之情感類別，此外，該研究同時針對語句因否定詞之存在，而導致情感類別反轉之可能性建立一個調整規則，若語句中存在奇數個否定詞時，其語句之情感類別極可能與首次分析之結果相反，若為偶數則為一致。透過該研究情感分析方法及調整規則，可有效協助使用者於語句結構複雜且存在轉折詞彙(否定詞)之維吾爾語中，分析語句之情感類別。

除此之外，問答系統所回應資訊之品質，多數與語料庫(詞彙、語句之組合)建置完整度成正向相關，然而，語料庫之建構多數須透過問與答之訓練資料串連語句間之關係，方可形成專業之語料庫，因此，Hao 等人 (2007) 乃提出一個基於問題類型導向之語料庫建立方法。該研究利用語意知識庫 (Chinese FrameNet; CFN) 標記訓練資料中所有語句目標詞、所屬之語意框架(依據語意知識庫定義之框架)、語句類型及語法，該研究同時定義八個問題類型(人物、地點等)，並將所有語句標記所屬之問題類型，其次，根據目標詞之重複性及多重語意關係建立語意框架，以使目標詞與語句之定義更能明確劃分，最後，加以標註含有多重目標詞之語句，將該語句延伸鏈結其他目標詞之關係，即透過上述之方法，可將詞彙與語句、歧義詞、問題類型之關係建立鏈結，以形成一個較完整之語料庫，進而協助使用者以語料庫為基礎給予需求者適當之回應內容。

綜合上述，針對社群文章語句重組技術中多數研究以主體概念及詞彙特徵為基礎分析文章之重點區塊，其次，藉由語句相似程度與主題關鍵句建立文章摘要，甚之，透過本，以使文章語句重組等相關技術與方法更加多元化，而本研究之目的即針對文章之

違規語句進行修正，以保留有討論價值之文章，因此，本研究乃透過相似語句分析技術、以及語句之語意關係分析技術，針對違規語句進行重組，以能修改文章內之違規字詞，亦能維持文章中各語句之原意。於此議題文獻之應用與技術分別彙整成表 2.3。

表 2.3、社群文章語句重組技術文獻彙整表

文獻議題	過去研究	使用技術或方法	分析方式與過程
文章重點區塊分析	Pang (2010)	動態共詞法、向量空間模型	依據主題詞彙、中英品牌、詞性關係等過濾之規則篩選與商品較無關之主題，以得知與商品文件最相關之評價
	Li 等人 (2008)	LingPipe 自然語言、Aprori 關聯規則	建構商品之特徵屬性並排序商品特徵詞彙之頻率，以依特徵語句頻率權重比例推論評論之情感傾向性，進而作為商品之主要特徵重點
	Chan (2006)	潛在語意分析、	透過詞彙之連結給予語句間之關聯程度及語句順序編號，同時以語句之詞彙向量加強相似語句之關聯權重，以依據語句生成文件之主旨
	Fan 等人 (2009)	條件隨機域	將主體概念之語句視為候選語句，並針對候選語句標記情感詞彙、轉折詞及連結詞等權重值之整合，篩選與主題最為相關之語句
文章摘要建立技術	Ko 與 Seo (2008)	雙連搜尋、詞頻計算法	將關鍵字詞作為語句相似度衡量依據，以取得最重要、最具代表性之語句以形成文章摘要
	Xu 等人 (2008)	階層式聚合分群法	將文章劃分數個局部主題將相似之局部主題合併為一個局部主題，藉由局部主題之關鍵句所形成之局部摘要形成完整摘要
	Zhang 等人 (2009)	詞類標記器	計算各語句之相似向量、長度及首尾權重三種分數並加總，即可依據分數高低區分語句重要順序以形成評論主旨
文章語句合適度分析	Win 及 Zhang (2006)	本體語義、知識檢索方法	將特殊字詞以標籤方式定義語意間之關係，並比對查詢語句與醫療知識之標題、關鍵詞之相似程度，以藉由相似語句之查詢給予對應之醫療知識
	Huang 等人 (2012)	條件隨機域	透過語句各詞彙與情緒詞之序列位置，分析語句局部特徵與各情感類別之條件機率，以藉由各局部特徵中最大條件機率作為該語句之情感類別
	Hao 等人 (2007)	語意知識庫、問題類型導向之語料庫建立方法	定義八個問題類型，並將所有語句標記所屬之問題類型，同時根據目標詞之重複性及多重語意關係建立語意框架，以使目標詞與語句之定義更能明確劃分

2.3.2 社群文章評級技術

於社群文章評級技術議題中，本研究乃針對「以前後文關係分析」、「以相異詞語分析法」以及「以語言文法特性分析」等三個主題進行文獻探討，期望從中瞭解文章評級議題所涉及之範圍與領域。

(A) 以前後文關係分析

為了比較自動文章輔助 (Automatic Essay Assessor; AEA) 及 K 鄰近分類法 (K-nearest-neighbor based; K-NN) 兩個文章語意分級方法之成效, Kakkonen 等人(2008) 透過國文教師將兩方法之各項權重參數整合一致, 並以教師文章作為訓練文章 (即較無特殊字詞或模糊語意), 以公平比較兩方法之分級成效。該研究利用自動文章輔助系統中潛在語意分析、機率潛在語意分析 (Probabilistic Latent Semantic Analysis; PLSA) 及潛在迪利克雷分配法 (Latent Dirichlet Allocation; LDA) 等三個分析法針對文章進行分級, 並各別與 K 鄰近分類法比較分級準確度, 其結果顯示, 自動文章輔助系統中三個分析法之文章分級效果皆優於 K 鄰近分類法, 且三個分析法中因所分析文章較為完整之因素, 故以潛在語意分析法之文章分級成效最佳, 是故, 透過該研究之比較結果, 有助於教育單位建置文章分級系統時, 推薦一個分析文章等級之參考方法。甚者, 由於網路自由評論 (影視、產品及新聞評論等) 多樣之寫作風格、格式及詞彙語意定義, 導致詞彙之情感傾向值僅適用於單一領域之評論中, 為能同時建立自由評論中詞彙之情感傾向, Yin 及 Peng (2012) 提出一個情感傾向性關係網路 (Sentiment Orientation Relationship Network; SORN) 概念, 藉以分析詞彙於自由評論之情感傾向值。該研究首先乃以特殊詞 (轉折詞、否定詞) 定義詞彙與前後文關係之規則, 並針對詞彙之序列長度給予路徑關係值以建立詞彙與所有詞彙 (包含情緒詞) 之路徑關係圖, 同時透過詞彙與前後文中所有情緒詞之詞頻及多重前後文語意之關係調整詞彙間之關係係數, 以精確取得詞彙間之路徑關係, 進而取得詞彙與情緒詞彙之情感傾向值。藉由該研究所提出之概念, 使用者可不須依靠情緒詞彙庫及人工標記情緒詞, 即可自動於各領域中自動建構所有詞彙與情感詞彙之傾向值, 以建立特定領域之情緒詞彙庫。

此外, 由於網路訊息常以簡短、不分段及無標點符號之形成發佈, 導致分析訊息常無法取得內文之重點與潛在之語意關係, 使得分類之結果不準確, 因此, Peng 等人(2007) 建立一個語意內積空間分析模型 (Semantic Inner Space Model), 分析訊息之重要部分與訊息之相似度, 以將相似之訊息歸為一類以達到分類之效果。該研究利用知網所提供之詞彙特性, 定義訊息中所有詞彙之上下文關係, 以針對該訊息建立原義概念圖, 並將訊息轉換為向量空間以計算詞彙、語句間之相似度, 同時依據詞語之相似度分配訊息內之候選代表點 (亦含有向量空間與相似值) 與聚類值, 以取最大之向量維度作為訊息代表點, 其次, 該研究依據訊息代表點再分為兩代表點 (最無相關之兩點) 將詞彙分類至對應之類別, 並再計算所有候選代表點內所有詞彙及語句之相似度, 以重新計算候選代表點之聚類值至不再變化為止, 進而取得訊息中相關度最低之兩集合向量空間, 最後, 再

將兩集合向量空間重新分配代表點並計算相似度，以藉由權重值將相似之代表點歸類為重點向量空間以作為訊息之重點，並根據重點向量空間與所有訊息加以比較，將所有訊息中相似之重點向量聚集為一類，即可分類非結構化且簡短之網路訊息。

另一方面，非英語系國家之使用者撰寫英文後，多數乃透過書本或人工批閱之方式，方可得知所有語句中文法之正確性，然而，此舉大多須花費大量時間及人力，因此，Lee 等人 (2011) 乃提出一個英文語句之文法分析機制，以幫助使用者立即判斷語句文法之正確性。該研究乃利用韓國教育部之英文文法辭典及詞彙建構一個語法語料庫，並同時透過 WordNet 中詞彙之語法架構，分析訓練語句上下文之同義詞、代名詞等規則，以加強語料庫之語法規則，其次，該研究以語法語料庫為基礎開發一個文法解析器，從中分析語句之拼寫、語句所延伸之語法以建立語句之語法依賴樹，最後，利用目標語句之語法依賴樹比對所有之訓練語句，以藉由相似語法依賴樹之語句文法對照目標語句，即可得知目標語句之文法正確性，以藉此幫助使用者快速得知語句文法之正確性。

(B) 以相異詞語分析法

為了協助使用者於大量之作文中，自動取得與題目不符或離題之作文內容，Ge 及 Chen (2009) 乃先將作文之數字、專有名詞等特徵詞歸類並過濾停用詞，以將作文內重要詞彙之詞根（不同時態或動作時詞彙之變化）還原，接著將所有作文之特徵詞及特徵詞之詞根以詞頻-逆向文件頻率法（Term Frequency-Inverse Document Frequencies；TF-IDF）方法計算特徵之向量，以取得所有特徵之特徵權重值，進而利用向量空間模型之餘弦函數分析所有作文間之相似度，最後，該研究以階層式聚合分群法將相似度相近之作文分群於一類，並重新計算該群集內所有作文之相似度，以將同群集內之作文再分群直至剩餘一個群集為止，藉此形成一個聚類樹，以將未分類於任一群集內之作文內容視為離題之作文。透過該研究之方法，使用者可於不同領域之作文題目中，自動取得偏離題目之作文內容，並可有效節省人工審核及過濾各篇作文之時間。

甚之，過去研究利用 N 連詞作為語句特徵分析語句之情感類別時，語句之特徵大多以連續出現之詞彙為主，時常忽略更為重要之單一詞彙，而重要語句乃以語句結構規則作為擷取依據，以致較無法取得非結構化且重要之語句，將可能導致分類結果錯誤或成效不佳，為了解決上述問題，Zhang 等人 (2011) 提出一個核樹（Tree Kernel）函數為基之語句情感分類方法，將所有語句作為核心樹並將詞彙特徵分支為數個子樹，以形成離散樹狀圖形，接著利用離散之向量空間計算兩語句間之相似程度，以將兩相似語句結

合為一個結合樹狀，最後，透過支援向量機整合結合樹狀內各詞彙特徵之情感值，以分析該結合樹之情感傾向，進而得知所有語句之情感類別，以解決使用者因忽略分析非結構化語句與詞頻數較少之詞彙等數據，以致分類語句情感類別時較不精準之問題。

另一方面，**Jiang 等人 (2011)** 乃以英文文法、型態等規則針對中國大學生撰寫之英文作文進行評級分類，以幫助審閱者節省人工批改之時間。該研究乃事先隔離文章中所有之字詞以擷取較為重要之詞語，並以史丹佛分析器 (Stanford Parser) 分析所有訓練文章之語法結構以形成語法規則，接著，利用詞頻-逆向文件頻率法、互信息 (Mutual Information; MI)、訊息增益法 (Information Gain; IG) 等統計方法，分析文章所有特徵詞以形成特徵詞向量權重，進而整合特徵向量距離及語法規則以作為分類之標準，最後，透過 K 鄰近演算法 (K-Nearest Neighbor; KNN) 分析目標文章之向量距離權重，再依據目標文章語法程度，將相似距離之特徵向量與語法歸為一類，即可幫助審閱者將相似語法程度之文章分為一類，進而節省審閱者須逐字批改及文法對照之時間。

(C) 以語言文法特性分析

為了協助使用者於大量英文文章中，以客觀角度評閱文章之級分數，**Huang 等人 (2009)** 提出一個基於投票演算法之多分類器文章評分系統，以協助使用者針對中國人所撰寫之英文文章自動評閱分數。該研究將具有標記錯誤文法、詞性特性之中國學習者英文語料庫 (Chinese Learner English Corpus; CLEC) 作為訓練文章庫，並以文章之表達特徵及寫作特性給予評分分類之等級分數，其次，以投票演算法將目標文章透過向量特徵及領域專家之觀察，計算文章於單純貝式 (Naïve Bayes)、K 鄰近及支援向量機三個分類技術之隸屬值，以將目標文章放置合適之分類器中，最後，依據分類器之評閱分數結果作為最終評閱結果，此外，該研究同時比較單一分類技術與整合技術 (即單純貝式、K 鄰近與支援向量機技術) 之分類效果，結果顯示整合技術之分類成效皆優於任單一技術之分類效果，即表示該研究所提之方法，可依照使用者不同之撰寫格式與風格，自動放置合適之評閱分類器中，且具有良好之分類效果，以回饋使用者精確之文章評閱結果。甚者，由於現有情緒詞彙庫所提供之情緒傾向值，可能隨著不同領域之語意及專有名詞之定義而有所改變，因而導致情緒詞彙庫不適用之問題，是故，**Wang 等人 (2011)** 乃以本體論為基提出一個情緒詞情緒傾向計算方法，先將特定領域之文章斷詞處理，以取得之文章特徵詞 (名詞、形容詞、動詞及副詞) 進而形成文章主題、語料之空間向量，接著分析知網建構之情緒詞與文章主題、語料特徵詞之相似度，並藉由特徵詞之詞頻及

情緒詞相似度之整合，初步計算特徵詞之情緒傾向正負值，之後依據情緒類別間之正反關係、否定詞及程度副詞予以相似度權重值，以調整最後情緒傾向值反方向之可能，以精確得知特定領域特徵詞之情緒傾向值，進而提供使用者依據特定領域之主題特徵建立特徵詞之情緒傾向值，並針對特定領域建立情緒詞彙庫。

除此之外，過去研究大多乃針對網路文章或大學生之寫作程度作為文章評級之標準，甚少研究根據國小學童之撰寫能力分析語句之評級分數，有鑑於此，Liao 等人(2012) 乃提出一個語句評級方法以幫助使用者針對國小學生所撰寫之語句，建構一個合適之語句評級標準。該研究首先藉由小學生以字造句(即給予數個詞彙由學生自行組合成語句)之語句作為訓練語句，並以潛在語意分析法及中央研究院漢語平衡語料庫(Sinica Corpus)所提供之詞彙語意，針對訓練語句個別分析詞彙及語句之語意權重以藉此建立評級標準，最後，該研究利用向量空間模型之餘弦函數，取得目標語句之所有相似語句，以整合所有相似語句之詞彙及個別語句之語意權重，進而取得目標語句之評級分數。透過該研究之方法，可協助使用者根據整體寫作程度建立評級標準，並以此為依據透過相似之語句語意程度分析語句之評級分數。

綜合上述，針對社群文章評級分析部份，過去研究大多以上下文語意之關係及詞彙特徵建立評級標準，甚者，透過詞語相似度分析文章之評級分數，以使文章評級分析技術應用範圍更為廣泛。本研究於此議題之探討得知相異詞語之分析法分析語句連貫性及流暢度之成效較佳，亦得知結合語意關係分析技術將使語句評級之成效更為完整，是故，本研究乃透過兩項技術之整合藉此分析文章內文之流暢性。而於此議題之相關文獻依其不同應用與技術分別彙整成表 2.4。

表 2.4、社群文章評級分析文獻彙整表

文獻議題	過去研究	使用技術或方法	分析方式與過程
以前後文關係分析	Kakkonen 等人 (2008)	K 鄰近分類法、潛在語意分析、機率潛在語意分析、潛在迪利克雷分配法	比較三個分析法針對文章分級之成效，並各別與 K 鄰近分類法比較分級準確度，其結果顯示，三個分析法分級效果皆優於 K 鄰近分類法，且以潛在語意分析法之文章分級成效最佳
	Yin 及 Peng (2012)	情感傾向性關係網路、	透過詞彙與前後文中所有情緒詞之詞頻，以及多重前後文語意之關係調整詞彙間之關係係數，取得詞彙與情緒詞彙之情感傾向值。
	Peng 等人 (2007)	語意內積空間分析模型	定義訊息中所有詞彙之上下文關係，並轉換為向量空間，以根據重點向量空間與所有訊息加以比較，進而將相似之重點向量聚集為一類
	Lee 等人 (2011)	文法分析機制	分析語句之拼寫、語句所延伸之語法以語法依賴樹，以藉由相似語法依賴樹之語句文法對照目標語句，即可得知目標語句之文法正確性
以相異詞語分析法	Ge 及 Chen (2009)	詞頻-逆向文件頻率法、向量空間模型、階層式聚合分群法	將相似度相近之作文分群於一類，並重新計算該群集內所有作文之相似度，以將同群集內之作文再分群直至剩餘一個群集為止，藉此將未分類於任一群集內之作文視為離題之作文。
	Jiang 等人 (2011)	核樹函數、支援向量機、	將兩相似語句結合為一個結合樹狀，並整合結合樹狀內各詞彙特徵之情感值，以分析該結合樹之情感傾向，進而得知所有語句之情感類別
	Jiang 等人 (2011)	詞頻-逆向文件頻率、互信息、訊息增益法、K 鄰近演算法	整合特徵詞向量距離及語法規則作為分類標準，再依據目標文章語法程度，將相似距離之特徵向量與語法歸為一類
以語言文法特性分析	Huang 等人 (2009)	投票演算法、單純貝氏、K 鄰近、支援向量機	以文章之表達特徵及寫作特性給予評分標準，並將目標文章放置合適之分類器中，以依據分類器之評閱分數結果作為最終評閱結果
	Wang 等人 (2011)	本體論、文章斷詞	藉由特徵詞之詞頻及情緒詞相似度之整合，初步取情緒傾向正負值，再透過正反關係、否定詞及程度副詞，分析準確之情感傾向值
	Liao 等人 (2012)	潛在語意分析、向量空間模型	針對訓練語句個別分析詞彙及語句之語意權重以藉此建立評級標準，並取得目標語句之所有相似語句，以整合所有相似語句之詞彙及各語句之語意權重，進而取得目標語句之評級分數

2.4 文章撰寫者行為探討

針對文章撰寫者行為探討之課題，本研究乃針對「文章撰寫者情感分類」、「文章撰寫者寫作習慣分析」以及「文章閱讀者閱讀感受分析」等三議題進行相關文獻探討，以更深層瞭解社群使用者撰寫與閱讀之習慣。

2.4.1 文章撰寫者情感分類

針對文章撰寫者情感分類議題中，本研究乃針對「以資料探勘進行分類」、「以詞語相似度進行分類」以及「以語意關係進行分類」等三主題進行相關文獻探討，期望從中

探討不同分類技術之成效與應用範疇。

(A) 以資料探勘進行分類

新聞評論所涉及之主題（人、事件及物等）甚為廣泛，以致於過去研究分類新聞評論之情緒類別時，會因主題之不同而有所誤差，有鑑於此，Yang 等人（2011）乃先將評論進行分詞處理以分析詞彙於新聞主題詞頻關係，進而取得語句與新聞主題之隸屬值，之後藉由共同詞彙、個別詞彙數及語句長度計算所有語句與新聞主題句之相似度，並將具有情緒詞及高相似度之語句作為候選主題情緒句，最後，該研究利用所有主題情緒句之情緒詞彙，與情緒類別之隸屬值加權平均，並加入主題情緒句與新聞主題之隸屬權重值，以計算評論之情緒類別係數，以達到新聞評論之情緒分類成效，進而協助使用者針對不同主題之新聞評論，取得評論之情緒類別。而為了從部落格、網路、論壇及新聞類型等新型文章中（即文句較簡短精簡之評論）取得使用者之情感傾向（針對積極及消極兩情感類別），Yang 等人（2010）利用網路最短覆蓋路徑演算法（Shortest Covering Path；SCP）針對多種類型之評論，以去除無意義之詞彙、按順序組成語句、計算語句之節點權重等處理過程，將各語句節點連結為序列網路模型（該評論類型之語意網路），並根據此模型分析評論中最短路徑之節點及長度，以藉由最短覆蓋路徑之隱馬爾可夫模型（Shortest Covering Path- Hidden Markov Model；SCP-HMM）計算該語句之路徑節點所隱含之情緒傾向性，進而推論評論所有路徑節點之情感傾向性，即可得知評論之情感傾向，是故該研究可有效針對語句簡短且多種類型之評論進行情感分析，且該研究可隨著訓練評論量之增加而提高情感分類之準確率。

此外，中日關係論壇之文章結構具有複雜且大量特徵詞詞性等特性，較無法以自然語言方法分析文章之情感類別，故 Wang 及 Li（2010）針對上述問題先以人工方式選取文章之指代（即文章之回應者乃承接他人之回應內容）、代字及縮寫詞三種結構，並透過訊息增益、互信息、 χ^2 統計量（Chi-square）三種特徵選擇法計算三種文章結構之特徵權重，以得知文章最為重要之結構，接著以支援向量機針對正反兩面之情感路徑加入三種特徵權重，以依據文章特殊結構之重要區塊計算兩情感類別之傾向，以協助使用者於具有特殊詞彙且複雜結構之論壇文章中取得情感類別。且過去研究分析詞彙之褒義及貶義，多數乃藉由已標記褒貶義強度詞彙之語料庫（如 HowNet），直接分析詞彙與語料庫之詞彙關係以進行分類，甚少研究藉由詞彙之同義詞（語言學中最小粒度之層級）關係分析詞彙之情感傾向，是故，Wang 等人（2009）利用點式交互訊息法解析目標詞彙

與所有詞彙之關係強度，藉以取得目標詞彙之同義詞集，其次，利用五種語料庫情緒詞彙之詞頻所形成之情感傾向值，個別分析同義詞之褒貶義關係強度，以透過同義詞之詞頻、褒貶義關係強度數據之整合，計算目標詞彙之情感傾向，以提供使用者於不同文章中取得詞彙褒貶義關係，進而於不同領域中建立詞彙之情感傾向表。

另一方面，過去研究多數以基於「語意規則」與「統計」兩方法分析文件之情緒類別，然而，基於規則方法可能因文件特徵之間相互矛盾（例如：文件主題為「天氣」，主要特徵可能為「冷」與「熱」兩矛盾特徵），使得特徵之結構錯誤，而基於統計方法則可能因情緒詞庫之不完整，以致於忽略比對文件中更為重要詞彙，有鑑於上述兩項分析方法之缺點可能導致情緒分類結果不精確，Zan 等人 (2011) 將此兩方法結合以互補優缺點，進而建構一個情感分析整合方法。該研究先將文件以人工定義規則（依據語句之完整性）判定至分句集合或整句集合中，接著透過詞彙之詞頻建構所有語句特徵之分佈向量，最後，透過貝葉斯 (Bayes) 分類法分析所有語句，以取得文件中最重要之語句特徵，並以情緒詞彙詞頻數判斷語句之情緒傾向以作為文件之情緒類別。透過該研究之整合方法，可有效減少擷取錯誤語句特徵之發生，亦可取得文件中最為重要之語句特徵，進而提升分類文件情緒類別之準確率。

(B) 以詞語相似度進行分類

短句（或詞彙）之情緒傾向常隨著領域之變化而有所不同，亦即短句於不同領域中將有多種解釋及語意表達方式，導致短句可能於不同領域之情緒分類結果有所誤差，有鑑於此，Lin 等人 (2011) 提出一個以點式交互訊息法為基之情緒分類模式，先將文章之形容詞與副詞利用 Google 搜尋引擎搜尋訓練詞彙之片段並擷取詞彙之情緒短句，透過點式交互訊息演算法計算情緒短句間之關聯性，同時利用關聯性較低之短句建立相反情緒之短句，並計算情緒短句之極性傾向（即正負情緒兩面之傾向）程度，以得知短句之情緒類別。透過該研究之方法，可以非監督學習方式於多元領域中快速分析短句之情緒傾向。甚之，有鑑於情緒詞彙於不同領域之情緒傾向值（正負面之情緒）將有所差異，為協助使用者於多元領域中自動建立情緒詞彙之情緒傾向，Liu 等人 (2009) 藉由知網之相似演算法計算詞彙各種語意（即詞彙於知網中所延伸之語意）之語意距離，接著分析詞彙之語意距離與知網所含正負情緒詞之相關程度，並藉由正面與負面情緒之傾向值差異程度取得詞彙之情緒傾向值。該研究表示，比對詞彙與知網所建構之語意關係於多元領域中取得重點語意後，再分析重點語意與正負面情緒之傾向程度，可提供使用者取

得特定領域詞彙之情緒傾向係數，進而幫助使用者建立特定領域之情緒詞彙庫。

另一方面，情緒詞會隨著領域之不同而有其他解釋與定義，尤以虛擬社群所涉及之領域最為廣泛，使得使用者較無法於社群中同時分析文章語句之情緒類別，有鑑於上述問題，Zhan 等人 (2011) 首先乃標記文章之語氣轉折點（即文章內之轉折詞），以取得轉折點前後情緒詞彙之序列，並依據文章中程度副詞、否定詞對於情緒詞之影響程度，加權情緒詞之情緒傾向係數，以藉由轉折點前後程度副詞、否定詞，分析情緒詞對於語句之影響值，以累加計算語句之情緒傾向值，進而取得各語句之情緒類別。透過該研究之方法，使用者可根據各領域之特性自動加權情緒詞之情緒傾向值，藉此於多領域中同時分析語句之情緒類別。最後，由於現有孟加拉語之情感詞彙庫完整度不足問題，導致使用者難以取得網路文章之情感傾向，因此，Das 及 Bandyopadhyay (2010) 利用部落格文章作為訓練資料，透過孟加拉語之詞彙與 WordNet 提供之詞彙相互對照，將孟加拉語之情感詞彙依據文章語句之層級（標題、主題、次主題及內文語句等）加以標籤化以給予詞彙之情緒權重值，透過主題語句所含之情感詞彙、情感詞彙於評論及各層級語句之頻率、各情感類別之最長語句權重值之整合，推論文章之情感類別，以幫助使用者於孟加拉語部落格中得知文章之情感類別。

(C) 以語意關係進行分類

語句之涵義常藉由上下文之串聯，方能得知語句欲表達之語意，但日本網路使用者常創造語意含糊之語句或詞彙（日本將此些語句及詞彙稱為「Wakamono Kotoba」），以致無法藉由語句涵義分析語句之情感類別，為解決上述之問題，Matsumoto 等人 (2011) 將網路常使用且可形容他人情緒之詞彙或語句，先行歸屬為 Wakamono Kotoba 之語料庫，並以 IPA、UniDic 及 Naist 日語詞彙字典之語意架構過濾 Wakamono Kotoba 內之詞彙，以建立 Wakamono Kotoba 情緒詞彙庫，最後透過貝氏分類法計算語句於各情緒類別之後驗機率，並結合累積分類法 (Accumulation Method) 計算語句中情緒詞彙之頻率，以整合兩分類法之計算數據推論語句之情緒類別。藉由該研究所建立之情緒詞彙庫，可協助網路社群中分析複雜、模糊且非結構化之語句或文件之情緒類別，並透過貝氏及累積分類法進行情緒分類以獲取良好分類效果。甚者，為協助中文論壇管理者了解使用者表達意見時之寫作情緒，Zhang 等人 (2009) 乃分析不同領域之中文論壇（中國亞馬遜論壇、音樂論壇等）使用者之意見，先行去除非情緒之語句，並利用調度規則 (Rule-Based) 語意分析法針對情緒語句建立依賴關係樹 (Dependency Tree) 及語意規則，以取得詞彙

間之情緒偏向，進而建立情緒語句之極性分數。該研究同時針對語句之轉折詞、否定詞、詞彙上下文關係分析語句情緒反轉之可能性，並針對情緒語句與標題之關係，調整情緒語句之極性分數。以聚合所有情緒語句之極性分數以得知使用者意見之情緒傾向。藉由該研究之方法可於不同領域之論壇中自動建立語意規則，且具有良好之情緒分類成效。

此外，Li 及 Liu (2012) 提出一個基於聚類之情感分析方法，以改善過去須花費大量時間人工訓練文件資料，以及利用詞彙之情感偏向分數加總平均導致分類準確率不佳之問題。該研究首先藉由 K 均值聚類演算法 (K-mean Clustering Algorithm) 針對文件庫之文件，依照形容詞之語意指向分為正向及負向情感之對照組，並利用詞頻-逆向文件頻率演算法計算目標文件與文件庫內所有文件之重要程度權重，以透過投票機制 (Voting mechanism) 反覆計算目標文件與兩對照組之重要程度，直至目標文件於其一對照組之重要程度趨近穩定時，即可取得目標文件之情感偏向，以幫助使用者於無訓練資料之情況下，自動判定文件之情感偏向，且具有良好之分類成效。

綜合上述，文章撰寫者情感分類方面所應用之技術甚廣，大多研究乃利用詞彙比對及語意規則等方法分析文章之情感類別，且透過不同分析之對象及語言而有所差異，因此，本研究所分析之語言為中文，並藉由此議題之探討得知，中文之語意結構及關係較為複雜，使得分析文章之情緒類別較為不易，是故，本研究乃利用過去文獻中較常分析中文之技術包含分詞處理、相似演算法、相似語句比對法等，分析文章主要語句之情緒類別，並透過相似語句之判定取得文章之情緒類別，以藉由此些技術之應用與分析，提升取得文章情緒類別之準確度。於此議題相關文獻之不同應用與技術分別彙整成表 2.5。

表 2.5、文章撰寫者情感分類文獻彙整表

文獻議題	過去研究	使用技術	研究方法
以資料探勘進行分類	Yang 等人 (2011)	分詞處理、詞頻關係	利用所有主題情緒句之情緒詞彙，與情緒類別之隸屬值加權平均，並加入主題情緒句與新聞主題之隸屬權重值，以計算評論之情緒類別係數
	Yang 等人 (2010)	網路最短覆蓋路徑、最短覆蓋路徑之隱馬爾可夫	將各語句節點連結為序列網路模型，並根據此模型分析評論中最短路徑之節點及長度，以計算該語句之路徑節點所隱含之情緒傾向性
	Wang 及 Li (2010)	訊息增益、互信息、 x^2 統計量、支援向量機	以三種特徵選擇法計算文章結構之特徵權重，接著以支援向量機針對正反兩面之情感路徑加入三種特徵權重，以依據文章特殊結構之重要區塊計算兩情感類別之傾向
	Wang 等人 (2009)	點式交互訊息法	透過同義詞之詞頻、褒貶義關係強度數據之整合，計算目標詞彙之情感傾向
	Zan 等人 (2011)	貝葉斯	透過貝葉斯分類法分析所有語句，以取得文件中最重要之語句特徵，並以情緒詞彙詞頻數判斷語句之情緒傾向以作為文件之情緒類別
以詞語相似度進行分類	Lin 等人 (2011)	點式交互訊息法	透過點式交互訊息演算法計算情緒短句間之關聯性，同時利用關聯性較低之短句建立相反情緒之短句，並計算情緒短句之極性傾向
	Liu 等人 (2009)	相似演算法	計算詞彙之語意距離，再分析詞彙與知網所含正負情緒詞之相關程度，以藉由正負面情緒之傾向值差異程度取得詞彙之情緒傾向值
	Zhan 等人 (2011)	情緒詞分佈權重	依據文章中程度副詞、否定詞對於情緒詞之影響程度，加權情緒詞之情緒係數，以分析情緒詞對於語句之影響值，進而取得語句之情緒傾向值
	Das 及 Bandyopadhyay (2010)	相似語句比對法	透過主題語句所含之情感詞彙、情感詞彙於評論及各層級語句之頻率、各情感類別之最長語句權重值之整合，推論文章之情感類別
以語意關係進行分類	Matsumoto 等人 (2011)	貝氏分類、累積分類法	透過貝氏分類法計算語句於各情緒類別之後驗機率，並結合累積分類法計算語句中情緒詞彙之頻率，以整合兩分類法之數據推論語句之情緒類別
	Zhang 等人 (2009)	調度規則、依賴關係樹、語意規則	利用調度規則語意分析法針對情緒語句建立依賴關係樹及語意規則，以取得詞彙間之情緒偏向，進而建立情緒語句之極性分數
	Li 及 Liu (2012)	K 均值聚類及詞頻-逆向文件頻率演算法、投票機制	透過投票機制反覆計算目標文件與兩對照組之重要程度，直至目標文件於其一對照組之重要程度趨近穩定時，即可取得目標文件之情感偏向

2.4.2 文章撰寫者寫作習慣分析

於文章撰寫者寫作習慣分析議題中，本研究乃針對「撰寫者寫作風格」以及「撰寫者用詞習慣」等兩主題進行相關文獻探討，期望從中探討文章撰寫者個人撰寫之風格與撰寫之習性。

(A) 撰寫者寫作風格

由於「微博」之評論具有多樣性、特有語意之風格等特性，常需以定義否定詞、連

接詞等方式自訂語意規則，方能對評論計算情緒指數，然而，過去研究甚少以機器學習方法（Exploiting Machine Learning Methods）自動產生語意規則以針對評論進行情緒分類，是故，Liu 及 Liu (2012) 乃以支援向量機、貝氏及 N 連詞三種分類方法，針對微博之電影評論進行情緒分類並比較三種分類方法之成效，同時比較微博與普通評論（無字數限制之評論）訓練評論庫之通用性；此外，該研究亦於支援向量機及貝氏分類中加入布爾（Presence）、詞頻（TF）及詞頻-逆向文件頻率法之特徵權重計算法，以比較三種特徵權重之成效差異。該研究結果表示，三種分類方法對於微博電影評論之情緒分類皆具極高之成效，其中採用支援向量機分類法及詞頻-逆向文件頻率法權重計算法之成效最佳，然而，微博與普通評論因寫作風格與字數限制之不同，兩者訓練評論庫互用後之情緒分類成效皆不佳。而網路評論常以複雜（數字、字數不定）且非結構化之方式呈現，導致分類評論情緒之傾向時，難以建立語意規則以致須以人工方式建立語意關係，因此，Bai (2011) 提出一個以馬可夫覆蓋（Markov Blanket Model）為基之情緒分類模式，先行利用馬可夫覆蓋模式自動建立詞彙與詞彙間之語意依賴條件圖，藉由語句之詞彙依賴關係取得詞彙與情緒類別之依賴程度，並利用 LrnMBC 演算法（LrnMBC Algorithm）偵測詞彙與重要語句間之情緒變化，以計算詞彙之情緒穩定值推論評論之情緒類別。透過該研究可不因字數之限制而影響情緒分類之效果，亦即表示該研究可於字數較少之評論中取得使用者欲表達之情緒類別。

另一方面，由於網路中各虛擬社群皆具有獨特之文化風格，使得社群中各網頁文章之情緒表達方式不一，以至於甚少研究可同時分析不同文化社群中，主題欲表達之情緒類別，是故，Fu 等人 (2013) 利用潛在迪利克雷分配之概念將主題以分層方式區分為文件層、主題層與文字層，並藉由使用者視窗滾動之軌跡參數（即虛擬社群使用者滑動捲軸之習慣），計算各層級之後驗機率以取得網頁文章重點主題（即取得文件、主題與文字層中重要主題），進而將重點主題所含詞彙與知網提供之情緒詞進行相似性比對，以計算重點主題之情緒相似傾向值，以得知重點主題之情緒類別，因此，透過該研究所提之方法，可不因虛擬社群之文化（如：字數限制）及使用者之表達習慣限制下，有效取得文章之重點主題，並可分類多文化虛擬社群中文章之情緒類別。甚之，網路商品之多樣性使得評論之標準、格式不一，導致難以建立使用者對商品及評論之情緒特徵，為能協助使用者可於非結構化且大量之評論中自動建立評論之情緒特徵並分類情緒類別，Wang 及 Jiang (2012) 提出一個基於粗糙集之情緒分析模式，首先透過粗規則將評論中具有詞袋（Bag of Words）特徵之詞句集合為評論之特徵屬性，以利用可變精度模

型 (Variable Precision Model; VPM) 建立特徵屬性與情緒詞之關聯性，並以特徵屬性與情緒詞之固定搭配過濾較無法凸顯商品特徵之屬性，以提高取得商品特徵之精確度，最後該研究藉由支援向量機整合特徵屬性與情緒詞彙之關聯，分析評論之情緒類別以可於非結構化之評論中，取得使用者對於商品之情感傾向，亦即透過該研究之方法可於不同類型之商品評論中，取得商品之特徵並可精確取得使用者對於商品真正之批判。

在最終方面，因網路評論乃由使用者自由撰寫，使得評論之內容有不同格式及寫作風格，然而，並非所有評論內容皆在描述主題（包含新聞、商品等）之特性，Liu 等人 (2010) 為了能從評論中取得真實描述主題特性之內容，提出一個以文件物件模型樹 (Document Object Model; DOM) 為基之訊息擷取模式，將網頁各區塊位置轉換為子節點樹，並以最大樹相似性演算法 (Max Tree Similarity Algorithm; MTS) 計算所有節點之相似度，再以最大序列匹配算法 (Max Sequence Similarity Algorithm; MSS) 比較所有子節點間之相似度以觀察各節點相似變化曲線，進而去除與評論無關之節點區塊，其次，利用各節點之語意特徵計算語意之標準差，以取得具有評論之子節點區塊，再將所有評論之區塊劃分為數個子節點，最後，計算評論內子節點區塊之語意標準差，將標準不一致之子節點區塊去除，以取得重點子節點區塊，並將此區塊即作為評論之重點，藉此協助使用者於不同網頁版面及寫作格式中，取得與主題較為相關之評論重點。

(B) 撰寫者用詞習慣

日文文件中常有語句結構相異但語意相似之情況，故為幫助使用者解決日文語句歧義之問題，以更精確取得重要語句，Okamoto 及 Ishizaki (2011) 乃利用聯想概念辭典 (Associative Concept Dictionary; ACD) 建立語句中形容詞、名詞及動詞等重點詞彙之語意關係，並利用上下文語意網路之概念 (Contextual Semantic Network; CSN) 分析詞彙之關聯距離並計算重要詞彙之重要分數，以藉由語句重要分數之比例提取文件之重要語句，此外，該研究語意關係乃根據上下文語意概念，自動於聯想概念辭典區分各詞彙之語意，即透過該研究可有效解決語句歧義之問題（去除語意相似之語句）。另外，網路論壇中熱門主題所引起之熱烈討論不乏含有正負面之評論，其中，當評論含有「持續破壞」類型之評論時常引起網路筆戰（以情緒字眼攻擊他人）之發生，為從評論中取得此類型評論，Ichifuji 等人 (2010) 乃利用型態分析法 (Morphological Analysis) 先行分析蓄意破壞評論之文句主詞、動詞、受詞之組合字詞及雙字詞彙，以建構蓄意破壞詞庫，並先將評論分為一般評論、無意破壞及蓄意破壞評論等，再利用貝氏過濾器 (Bayesian

Filter) 分析評論之雙字詞彙及詞性以取得蓄意破壞評論之評論，最後以一般評論之詞性、組合字詞及雙字詞等特徵比對蓄意破壞評論，藉此計算蓄意破壞評論可能為無意破壞評論之機率，以過濾蓄意破壞評論中隱含無意破壞之評論，進而更精確取得蓄意破壞之評論。透過該研究所提之方法可協助論壇使用者過濾與主題無關之評論，並可避免使用者於主題之評論中產生分歧之話題，以維持使用者對該主題之關注度。

除此之外，各領域之網路評論者皆有獨特之語意風格，導致分析評論之情感分類時常局限於單一領域中，有鑑於此，Li 等人 (2008) 利用超空間模擬語言 (Hyperspace Analogue to Language; HAL) 與訊息推理方法 (Information Inference) 整合一套多領域評論之情感分析模式。該研究乃以具有多領域特性之新聞評論作為分析對象，藉由超空間模擬語言將多領域評論中各語句之語意空間整合鏈結，以制定語句於多領域中語意之關係，並以訊息推論之概念分析評論中各語句與多領域之情感傾向性，藉此推論語句與情緒類別之後驗機率，以將所有語句之後驗機率正規化處理進而取得評論之情感傾向。透過該研究之整合方法可分析各領域語意鏈結之關係，並自動串聯各領域之情感詞彙 (即無須人工建立情緒詞彙)，即可幫助使用者於各種領域之評論中推論情感傾向。甚者，各式虛擬社群所制訂之評論格式多數不盡相同，導致無法明確定義標題、主題及語句內容與情緒間之語意關係，以致無法同時分析各式評論之情緒傾向性，為能幫助使用者針對不同格式之評論分析情緒傾向，Wang 等人 (2010) 提出一個以本體論 (Ontology) 為基之情緒分類方法，該研究首先藉由知網所提供之正面、反面及程度詞彙 (影響情緒語氣之詞彙) 建構情緒詞彙庫，並透過本體論之特性選取情緒詞彙庫與特性領域最為相關之情緒詞彙 (分別取得正、反及中性之情緒詞)，以過濾與該特定領域不相關之情緒詞，其次，該研究利用向量空間模型計算評論中所有詞彙與情緒詞彙之相似程度，以針對該特定領域建構詞彙與情緒之關係係數，最後，該研究藉由支援向量機分析評論中所有詞彙與情緒詞彙、程度詞彙情緒傾向權重，以取得該評論於特定領域之情緒傾向性，進而協助使用者於複雜且多樣之評論格式中取得情感傾向。

甚之，有鑑於中國新浪微博中訊息簡短且所發佈之格式不一致等特性，以致於分類訊息情緒傾向之成效不佳，是故，Xie 等人 (2012) 乃提出三種文章情緒分類方法，首先，計算微博文章特有之情緒表情符號數，依據正、中及負向符號數推論文章之情緒類別，其次，透過三類別情緒詞彙之詞頻數，由詞頻數最多者作為文章之情緒類別，最後，該研究乃將訊息先分為整體訊息及分句訊息，並於整體訊息中擷取主、客觀特徵及情緒詞彙，再由分句訊息中語句之情緒詞彙數作為訓練資料，並同時加入主題詞與訊息內詞

彙頻率權重值，以藉由支援向量機分析訊息之情緒傾向，此外，該研究亦同時比較三種分類方法之成效，則以支援向量機整合訊息特徵分類方法之成效最佳。透過該研究之方法，使用者可有效於簡短且非結構化之訊息中取得情緒傾向，亦可依據訊息之格式於三種分類法中選擇分類成效最佳之方法。且目前烏爾都語中尚無較完整且標註情緒傾向明確之詞彙庫，雖可利用英文情緒詞彙庫相互對照以對應情緒詞之情緒傾向值，但因烏爾都語中存有修飾詞與特殊形容詞之關係，導致對照後之成效不準確，為協助使用者建立烏爾都語之情緒詞彙庫，Afraz等人(2010)首先利用訓練網頁文件定義特殊形容詞與修飾詞之屬性，將表達數量、質量之形容詞及前後呼應之修飾詞加以標記，以明確定義形容詞與修飾詞間之關係，接著以烏爾都語之型態規則定義所有情緒詞彙之語法型態，以分析所有詞彙與情緒詞於不同時間及動作時之解釋，並利用同義與反義詞與情緒詞之組合進一步延伸詞彙之語意解釋，最後，該研究以人工方式與詞頻關係，將情緒詞與所有語法型態、同義及反義詞之組合標記情緒傾向值（正負面兩情緒類別），並依據形容詞與修飾詞之關係調整情緒詞之情緒傾向值，以更精確取得所有情緒詞彙之情緒傾向值。藉由該研究之方法，使用者可依據各種語法定義規則及詞頻關係，有效於特殊語法規則及詞彙之語言中，建立情緒詞彙之情緒傾向值。

最後，因網路產品評論所使用之語法及習慣用語多數不同，使得情緒詞彙庫多數僅適用於單一領域中，是故，為了協助使用者依據特定領域中評論用語之特性以建立情緒詞彙庫，Hamouda等人(2011)乃訂定收集訓練資料之規則以建構一套基於機器學習法之情緒詞彙庫(Machine Learning Based Senti-word Lexicon; MLBSL)，該研究乃以亞馬遜(Amazon)論壇之評論作為訓練資料，先行標記評論之情緒代表詞，並以GATE(General Architecture for Text Engineering)自然語言處理工具，標記所有詞性中詞彙之延伸定義(如「跑」與「跑步」、「運行」相關)，以藉由詞頻-逆向文件頻率法給予所有詞彙及延伸詞彙定義之分群權重，最後，透過支援向量機計算所有詞彙之分群向量，以取得所有詞彙之情緒傾向值，進而幫助使用者取得特定領域之情緒詞彙庫，並可以此情緒詞彙庫為基分析評論之情緒類別。

綜合上述，針對文章撰寫者寫作習慣分析中發現大多研究對於非結構化且複雜之用詞，乃以潛在語意分析語意結構，甚之，透過特殊語意規則針對不同之寫作風格分析語意內容，以不同方法分析使用者寫作之習慣。而本研究藉由此議題之探討得知，各虛擬社群之文章格式與寫作習慣皆不相同，且中文含有大量之歧義句及特殊之語意結構，以致於過去研究大多乃針對特定社群分析文章之內容及語意架構，因此，本研究即發展一

個適用於一般虛擬社群之文章分析方法，並可不受撰寫者之用詞與語法結構之影響，而使得應用範圍有所限制。於此議題相關文獻之分析類型與過程分別彙整成表 2.6。

表 2.6、文章撰寫者寫作習慣分析文獻彙整表

文獻議題	過去研究	研究目的	分析類型	分析過程
撰寫者寫作風格	Liu 及 Liu (2012)	以機器學習方法自動產生語意規則，以針對評論進行情緒分類	多樣性、特有語意風格之評論	以支援向量機、貝氏及 N 連詞三種分類方法進行分類
	Bai (2011)	於字數較少之評論中取得使用者欲表達之情緒類別	數字、字數不定且非結構化之評論	利用 LrnMBC 演算法偵測詞彙與重要語句間之情緒變化，以評論之情緒類別
	Fu 等人 (2013)	同時分析不同文化社群中，文章主題欲表達之情緒類別	情緒表達方式不一之文章	藉由重點主題所含詞彙與情緒詞分析相似性，以取得重點主題之情緒傾向值
	Wang 及 Jiang (2012)	協助使用者於非結構化且大量之評論中自動分類情緒類別	多樣性、格式不一之評論	藉由支援向量機整合特徵屬性與情緒詞彙之關聯，分析評論之情緒類別
	Liu 等人 (2010)	從評論中取得真實描述主題特性之內容	使用者自由撰寫之評論	分析評論內子節點區塊之語意標準差，將不一致之區塊去除以取得評論之重點
撰寫者用詞習慣	Okamoto 及 Ishizaki (2011)	於語句結構相異但語意相似之文章中，取得重要語句	歧義之語句	利用上下文語意概念，自動於聯想概念辭典區分詞彙之語意，以分析語句之涵義
	Ichifuji 等人 (2010)	協助論壇使用者過濾與主題無關之評論	持續破壞之評論	利用貝氏過濾器分析評論之雙字詞彙及詞性以取得蓄意破壞評論之評論
	Li 等人 (2008)	幫助使用者於各種領域之評論中推論情感傾向	獨特之語意風格	將所有語句之後驗機率正規化處理以取得評論之情感傾向
	Wang 等人 (2010)	協助使用者於複雜多樣之評論格式中取得情感傾向	格式不相同之評論	藉由支援向量機分析評論中所有詞彙與情緒詞彙、程度詞彙情緒傾向權重
	Xie 等人 (2012)	幫助使用者可有效於簡短且非結構化之訊息中取得情緒傾向	簡短且發佈格式不一致之訊息	計算微博文章特有之情緒表情符號數，依據正、中及負向符號數推論文章之情緒類別
	Afraz 等人 (2010)	協助使用者建立烏爾都語之情緒詞彙庫	不同情緒係數之詞彙	透過支援向量機計算所有詞彙之分群向量，以取得所有詞彙之情緒傾向值
	Hamouda 等人 (2011)	協助使用者依據特定領域中評論用語之特性以建立情緒詞彙庫	用語習慣及語法不相同之評論	透過支援向量機計算所有詞彙之分群向量，以取得所有詞彙之情緒傾向值

2.4.2 文章閱讀者閱讀感受分析

於文章閱讀者閱讀感受分析議題中，本研究乃針對「文章篇幅之於閱讀感受」以及「圖表之於閱讀感受」等兩主題進行相關文獻探討，期望從中探討文章閱讀者之閱讀習

慣與感受。

(A) 文章篇幅之於閱讀感受

文章閱讀者閱讀網路文章時，常以首句詞語判斷主題與文章內容之相符程度，但文章閱讀者之生長環境與擅長領域不同，其解讀語句之涵義將與文章撰寫者原先表達之涵義有所差異，為了驗證此差異性之現象，David 等人 (2011) 針對 17 位測試者以實驗法方式測試使用者對於語句之認知。該研究之受測者皆屬於正常成人、平均年齡為 25 歲且皆為右撇子，當受測者於文章中看見第一個語句時，該研究以腦磁圖追蹤受測者大腦瞬間判斷此語句與主題相關與否之反應，並提供「相關」、「不相關」之按鈕供受測者判斷，以藉此觀察受測者大腦判斷時間、腦物質增益性等數據，以及受測者認知語句與主題相關程度，進而加以分析認知數據。該研究結果顯示，不同受測者對於第一個語句之認知涵義將有所差異，但若制定規範或結構化之方式呈現文章時，受測者對於語句之認知差異性會有明顯之縮小。甚者，多文件摘要技術大多乃取目標文件之重要部分，再藉由其他文件之語句加以排序以形成摘要，然而，此方法常有語句不通順、不連貫或不符合事件（即兩連續語句所指之事件不同）之問題，是故，Danushka 等人 (2012) 乃針對新聞文件發佈時間定義語句之順序權重，接著比較新聞事件與相似事件之新聞文件，從中分析兩相似新聞中起頭、結論與過程等語句寫法，以給予目標新聞文件所有語句之順序機率值，其次，以動詞與專有名詞作為局部主題，將語句分類至對應之主題中，並根據主題與事件之關係及語句順序權重分析所有語句之優先順序，最後，利用語句排序演算法 (Sentence Ordering Algorithm) 計算所有語句與事件之關係函數，從中取得最合適之語句順序，以幫助使用者取得語句通順且與目標新聞相關之語句組合。

此外，目前電子新聞彙整機制乃以新聞標題與關鍵字方式呈現，但此方式缺乏具體主題以描述事件起始與重點等，待時間過去後閱讀者較難以追溯該新聞之事件。是故，Lin 及 Liang (2008) 乃提出將事件主軸納入摘要建立條件之機制 (Story-line based Topic Retrospection, SToRe)，透過事件主軸能使讀者更了解事件之發展與概念。該研究乃先界定事件再由事件建立議題，透過議題發展延伸脈絡並去除較無相關性之事件，從中再擷取代表性語句並以議題主軸為依據形成摘要，此外，該機制透過議題延伸發展為一個新的議題時，將可作為索引以提供閱讀者作為搜尋依據。透過該研究所提之機制，可更精確取得新聞之重點，同時藉由摘要之彙整提高閱讀者閱讀之效率，並透過議題之索引方式加強閱讀者搜索新聞之效率。其次，有鑑於網路新聞所涉及之主題包含人、事件、

物質等屬性，且此些屬性之新聞評論者因領域之不同，導致內容之風格及造詞特色有所差異，以致於難以分類新聞評論之情感類別，因此，Zhou 等人 (2010) 提出一個機器學習之情感分類方法，先將評論進行去除重複評論（同一評論者短時間再發佈評論）、規範格式（字母大小寫、簡繁轉換等）、情感詞標記等作業以形成訓練評論庫，之後利用空間向量模型將評論內容根據情緒語料庫之特徵形成文章之特徵向量，並以支援向量機計算各特徵向量之情緒機率值，再加入詞頻-逆向文件頻率法及布爾型（Boolean）兩種權重計算方法並比較分類成效，其結果顯示，以詞頻-逆向文件頻率法作為權重計算時，分類效果將優於布爾型。該研究之支援向量機與詞頻-逆向文件頻率法整合分類方法，可便於使用者於相異領域之新聞中，有效取得評論之情感類別，並提供使用者一個權重計算之參考建議。

另一方面，醫療人員依據經驗所撰寫之醫療文件常有篇幅過長與專業術語過多之情況，為協助使用者快速取得醫療文件之重點，Zhou 等人 (2006) 基於自然語言處理提出一套自動化擷取醫療術語模式，藉由專門於醫療領域之自然處理語言（Medical Language Extraction and Encoding, MedLEE）建立語意結構之語料庫，並經由圖形介面工具建立視覺化摘要以產生醫療文件之概念與主題，同時透過主題樹即可得知專業術語之語意類型並擷取作為陳述句子，因此，利用主題概念與專業術語之陳述將可形成醫療文件之重點，以提供使用者參考。透過該研究所提之方法將可改善人工建立醫療術語方式，透過自然語言處理將文件摘要視覺化呈現亦可減輕開發人員之負擔，同時藉由語料庫資料之增長，將使文件產生之重點更為完整。

(B) 圖表之於閱讀感受

使用者對於網路圖片之訊息來源，多數與圖像特徵（紋理、顏色）無相關之文字，亦即圖片之文字描述多數乃針對圖片中人、物之外觀或輪廓，是故，為了探討使用者以知覺描述圖片特徵之情況，Rorissa (2008) 乃以圖像索引個人圖像及圖像檢索群體圖像為分析之主軸。於此前，該研究認為使用者比較個人與群體標籤之認知模式較為有限，因此，該研究使用基層理論為整體框架，以 180 名測試者比較圖像說明及標籤分配，並製成三項層次（即上層、中層、下層）之圖表。該研究分析結果顯示，群體圖象（如汽車、工作等具體名稱）落點為中層之上，然而，個人圖像描述（即傳達表面性質有關之形像，參考紋理或色彩）其主要落點於下層。藉由該研究之探討可了解使用者瀏覽介面圖片之描述習慣，進而提供決策者針對圖像之特性給予使用者不同之描述工具。甚之，

Oh 等人 (2008) 乃探討網路店面呈現方式對於商品形象之影響，並同時探討消費者瀏覽網頁後對於商品之期望程度。該研究利用刺激有機體反應理論 (Stimulus Organism Response; S-O-R) 設計兩種網頁介面類型 (主題與非主題式) 與兩種訊息呈現方式 (文字與圖片)，以針對 307 大學生提供實測四種組合設計並發放問卷，以從中了解消費者最傾向之組合，透過問卷分析結果顯示，以主題式介面與圖片呈現方式乃消費者最傾向之方式，亦對商品之形象具有良好之反應 (即正面評價居多)，同時透過圖片瀏覽之方式將可提升消費者對於商品之期待性，因此，藉由該研究分析之結果，可提供網路賣家一個呈現網頁之參考方法，亦可透過圖片之呈現提升商品形象，以增加消費者購買意願。

且由於目前之圖像檢索技術仍未成熟，導致使用者對於欲搜尋圖片需求至今仍無法滿足，因此，Hollink 等人 (2004) 為獲取使用者搜尋圖像時所描述之訊息，制訂一圖像描述之分類模式，該模式以圖像中三項觀點，包含非視覺化之數據、圖像之概念及感性之概念描述作為分類依據，再以 30 名受測者對於圖片之顏色、形狀等五大圖片特徵分別描述與指定其關係，最後再加以分析統整，其分析結果顯示，使用者較趨向於一般描述 (圖中事件與對象之關係) 而並非以具體、抽象之概念分類圖像。透過該研究可了解使用者對於圖片之描述概念，以提供圖片搜尋開發者釐清圖片之關鍵字定義及語句之描述。其次，使用者瞬間閱讀網頁之首次印象乃影響網頁引人注目之關鍵，是故，為了探討不同網頁類型對於使用者之吸引程度，Tractinsky 等人 (2006) 乃以兩項對比實驗並針對 40 位大學生 (對於網頁設計較無概念之群體學生) 進行實測，第一項實驗中，該研究乃提供受測者 25 個不知名且內容較具有視覺效果 (具有多種色彩背景、圖片) 之網站，每個網頁供受測者 5 秒觀看時間以分析該類型網站之吸引力，於第二項實驗中則以 25 個以文字、正規排版且不知名之網站供受測者測試，亦以 5 秒之觀看時間分析吸引數據，最後，兩項實驗之分析結果顯示，兩項類型之網站皆有達到吸引之效果，但具有視覺效果類型之網站，其吸引程度較高於以文字與正規排版類型之網站，是故，透過該研究之實驗結果，將可提供網站設計者一個設計概念，以針對特定群體之使用者呈現最合適之網站類型。

最後，圖片之人、事、物等具體之意象將可提供知識需求者作為關鍵字搜尋之條件，然而現今大量圖片存於網路中，使得圖片之定義、註解及描述不完全，導致知識需求者利用圖片意象所建構之關鍵字進行搜索時之效率低下，有鑑於此，Liu 等人 (2008) 提出一項自動影像註解架構，該研究利用三種關係加強及改善關鍵字之圖像檢索。首先，針對圖片三種關鍵字串聯所有關鍵字，以加強關鍵字聯想之延展性，接著，將關鍵字分

解重組並針對字詞之語意架構加以標註，最後，以人工方式針對圖片間意象建立相似關係，以將相似之圖片歸於一類，是故，藉由該研究漸進式之分析方法，可協助使用者欲檢索圖片且註記不清時，能提升檢索圖片之情況，並可有效取得所需之圖片。

綜合上述，針對文章閱讀者閱讀感受分析中許多研究證實文章篇幅將影響閱讀者之閱讀效率，且閱讀者對於文章之具體意象不同時亦會影響閱讀之感受。本研究藉由此議題文獻之探討得知文章之流暢與連貫性將會影響閱讀者之閱讀效率，因此本研究即針對語句之結構進行分析並修改，期望提升語句之流暢性以提升閱讀者之閱讀效率。於此部分之文獻彙整如表 2.7。

表 2.7、文章閱讀者閱讀感受分析文獻彙整表

文獻議題	過去研究	分析方法	閱讀影響
文章篇幅之於閱讀感受	David 等人 (2011)	以實驗法分析 17 位測試者之腦物質數據，以測試使用者對於語句之認知	制定規範或結構化之方式呈現文章時，受測者對於語句之認知差異性會有明顯之縮小
	Danushka 等人 (2012)	分析兩相似新聞中起頭、結論與過程等語句寫法，以重組目標新聞文件所有語句之順序	多文件摘要形成新聞重點，對使用者常有語句不通順、不連貫之閱讀影響
	Lin 及 Liang (2008)	透過議題發展延伸脈絡並去除較無相關性之事件，從中再擷取代表性語句並以議題主軸為依據形成摘要	電子新聞彙整待時間過去後閱讀者較難以追溯該新聞之事件
	Zhou 等人 (2010)	整合支援向量機與詞頻-逆向文件頻率法分類方法，取得相異領域評論之情感類別	內容之風格及造詞特色使使用者難以分辨評論之情感傾向
	Zhou 等人 (2006)	透過主題樹得知專業術語之語意類型並擷取作為陳述句子以形成醫療文件之重點	醫療文件篇幅過長與專業術語過多，閱讀者難以取得文件重點
圖表之於閱讀感受	Rorissa (2008)	透過 180 名測試者比較圖像說明及標籤分配，以了解使用者瀏覽介面圖片之描述習慣	使用者知覺描述圖片特徵之情況
	Oh 等人 (2008)	針對 307 大學生提供實測四種組合設計並發放問卷，結果得知以主題式介面與圖片呈現方式乃消費者最傾向之方式	網路店面呈現方式對消費者瀏覽之影響
	Hollink 等人 (2004)	制訂一圖像描述之分類模式，以 30 名受測者對於圖片之顏色、形狀等五大圖片特徵分別描述與指定其關係	使用者搜尋圖像時，對於圖片所描述之特徵、訊息
	Tractinsky 等人 (2006)	以兩項對比實驗並針對 40 位大學生進行實測，以比較視覺效果及正規排版網站之吸引力	使用者瞬間閱讀網頁之首次印象
	Liu 等人 (2008)	以漸進式之分析方法，協助使用者欲檢索圖片且註記不清時，提升檢索圖片之情況，並可有效取得所需圖片	知識需求者利用圖片意象所建構之關鍵字進行搜索時之效率不佳

2.5 小結

本研究所建立「以論壇文章蘊含情緒為基之文章重組模式」，主要目的乃協助論壇管理者快速取得違規之文章，並針對違規但可增進社群使用者持續討論之文章重組其內容。本研究乃針對「虛擬社群之管理機制探討」、「社群文章資料探勘」與「文章撰寫者行為探討」等三大議題進行比較，並歸納本研究與過往研究之差異，如表2.8所示。

於2.2節「虛擬社群管理機制探討」議題中得知，過去研究主要著重於協助使用者快速取得正確之文章，同時透過文章審核機制過濾違規之文章，藉此管理及控管虛擬社群整體之品質。但過去之研究對於違規文章之審核多數僅能以文字比對技術，將出現違規字詞之文章移除，故本研究除針對違規字詞外，亦分析文章撰寫者撰寫時之情緒感受，藉此審核帶有偏激或惡意情緒之文章，進而協助管理者於過量之文章取得違規之文章。

而於2.3節「社群文章資料探勘」之相關文獻中，可發現過去研究針對文章語句重組技術大多藉由語意規則及語句相似特徵重組語句之架構，而對於文章評級分析方面，過去多數研究乃透過相似詞彙比對並藉由文章分類法分析文章之評級類別，以分析文章語句之連貫程度。故本研究乃藉由相異詞語分析法及語句相似特徵之整合，建構「發文者文章語句重組模組」針對違規之文章重組文章之內容架構。

最後於2.4節「文章撰寫者行為探討中」之相關文獻中，乃針對使用者撰寫習性與閱讀習慣進行探討，並從文章撰寫者情緒分類議題中，可得知以文章主題、語句、詞彙或語意關係可作為情緒之主要分析因素，故本研究乃以此為依據建構「文章表達情緒判定模組」以分析文章撰寫者之情緒類別，從中取得惡意情緒之文章。

綜上所述，本研究所建立「以論壇文章蘊含情緒為基之文章重組模式」乃透過相似語句分析文章內容所含之情緒特徵，並透過文章之情緒特徵區分文章所屬情緒，此外，本研究亦針對帶有惡意情緒之文章，先以相異詞語分析文章之語句連貫程度，藉此分析文章重組之合適性，再藉由相似語句延伸之多重組合句重組文章之架構，最終將可協助論壇管理者快速取得違規之文章並改寫其內容，以協助社群保留具有討論價值之文章。

表 2.8、本研究與過去文獻差異彙整表

比較研究	分析對象	研究類型	應用範圍	文章分類	文章結構	語意解析	發展新技術	使用者行為分析	使用者感受	使用者觀點調查	優勢
本研究	網頁文章	演繹法	一般領域	✓	✓	✓		✓	✓		協助社群取得違規文章，並自動改寫文章內容，以保留有意義之文章。
Peng等人(2008)、	搜尋語句	演繹法	特定領域		✓	✓	✓				幫助使用者串聯搜尋語句與各領域專業詞彙之關聯性
Jung等人(2005)	網頁文章	演繹法	一般領域		✓	✓		✓			無須人工訓練資料，即可自動串聯問題與知識之語意關係
Rorissa(2008)、 Hollink等人(2004)	視覺受測者	歸納法	一般使用者				✓	✓	✓	✓	以基層理論製成三項層次，以探討使用者對於圖像之描述
Oh等人(2008)	大學生	歸納法	特定使用者		✓				✓	✓	設計四種組合問卷，探討消費者瀏覽網頁後對於商品之期望程度
Li等人(2006)	搜尋語句	演繹法	一般領域		✓	✓	✓	✓			以簡短之語句取得語意相似之語句
Kaila等人(2006)	網頁文章	演繹法	一般領域	✓	✓			✓			以三項功能平台過濾知識
Wang等人(2008)	網頁文章	演繹法	一般領域	✓	✓	✓	✓	✓			將知識庫依功能級別分類，並分類知識分享者
Zhang(2009)	過去文獻	歸納法	一般使用者					✓	✓	✓	將知識共享因素分為三個研究導向並彙整
Fang等人(2010)	論壇成員	歸納法	特定使用者					✓	✓	✓	分析成員持續與參與共享知識之因素
Li及Wu(2010)	網頁文件	演繹法	一般領域	✓	✓		✓		✓		以樹狀分類結構圖取得熱門主題
Teo等人(2003)	學生	歸納法	特定使用者				✓	✓		✓	以科技接受模型為基加入兩項分析因素
Wang與Noe(2010)	過去文獻	歸納法	特定使用者	✓			✓			✓	整合定性與定量相關研究

Gu 與 Grossman (2010)	知識流	演繹法	一般領域	✓	✓			✓			可於大量之知識流中彙整與過濾知識之專有領域
Matsubara 等人(2006)	知識檔案	演繹法	一般領域	✓	✓		✓	✓			以中央管理器控制虛擬目錄之知識檔案
Ercan及Cicekli(2007)	知識文章	演繹法	一般領域		✓	✓	✓				以詞彙鏈方式取得關鍵字
Li 等人 (2012)	網頁文章	演繹法	特定領域	✓	✓		✓	✓			推薦合適之知識專家
Yu 等人 (2006)	知識文章	演繹法	一般領域		✓	✓	✓	✓			以潛在語意分析，回饋更精確之答案
Li 等人 (2011)	網頁文章	演繹法	特定領域	✓	✓	✓	✓				取得真正用以描述產品觀點之屬性詞
Jonathan及Yiu (2006)	網頁文章	演繹法	一般領域		✓	✓		✓			結合公制群集分析法與模糊集檢索法取得相似文件
Lai (2007)	知識文件	演繹法	一般領域	✓	✓	✓	✓	✓			協助知識管理者管理龐大且變動頻繁之知識
Pérez 等人 (2007)	知識文件	演繹法	一般領域	✓	✓		✓		✓		分散計算資料存取之過程
Pang (2010)、 Li 等人 (2008)	網路評論	演繹法	一般領域	✓	✓	✓		✓	✓		快速取得商品之最新評價
Chan (2006)	網頁文章	演繹法	一般領域		✓	✓			✓		生成文章之主旨
Fan 等人 (2009)、 Zhang 等人 (2009)	網路評論	演繹法	特定領域			✓	✓	✓	✓		得知評論之重點主題
Ko 與 Seo (2008)	網頁文章	演繹法	一般領域		✓	✓	✓				多文章摘要建立
Xu 等人 (2008)	網頁文章	演繹法	特定領域		✓	✓	✓				以階層式聚合分群法形成摘要
Win 及 Zhang (2006)	醫療文件	演繹法	特定領域		✓	✓	✓	✓			以簡短搜尋之字串取得領域知識
Huang 等人 (2012)	文章語句	演繹法	特定領域	✓			✓	✓	✓		於結構複雜語言中分析詞彙之情緒類別
Hao 等人 (2007)	知識文件	演繹法	一般領域		✓	✓	✓				將詞彙與語句、歧義詞、問題類型之關係建立鏈結
Kakkonen 等人(2008)	教師文章	演繹法	特定領域	✓	✓	✓		✓			比較三個文章評級分析方法之成效
Yin 及 Peng (2012)	網路評論	演繹法	特定領域	✓	✓		✓	✓	✓		以關係網路建立情緒詞庫

Peng 等人 (2007)	網路訊息	演繹法	特定領域	✓	✓	✓		✓			分類非結構化之網路訊息
Lee 等人 (2011)	英文文章	演繹法	特定領域		✓	✓		✓			建立語法依賴樹以得知文法之正確性
Ge 及 Chen (2009)	一般作文	演繹法	一般領域	✓	✓	✓		✓			自動取得偏離題目之作文
Zhang 等人 (2011)	文章語句	演繹法	一般領域	✓		✓	✓		✓		以結合樹提升分類準確率
Jiang 等人 (2011)	大學作文	演繹法	特定領域	✓	✓	✓		✓			將相似語法程度之文章分為一類
Huang 等人 (2009)	英文文章	演繹法	一般領域	✓	✓	✓		✓			依寫作特性放置合適之分類器中分類文章
Wang 等人 (2011)、 Liu 等人 (2009)、 Zhan 等人 (2011)	網頁文章	演繹法	特定領域	✓		✓		✓	✓		建立特定領域之情緒詞庫
Liao 等人 (2012)	國小作文	演繹法	特定領域	✓		✓		✓			根據整體寫作程度建立評級標準
Yang 等人 (2011)	網路評論	演繹法	特定領域	✓		✓		✓	✓		以主題句取得評論之情緒類別
Yang 等人 (2010)	網路評論	演繹法	特定領域	✓	✓			✓	✓		隨著訓練評論量之增加而提高情感分類之準確率
Wang 及 Li (2010)、 Matsumoto 等人 (2011)、 Bai (2011)	網頁文章	演繹法	特定領域	✓	✓	✓			✓		取得非結構化文章之情緒類別
Wang 等人 (2009)	情緒詞彙	演繹法	一般領域	✓		✓	✓				以最小粒度之層級分析詞彙之情感傾向
Zan 等人 (2011)	網頁文章	演繹法	一般領域	✓		✓	✓		✓		結合語意規則與統計方法分析文件之情緒類別
Lin 等人 (2011)	文章語句	演繹法	一般領域	✓	✓				✓		以非監督學習方式分析語句情緒傾向
Das 及 Bandyopadhyay (2010)	網頁文章	演繹法	特定領域	✓		✓	✓		✓		針對孟加拉語分析文章情緒類別
Zhang 等人 (2009)	網頁文章	演繹法	特定領域	✓	✓	✓	✓		✓		自動建立語意規則
Li 及 Liu (2012)	網頁文章	演繹法	一般領域	✓			✓	✓	✓		於無訓練資料之情況下，自動判定文章之情感偏向

Liu 及 Liu (2012)	網路評論	演繹法	特定領域	✓	✓			✓	✓		比較過去研究常用三種情緒分類方法之成效
Fu 等人 (2013)	網頁文章	演繹法	一般領域	✓				✓	✓	✓	取得文章之重點主題
Wang 及 Jiang (2012)	網路評論	演繹法	一般領域		✓	✓		✓		✓	於同類型之商品中取得商品之特徵
Liu 等人 (2010)	網路評論	演繹法	一般領域	✓	✓	✓	✓	✓		✓	取得與主題較為相關之評論重點
Okamoto 及 Ishizaki (2011)	日文文章	演繹法	一般領域	✓	✓				✓		解決語句歧義之問題
Ichifuji 等人 (2010)	網路評論	演繹法	一般領域	✓	✓	✓		✓	✓		過濾與主題無關之評論
Li 等人 (2008)、Wang 等人 (2010)、Xie 等人 (2012)	網路評論 網路訊息	演繹法	一般領域	✓	✓		✓		✓		無須人工建立情緒詞彙，即可取得情緒傾向
Afraz 等人 (2010)	烏爾都語文件	演繹法	一般領域	✓	✓		✓				建立專屬烏爾都語之情緒詞彙庫
Hamouda 等人 (2011)	網路評論	演繹法	一般領域	✓	✓		✓	✓			以分群概念分析情緒
David 等人 (2011)	正常成人	歸納法	一般使用者		✓				✓	✓	測試使用者對語句之認知
Danushka 等人 (2012)	新聞文件	演繹法	特定領域		✓		✓	✓			取得目標新聞相關之語句
Lin 及 Liang (2008)	新聞文件	演繹法	特定領域		✓	✓		✓		✓	加強閱讀者搜索新聞效率
Zhou 等人 (2010)	網路評論	演繹法	特定領域	✓	✓				✓		依造詞特色取得評論之情緒類別
Zhou 等人 (2006)	醫療文件	演繹法	特定領域		✓	✓	✓		✓	✓	改善人工建立醫療術語方式形成醫療文件之重點
Tractinsky 等人 (2006)	大學生	歸納法	特定使用者		✓				✓	✓	探討不同網頁類型對於使用者之吸引程度
Liu 等人 (2008)	圖片 關鍵字	演繹法	一般領域		✓	✓	✓		✓		自動影像註解架構加強圖像檢索效率

第三章、以論壇文章蘊含情緒為基之文章重組模式

本研究所提之「以論壇文章蘊含情緒為基之文章重組模式」乃以虛擬論壇之論壇文章為分析對象，先行經中文詞知識小組（Chinese Knowledge Information Processing Group；CKIP）之中文斷詞系統斷詞後，從中取得論壇文章之候選事件，以利用候選事件之詞性、前後文及標題相關性分數分析論壇文章之代表事件，並藉由代表事件之語句比對訓練文章之語句相似程度，分析論壇文章之情緒類別，進而使後續發文者文章語句重組模組重組內文之架構後，可與原先文章撰寫者欲表達之情感一致。透過論壇文章評閱分數取得合適改寫之文章，並去除語意相似及情緒與原文不一致之語句，分析論壇文章所需之候選填充句與組合句，並進行多重組合以重組論壇文章架構，最終形成一篇語意完整之論壇文章內容。本研究之主要流程可分為兩大部份，分別為如圖 3.1 之 Part1「文章表達情緒判定模組」與 Part2「發文者文章語句重組模組」所示。

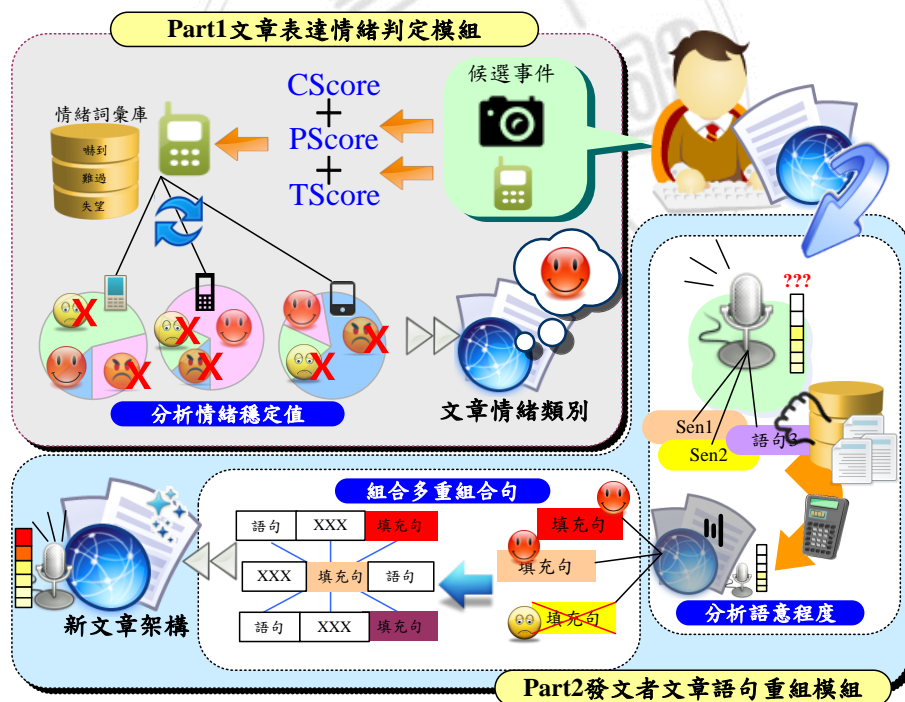


圖 3.1、以論壇文章蘊含情緒為基之文章重組模式之流程架構

3.1 文章表達情緒判定模組

根據第一章研究動機與目的之期望模式，於「文章表達情緒判定模組」之目的為自動推論文章之情緒類別，於最終推論情緒類別之步驟乃借鑒並改良 Zhao 等人（2010）

之方法，該研究主要以非結構化之文章為分析對象，因此將可適用本模組中作為最終推論方法，而推論情緒類別方法時則須以文章之主題事件為基礎進行推論，因此本模組乃借鑑 Tung 與 Lu (2010) 之方法規劃至步驟(A1)至步驟(A3)中，以此三步驟取得文章之主題事件，此外，本研究情緒類別推論乃以情緒詞彙為基礎，並考量過去研究大多以人工方式定義情緒詞彙之隸屬係數，因此乃借鑒與改良 Miao 等人 (2012) 之方法規劃至步驟(A4)中，以於推論文章情緒類別前可先行自動推論情緒詞彙之隸屬係數，完成上述步驟後，即將 Zhao 等人 (2010) 改良後之方法規劃至步驟(A5)至步驟(A7)中。

本模組乃以論壇發文者文章為分析之基礎，以判定論壇文章所表達之情緒類別，進而作為後續發文者文章語句重組模組替換論壇文章情緒字詞之用。首先乃藉由「知網」(HowNet) 所含 2090 個正負面情緒字詞以人工方式分類情緒類別，以建構情緒類別詞彙庫，但「知網」所提供之情緒詞字體為簡體中文，有鑑於簡體中文與正體(繁體)中文之方言差異，故依據教育部國語、成語字典篩選具方言之情緒字詞(如表 3.1 範例所示)，以保持情緒類別詞彙庫之語意一致性；其次，利用中文詞知識小組 (Chinese Knowledge Information Processing Group; CKIP) 之中文斷詞系統挑選論壇文章候選事件，其中候選事件多數可視為論壇文章之代表詞彙，並依據 Tung 與 Lu (2010) 之方法利用論壇文章候選事件之詞性、前後文配對及標題分數計算文章各候選事件之權重分數，以取得論壇文章代表事件，並參考 Zhao 等人 (2010) 以向量空間模型 (Vector Space Model; VSM) 之餘弦函數 (Cosine) 方式計算目標論壇文章代表語句與所有訓練文章之語句相似程度後，再以帕茲模型 (Potts Model) 取得目標論壇文章代表語句所表達之情緒類別，以作為目標論壇文章表達情緒類別之依據(如圖 3.2 所示)。本模組所使用之符號定義如下：

表 3.1、方言情緒字詞之比較

情緒詞類別	簡體字體	正體字體	是否為方言	教育部國語辭典之詞義	地方方言之詞義
正面情緒	留个心眼儿	留個心眼兒	是	-	「小心」之意
	不离儿	不離兒	是	-	「差不多」之意
	敞开儿	敞開兒	否	恣意、儘量	-
	把稳	把穩	否	主意堅定，不可動搖	-
負面情緒	诧异	詫愕	是	-	「驚奇」之意
	堵得慌	堵得慌	否	「悶得慌」之意	-
	发怔	發怔	否	因心神不貫注而眼睛呆視的樣子	-
	忿恚	忿恚	否	怨恨發怒	-

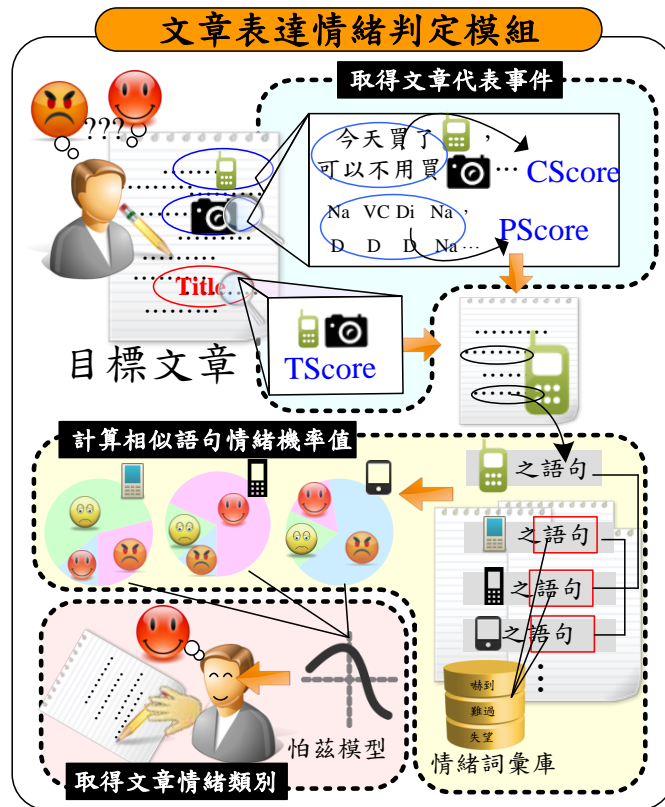


圖 3.2、文章表達情緒判定示意圖

符號定義

A_T	目標論壇文章 (已經斷詞處理之文章)
ACE_i	目標論壇文章中第 i 個候選事件 (即從文章中挑選具專有名詞或地方詞之代表詞彙)
$Fre(ACE_i)$	目標論壇文章中第 i 個候選事件於文章中出現頻率
$N(ACE_i)$	目標論壇文章中候選事件總個數
CES_i^F	第 i 個候選事件前一個語句 (即標點符號與標點符號間之語句)
CES_i^B	第 i 個候選事件後一個語句
$CES_{i,j}^F$	第 i 個候選事件前一個語句 CES_i^F 中第 j 個觸發詞 (即觸發事件之詞彙, 該詞彙乃語句經斷詞處理後依據前後語意關係擷取語意較合適之詞彙)
$CES_{i,k}^B$	第 i 個候選事件後一個語句 CES_i^B 中第 k 個觸發詞
$N(CES_i^F)$	第 i 個候選事件前一個語句 CES_i^F 之觸發詞個數
$N(CES_i^B)$	第 i 個候選事件後一個語句 CES_i^B 之觸發詞個數
$N(CES^F)$	所有候選事件前一個語句之觸發詞總數

$N(\text{CES}_i^B)$	所有候選事件後一個語句之觸發詞總數
$\text{Fre}(\text{CES}_{i,j}^F)$	第 i 個候選事件前一個語句 CES_i^F 中第 j 個觸發詞於語句 CES_i^F 中出現次數
$\text{Fre}(\text{CES}_{i,k}^B)$	第 i 個候選事件後一個語句 CES_i^B 中第 k 個觸發詞於語句 CES_i^B 中出現次數
$N(\text{CES}_{i,j}^F)$	第 i 個候選事件前一個語句 CES_i^F 中第 j 個觸發詞於文章中出現次數
$N(\text{CES}_{i,k}^B)$	第 i 個候選事件後一個語句 CES_i^B 中第 k 個觸發詞於文章中出現次數
$\text{Fre}(\text{ACE}_i, \text{Title})$	第 i 個候選事件於標題詞彙中出現次數
POS_m	訓練文章中所有候選事件所含之第 m 種詞性種類
$\text{Sdis}(\text{POS}_\cdot)$	詞性距離權重值, 範圍介於 1 與訓練文章候選事件所含詞性種類之自然數
$\text{VEF}_{i,n}$	第 i 個候選事件向前推算之第 n 個詞彙, $n \leq \text{Sdis}(\text{POS}_\cdot)$
$\text{Set}(\text{VEF})$	所有候選事件及所有候選事件向前推算 n 個詞彙之集合, $n \leq \text{Sdis}(\text{POS}_\cdot)$
$\text{Fre}[\text{Set}(\text{VEF}_i)]$	集合 $\text{Set}(\text{VEF})$ 中第 i 個候選事件重複出現次數
$\text{Fre}[\text{Set}(\text{VEF}_{i,n})]$	集合 $\text{Set}(\text{VEF})$ 中第 i 個候選事件向前推算之第 n 個詞彙出現頻率
$N(\text{VEF}_{i,n})$	第 i 個候選事件向前推算之第 n 個詞彙於論壇文章中出現次數, $n \leq \text{Sdis}(\text{POS}_\cdot)$
$\text{PScore}(\text{ACE}_i)$	論壇文章中第 i 個候選事件之詞性相關性平均分數
$\text{TCScore}(\text{ACE}_i)$	論壇文章中第 i 個候選事件之前後文配對及標題加權分數
$\text{EScore}(\text{ACE}_i)$	論壇文章中第 i 個候選事件之代表分數
$\text{PPScore}(\text{ACE}_i)$	論壇文章中第 i 個候選事件之詞性相關性平均分數比重值
$\text{PTCcore}(\text{ACE}_i)$	論壇文章中第 i 個候選事件之前後文配對及標題加權分數比重值
L_p	既有訓練文章庫中之第 p 份文章
$L_{p,b}$	第 p 份訓練文章中第 b 個語句
L_b^ω	第 b 個文章語句集合向量
ADS_q	目標論壇文章中第 q 個代表語句 (即具有代表事件之語句)
ADS_q^ω	目標論壇文章代表語句之集合向量
$\text{Sim}(\text{ADS}_q, L_b)$	目標論壇文章中第 q 個代表語句與第 b 個文章語句集合之相似值
$\omega(\text{ADS}_q, L_b)$	目標論壇文章中第 q 個代表語句與第 b 個文章語句集合之篩選門檻值

	(若相似值 $\text{Sim}(\text{ADS}_q, L_b)$ 大於門檻值即與該語句相似)
$\text{Fi}(\text{ADS}_q, S_u)$	與目標論壇文章第 q 個代表語句相似之語句中，經篩選後之第 u 個相似語句
$\text{Sim_Fi}(\text{ADS}_q, S_u)$	目標論壇文章第 q 個代表語句與經篩選後，第 u 個相似語句相似值
ADS_DI_q	目標論壇文章第 q 個代表語句之極性項目組合(由情緒詞彙與重要詞彙(名詞或地方詞)所組成，即決定語句表達目的之項目組合，例如：語句「足以影響一般民眾的恐慌」中重要詞彙「民眾」與情緒詞彙「恐慌」之組合)
$\text{R}[\text{ADS_DI}_{q,y}, S_w]$	目標論壇文章第 q 個代表語句中第 y 個極性項目與第 w 個情緒類別之相關係數
NANC	重要詞彙之詞性(地方詞或名詞)
$\text{Set}(\text{NW})$	既有訓練文章庫中經人工標記否定詞(例如：不、不是)後所形成之否定詞集合
$\alpha(\text{NW})$	語句否定詞頻率參數值，用以判斷相似語句之原意
TW	既有訓練文章庫中所有語句經人工標記之轉折詞(Transitional Words)集合(例如：但是、然而)
$\text{I}[\text{Fi}(\text{ADS}_q, S_u), \text{TW}]$	為一指標函數，用以判斷相似語句中是否存在轉折詞(若 $\text{I}[\text{Fi}(\text{ADS}_q, S_u), \text{TW}] = 1$ ，即表示第 u 個相似語句中無轉折詞存在；反之，若 $\text{I}[\text{Fi}(\text{ADS}_q, S_u), \text{TW}] = 0$ ，則第 u 個相似語句中存在轉折詞)
$\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_w]$	目標論壇文章第 q 個代表語句經篩選後之第 u 個相似語句，偏向第 w 個情緒類別之機率值
$\text{ST}[\text{ADS}_q, S_w]$	目標論壇文章第 q 個代表語句對於第 w 個情緒類別之穩定值
DE_d	情緒詞彙庫中第 d 個情緒詞彙(一般人常用之情緒詞彙)
S_w	第 w 種情緒類別
$\text{LS}[S_w, \text{DE}_d]$	第 w 種情緒類別中第 d 個情緒詞彙
$\text{N}[L, \text{LS}(S_w, \text{DE}_d)]$	訓練文章中出現第 w 種情緒類別中第 d 個情緒詞彙之次數
SS_Set_w	第 w 種情緒語句之集合
$\text{ReS}'(\text{DE}_d, S_w)$	第 d 個情緒詞彙與第 w 種情緒類別之相關係數
$\text{ReS}(\text{DE}_d, S_w)$	第 d 個情緒詞彙與第 w 種情緒類別之隸屬係數

步驟(A1)—計算文章事件之前後文配對分數及標題加權分數

根據候選事件 ACE_i 前後語句（即標點符號與標點符號間之語句）與事件觸發詞之關係，可得知觸發詞與候選事件 ACE_i 乃具有分佈規則，故本步驟乃計算各觸發詞於目標候選事件前後語句及論壇文章內之分佈比例並相加取平均值後，即可取得目標候選事件 ACE_i 前後文配對平均分數，其值愈大即表示該候選事件為文章事件之代表性愈高。此外，論壇文章之標題可能代表文章之主旨，因此，將論壇文章候選事件 ACE_i 與標題之關係計算標題加權分數，以提升取得論壇文章之代表事件精確度，最後將兩分數相加即為候選事件之前後文配對分數及標題加權 $TCScore(ACE_i)$ ，如公式(3.1)所示。

$$TCScore(ACE_i) = \left[\sum_{\text{all } j} \left(\frac{Fre(CES_{i,j}^F)}{N(CES_{\cdot,\cdot}^F)} + \frac{N(CES_{i,j}^F)}{N(ACE_{\cdot})} \right) + \sum_{\text{all } k} \left(\frac{Fre(CES_{i,k}^B)}{N(CES_{\cdot,\cdot}^B)} + \frac{N(CES_{i,k}^B)}{N(ACE_{\cdot})} \right) + Fre(ACE_i, Title) \right] \cdot \left[\frac{1}{N(CES_{i,\cdot}^F)} \frac{1}{N(CES_{i,\cdot}^B)} \frac{1}{Fre(ACE_i)} \right] \quad (3.1)$$

步驟(A2)—取得論壇文章事件之詞性相關性平均分數

由於事件之詞性與事件前後詞彙之詞性具有關聯規則，故根據中研院漢語平衡語料庫（Sinica Corpus）所制定之詞性，先行統計所有文章中所有詞彙之詞性所佔比率，並標記詞性距離參數值 $Sdis(POS_{\cdot})$ （如表 3.2 所示），以訓練詞性距離之精準度。事件之詞性相關性平均分數計算公式如公式(3.2)、公式(3.3)所示，首先乃建立所有候選事件向前推算 $Sdis(POS_{\cdot})$ 個詞彙集合 $Set(VEF)$ ，並統計目標候選事件向前推算之各詞彙於詞彙集合 $Set(VEF)$ 及文章內出現次數，以取得各詞彙之詞性相關性機率，進而計算候選事件之詞性相關性平均分數 $PScore(ACE_i)$ 。

表 3.2、訓練文章所有候選事件詞性種類之統計

詞性種類	數量	所佔比率	詞彙距離
POS ₁ =Na (普通名詞)	36	34.6%	Sdis(POS)=1
POS ₂ =VE (動作句賓動詞)	16	15.4%	Sdis(POS)=2
POS ₃ =NH (代名詞)	16	15.4%	Sdis(POS)=3
POS ₄ =VC (動作及物動詞)	14	13.5%	Sdis(POS)=4
POS ₅ =ND (時間詞)	7	6.7%	Sdis(POS)=5
POS ₆ =VA (動作不及物動詞)	7	6.7%	Sdis(POS)=6
POS ₇ =VH (狀態不及物動詞)	5	4.8%	Sdis(POS)=7
POS ₈ =VB (動作類及物動詞)	2	1.9%	Sdis(POS)=8
POS ₉ =NC (地方詞)	1	1%	Sdis(POS)=9

$$\text{Set}(\text{VEF}) = \begin{bmatrix} \text{VEF}_{1,1} & \text{VEF}_{2,1} & \cdots & \text{VEF}_{i,1} \\ \text{VEF}_{1,2} & \text{VEF}_{2,2} & \cdots & \text{VEF}_{i,2} \\ \vdots & \vdots & & \vdots \\ \text{VEF}_{1,n} & \text{VEF}_{2,n} & \cdots & \text{VEF}_{i,n} \end{bmatrix} \quad (3.2)$$

$$\text{PScore}(\text{ACE}_i) = \frac{\sum_{\text{all } n} \frac{\text{Fre}[\text{Set}(\text{VEF}_{i,n})]}{N(\text{VEF}_{i,n})}}{\beta(\text{POS}_i) \cdot N(\text{ACE}_i) \cdot \text{Fre}[\text{Set}(\text{VEF}_i)]} \quad (3.3)$$

步驟(A3)—取得論壇文章代表性事件

因考量論壇文章類型之語意及標題重要性不同，故本步驟將透過步驟(A1)及步驟(A2)所取得候選事件之詞性相關性平均、前後文配對分數及標題加權分數加上比重分數並相加後，即可取得各候選事件之代表分數，其中代表分數最大之候選事件即為文章之代表事件，如公式(3.4)所示。

$$\begin{aligned} \text{EScore}(\text{ACE}_i) &= \text{PPScore}(\text{ACE}_i) \cdot \text{PScore}(\text{ACE}_i) \\ &+ \text{PTCScore}(\text{ACE}_i) \cdot \text{TCScore}(\text{ACE}_i) \end{aligned} \quad (3.4)$$

Where $\text{PPScore}(\text{ACE}_i) + \text{PTCScore}(\text{ACE}_i) = 1$

步驟(A4)—建立情緒詞彙與情緒類別之隸屬係數

本步驟藉由 Gregory 及 Daniel (2008) 中所建構之 482 個常用英文情緒詞彙及所對應之情緒類別轉換為中文字詞，並將情緒類別 S_w 分為憤怒 (Angry)、焦慮 (Anxiety)、

厭惡 (Disgust)、恐懼 (Fear)、快樂 (Happiness)、悲傷 (Sadness)、驚奇 (Surprise) 等七個類別，同時依據既有情緒詞彙所對應之情緒類別 (如表 3.3 所示，各類別以 2 個詞彙表示) 建立情緒語句集合 SS_Set_w (如公式(3.5)所示)，之後參考 Miao 等人 (2012) 之方法計算目標情緒詞彙於該情緒語句集合 SS_Set_w 中出現頻率，並統計訓練文章中各情緒類別之情緒詞彙總數後，即以公式(3.6)計算目標情緒詞彙與各情緒類別之相關係數 $ReS'(DE_d, S_w)$ ，並將相關係數予以正規化如公式(3.7)所示，即可得知目標情緒詞彙與情緒類別之隸屬係數，其結果整理如表 3.4 所示，此係數值愈大即代表情緒詞彙愈偏向該對應情緒類別。

$$SS_Set_w = L_{p,b} \text{ where } LS[S_w, DE_d] \text{ exist in } L_{p,b} \quad \forall d \quad (3.5)$$

$$ReS'(DE_d, S_w) = \frac{N(DE_d \cap SS_Set_w)}{\sum_{all d} N[L., LS(S_w, DE_d)]} \times \log_2(N(DE_d \cap SS_Set_w) + 1) \quad (3.6)$$

$$ReS(DE_d, S_w) = \frac{ReS'(DE_d, S_w)}{\sum_{all w} ReS'(DE_d, S_w)} \quad (3.7)$$

表 3.3、情緒類別所對應之情緒詞彙

情緒類別	英文情緒詞	中文情緒詞
憤怒 (Angry)	Contempt	鄙視、輕視、藐視
	Violent	激烈、暴力
焦慮 (Anxiety)	Awkward	尷尬、笨拙、不熟練的
	Uneasy	不安、不自在、擔心
厭惡 (Disgust)	Villain	壞人、惡棍
	Stinking	惡臭、非常討厭
恐懼 (Fear)	Horror	恐怖、毛骨悚然
	Doom	死亡、毀滅、惡運
快樂 (Happiness)	Friendly	友好、友善、親切
	Pleasant	愉快、爽快、舒適
悲傷 (Sadness)	Despair	失望、絕望
	Grief	悲苦、悲痛、哀痛
驚奇 (Surprise)	Amazed	吃驚、驚詫
	Shocked	震驚

表 3.4、情緒詞彙與各情緒類別之隸屬係數

情緒詞彙 情緒類別	DE ₁	DE ₂	...	DE _d	...
S ₁	ReS[DE ₁ , S ₁]	ReS[DE ₂ , S ₁]	...	ReS[DE _d , S ₁]	...
S ₂	ReS[DE ₁ , S ₂]	ReS[DE ₂ , S ₂]	...	ReS[DE _d , S ₂]	...
...
S _w	ReS[DE ₁ , S _w]	ReS[DE ₂ , S _w]	...	ReS[DE _d , S _w]	...
...

步驟(A5)–計算論壇文章代表語句與所有訓練文章內語句之相似值

透過前三步驟所取得論壇文章之代表事件後，本研究乃將訓練文章之語句向量集合 L_b^{ω} 及目標文章代表語句 ADS_q （即具有代表事件之語句）之集合向量 ADS_q^{ω} ，以向量空間模型之餘弦函數計算目標論壇文章之代表語句與所有訓練文章內之語句相似度，並以公式(3.8)判斷代表語句 ADS_q 與所有語句之相似值 $Sim(ADS_q, L_b)$ 。

$$\begin{aligned}
 L_{p,b} &= \{L_{p,1}, L_{p,2}, L_{p,3}, \dots, L_{p,b}, \dots\} \\
 L_b^{\omega} &= [w_1, w_2, \dots, w_b]^T \\
 ADS_q^{\omega} &= [w_1, w_2, \dots, w_q]^T \\
 Sim(ADS_q, L_b) &= \frac{ADS_q^{\omega} \cdot L_b^{\omega}}{|ADS_q^{\omega}| \cdot |L_b^{\omega}|}
 \end{aligned} \tag{3.8}$$

步驟(A6)–計算相似語句之情緒類別機率值

當取得論壇文章代表語句 ADS_q 與所有語句之相似值 $Sim(ADS_q, L_b)$ 後，若相似值 $Sim(ADS_q, L_b)$ 大於門檻值 $\omega(ADS_q, L_b)$ 且愈趨近於 1，即表示該文章語句與代表語句 ADS_q 具有相似之語意，但語句常因否定詞之存在而導致語句與原意相反，故若語句否定詞頻率大於參數值 $\alpha(NW)$ ，即視該語句與代表語句不為相似，此外，本步驟亦擷取相似語句中情緒詞彙與重要詞彙形成代表語句之極性項目組合，若相似語句中無名詞 N_a 或地方詞 N_c 以及情緒詞彙之存在，將忽略該相似語句之極性項目組合，以保持代表語句之情緒詞彙與重要詞彙關聯性，如公式(3.9)所示，並將極性項目所含情緒詞彙與情緒類別之隸屬係數 $ReS(D_d, S_w)$ ，作為該極性項目與情緒類別之隸屬係數 $R[ADS_{-q,y}, D_d, S_w]$ ，如公式(3.10)所示，最後即可以公式(3.11)計算所有相似語句偏向各情緒類別之機率值 $SP[Fi(ADS_q, S_u), S_w]$ ，其計算結果整理如表 3.5 所示。

IF $\text{Sim}(\text{ADS}_q, L_b) \geq \omega(\text{ADS}_q, L_b)$ and $N(L_b \cap \text{Set}(\text{NW})) < \alpha(\text{NW})$ and
 NANC exist in L_b and DE_d exist in L_b Then $\text{Sim}(\text{ADS}_q, L_b) \in \text{Sim_Fi}(\text{ADS}_q, S_u)$ (3.9)
 and $\text{NANC}, \text{DE}_d \in \text{ADS_DI}_q \forall d$

$$R[\text{ADS_DI}_{q,y}, S_w] = \text{ReS}(\text{DE}_d, S_w) \text{ where } \text{DE}_d \text{ exist in } \text{ADS_DI}_{q,y} \forall d \quad (3.10)$$

$$\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_w] = \frac{\exp \left(I[\text{Fi}(\text{ADS}_q, S_u), \text{TW}] + \text{Sim_Fi}(\text{ADS}_q, S_u) \times \frac{\sum_{\text{all } y} R[\text{ADS_DI}_{q,y}, S_w]}{N(\text{ADS_DI}_{q,y} \cap \text{Fi}(\text{ADS}_q, S_u))} \right)}{\sum_{\text{all } w} \exp \left(I[\text{Fi}(\text{ADS}_q, S_u), \text{TW}] + \text{Sim_Fi}(\text{ADS}_q, S_u) \times \frac{\sum_{\text{all } y} R[\text{ADS_DI}_{q,y}, S_w]}{N(\text{ADS_DI}_{q,y} \cap \text{Fi}(\text{ADS}_q, S_u))} \right)} \quad (3.11)$$

where $I[\text{Fi}(\text{ADS}_q, S_u), \text{TW}] = \begin{cases} 0, N(\text{Fi}(\text{ADS}_q, S_u) \cap \text{TW}) > 0 \\ 1, \text{otherwise} \end{cases}$ and

$\text{ADS_DI}_{q,y}$ exist in $\text{Fi}(\text{ADS}_q, S_u) \forall y$

表 3.5、相似語句與情緒類別之機率值

相似語句 情緒類別	$\text{Fi}(\text{ADS}_q, S_1)$	$\text{Fi}(\text{ADS}_q, S_2)$...	$\text{Fi}(\text{ADS}_q, S_u)$...
S_1	$\text{SP}[\text{Fi}(\text{ADS}_q, S_1), S_1]$	$\text{SP}[\text{Fi}(\text{ADS}_q, S_2), S_1]$...	$\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_1]$...
S_2	$\text{SP}[\text{Fi}(\text{ADS}_q, S_1), S_2]$	$\text{SP}[\text{Fi}(\text{ADS}_q, S_2), S_2]$...	$\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_2]$...
...
S_w	$\text{SP}[\text{Fi}(\text{ADS}_q, S_1), S_w]$	$\text{SP}[\text{Fi}(\text{ADS}_q, S_2), S_w]$...	$\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_w]$...
...

步驟(A7)—取得論壇文章之情緒類別

透過相似語句與各類別情緒之機率值 $\text{SP}[\text{Fi}(\text{ADS}_q, S_u), S_w]$ 即可以怕茲模型（如公式 (3.12) 所示）判斷論壇文章代表語句對於情緒類別之穩定值 $\text{ST}[\text{AL_DI}_q, S_w]$ （整理如表 3.6 所示），若該值愈趨近於 0 即表示文章代表語句對於該情緒類別偏向性愈高，但由於文章中大多包含數個代表語句，且論壇文章中常隱含數種情緒類別，故本研究乃將各代表語句之情緒類別視為目標論壇文章所表達之情緒類別。

$$\begin{aligned}
ST[ADS_q, S_w] = & -\sum_{allu} SP[Fi(ADS_q, S_u), S_w] \times I[Fi(ADS_q, S_u), TW] \\
& -\sum_{allu} SP[Fi(ADS_q, S_u), S_w] \times Sim_Fi(ADS_q, S_u) \times I[Fi(ADS_q, S_u), TW] \quad (3.12) \\
& -\sum_{allu} -SP[Fi(ADS_q, S_u), S_w] \times \log(SP[Fi(ADS_q, S_u), S_w])
\end{aligned}$$

表 3.6、代表語句與情緒類別之穩定值

代表語句 情緒類別	ADS ₁	ADS ₂	...	ADS _q	...
S ₁	ST[ADS ₁ , S ₁]	ST[ADS ₂ , S ₁]	...	ST[ADS _q , S ₁]	...
S ₂	ST[ADS ₁ , S ₂]	ST[ADS ₂ , S ₂]	...	ST[ADS _q , S ₂]	...
...
S _w	ST[ADS ₁ , S _w]	ST[ADS ₂ , S _w]	...	ST[ADS _q , S _w]	...
...

於「文章表達情緒判定」模組中乃將文章中具有代表事件之語句作為論壇文章代表語句，並分析代表語句與所有語句之關聯性，從中篩選與代表語句相似之語句，以藉由相似語句中情緒詞彙與情緒類別之關係，推論相似語句之情緒類別偏向性，再以怕茲模型推論文章代表語句之情緒類別偏向，以視為論壇文章所表達之情緒類別，亦即透過本模組可幫助使用者了解論壇文章所隱含不同之情緒類別。

3.2 發文者文章語句重組模組

本研究乃將「發文者文章語句重組模組」規劃為六大步驟，並於最終可重組文章之語句內容。於步驟(B1)與步驟(B2)中乃考量重組文章語句之內容後，期望可與原文章之語句流暢程度相似，並可以於最終推論之數個語句架構中，取得語句流暢程度最高之文章作為最終結果，故乃參考Chen等人(2010)判定語句流暢程度之方法規劃至步驟(B1)與步驟(B2)中，待取得文章之語句流暢程度後，考量重複分析語意相似之語句時，將影響後續推論之結果，因此於步驟(B3)中乃參考Kuo(2007)之方法規劃去除相似語句推論；此外，由於過去研究大多以相似語句或詞彙替換作為重組語句之方法，因此，規劃步驟(B4)至步驟(B5)時乃參考同為分析論壇與非結構化文章Liu等人(2011)之方法，將相似語句、填充詞與候選填充句取得方法規劃於此兩步驟，完成上述步驟後，即可滿足

語句重組之基礎條件，最後於步驟(B6)規劃中，本研究乃提出一方法依據語句順序重組語句之內容。

過去研究針對文章語句重組部份，改寫後大多無法與原先欲傳達之情緒、情感一致，而無法滿足使用者之需求，且論壇發文者發表文章時，內文完整性與流暢性乃吸引讀者注目關鍵之一。針對發文者語意表達程度部分，本研究先行參考Chen等人(2010)之方法針對目標論壇文章之間接特徵(文章所使用之詞語種類數)給予初始分數，再以投票演算法(Voting Algorithm)將文章間之相似程度以累加式方式，持續修正目標論壇文章評閱分數值直至收斂穩定狀態，以判斷該論壇文章是否須進行後續改寫內容之程序。本研究針對文章語句重組部分乃替換文章撰寫者所描述之情緒語句及重要語句，故本研究參考Kuo(2007)之方法先行針對各語句詞義消歧部分去除語意相近之句子，以取得文章之重要語句，接著本研究乃將重要語句及代表語句(具代表事件之語句)與訓練文章庫之所有語句進行相似分析，以取得論壇文章重組時所需之候選填充句，但因候選填充句、重要語句、代表語句組合後可能產生句子不連貫，是故，本研究乃藉由訓練文章庫中所有文章之語句彙集候選填充句之非關鍵詞(即填充詞)，以形成候選填充句之多重組合句，最後，透過多重組合句之組合形成數個論壇文章之語句架構，並依據最高評閱分數之語句架構，作為目標論壇文章重組後之內容。本模組之語句重組流程如圖3.3所示，其所使用之符號定義如下：

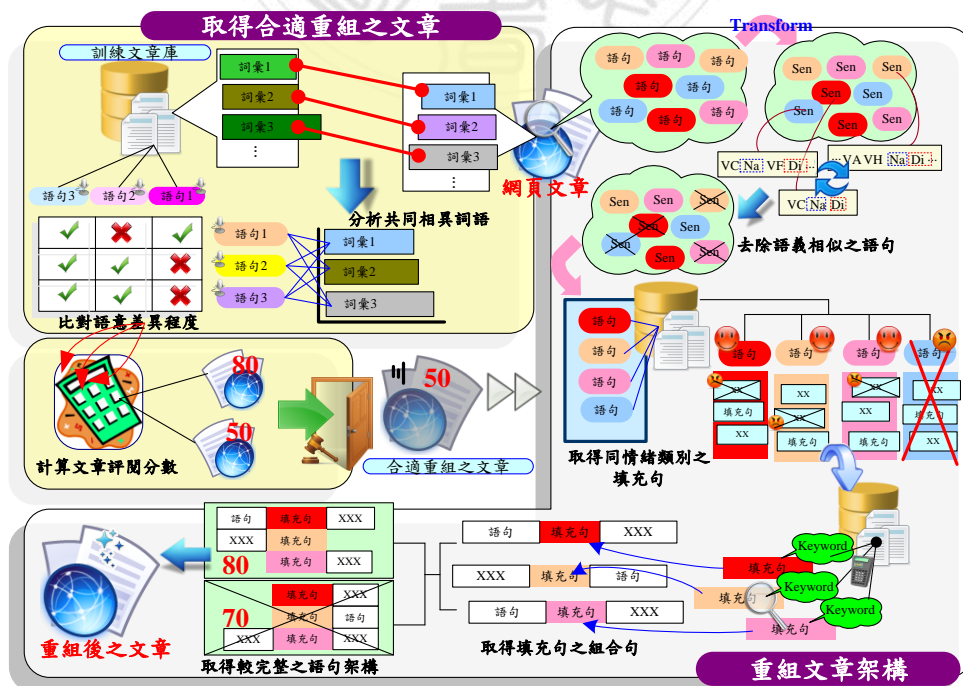


圖 3.3、發文者文章語句重組示意圖

符號定義

$L_{p,k}$	第 p 份訓練文章中第 k 個詞彙
LDW_Set_p	第 p 份訓練文章之相異詞語集合
AL_j	目標論壇文章中第 j 個詞彙
ADW_Set	目標論壇文章之相異詞語集合
$N(LDW_Set_p)$	第 p 份訓練文章中相異詞語集合之詞彙總數
$N(ADW_Set)$	目標論壇文章中相異詞語集合之詞彙總數
L_*	訓練文章庫之文章總數
$PGrade$	目標論壇文章之評閱分數
$LPGrade_p$	第 p 份訓練文章之評閱分數
$Sim(A_T, L_p)$	目標論壇文章與第 p 份訓練文章之相似度
Th_PGrade_u	第 u 種評閱分數門檻值，用以判斷論壇文章是否須執行改寫程序， $u=1, 2, 3, 4, 5$ ； Th_PGrade_1 為平均值門檻值、 Th_PGrade_2 為中位 數門檻值、 Th_PGrade_3 為直接定義門檻值、 Th_PGrade_4 為最小值門檻 值， Th_PGrade_5 為四分位數門檻值
$Sd(A_T, L_*)$	目標論壇文章與訓練文章之相似程度標準差
$Sem(A_T, L_p)$	目標論壇文章與第 p 份訓練文章之語意差異程度
CRT_Set	合適語句重組之論壇文章集合
AS_m	目標論壇文章中第 m 個語句
$ASL_{m,i}$	目標論壇文章第 m 個語句中第 i 個詞性之代表詞彙(經由 Denisowski's CEDICT 翻譯器處理後之英文詞彙)，例如語句「不管走路或開車真的 都要很小心」中包含有「Cbb、Va、Caa、D、Dfa、VK」等詞性。
$ASL_{*,i}$	目標論壇文章所有語句之第 i 個詞性之代表詞彙總數
$ASLE_Set_{m,i}$	目標論壇文章第 m 個語句中，第 i 個詞性代表詞彙之延伸定義詞彙集 合(利用 WordNet 所取得之延伸詞彙)
CL_Set_i	目標論壇文章中第 m 個語句與第 k 個語句中，第 i 個詞性代表詞彙共 同延伸詞彙集合
$CL_Set_i[EL_u]$	第 m 個語句與第 k 個語句中，第 i 個詞性代表詞彙共同延伸詞彙集合 之第 u 個詞彙

$N(\text{CL_Set}_i[\text{EL}_u])$	目標論壇文章中第 m 個語句與第 k 個語句中，第 i 個詞性代表詞彙共同延伸詞彙集合之第 u 個詞彙出現頻率
$\text{Re}[\text{ASL}_{m_i}, \text{ASL}_{k_i}]$	目標論壇文章中第 m 個語句與第 k 個語句中，第 i 之詞性代表詞彙相關程度
$\text{amp}[\text{AS}_m, \text{AS}_k]$	目標論壇文章中第 m 個語句與第 k 個語句之歧義相似分數
Ex_Set	目標論壇文章之擴充語句集合
$\text{Ex_Set}[\text{IS}_g]$	目標論壇文章之擴充語句集合中第 g 個語句
$\text{Sim}(\text{Ex_Set}[\text{IS}_g], L_b)$	目標論壇文章之擴充語句集合中第 g 個語句，與第 b 個論壇文章語句集合之相似值
Con_Set	目標論壇文章之候選填充句集合
$\text{LS}_{p,b,d}$	第 p 份訓練文章中第 b 個語句之第 d 個情緒詞彙
AS_Set_v	目標論壇文章所表達第 v 種情緒類別之所有情緒詞彙集合
$\text{Con_Set}[\text{CS}_{q,r}]$	目標論壇文章之候選填充句集合中，第 q 個候選填充句之第 r 個關鍵詞（即普通名詞、情緒詞彙、地方詞及形容詞）
$\text{LF}_{p,b,x}$	第 p 份訓練文章中第 b 個語句之第 x 個填充詞（關鍵詞與關鍵詞間之詞彙）
$\text{CS_MUS}_{q,z}$	第 q 個候選填充句之第 z 個組合句
Rec_A_y	目標論壇文章重組後第 y 個語句架構
Rec_Grade_y	目標論壇文章重組後第 y 個語句架構之評閱分數

步驟(B1)–取得論壇文章之初始評分

本步驟乃利用目標論壇文章 A_T 之表面特徵「相異詞語數」作為初始評分之依據，首先乃去除文章重複出現之詞語，將相同之詞語視為1種詞語種類，即可取得目標論壇文章真正所應用之相異詞語數 $N(\text{FAL_Set})$ ，此外，本步驟亦同時計算訓練文章庫所有論壇文章之相異詞語數 $N(\text{LDW_Set}_p)$ ，以便於後續步驟與目標論壇文章 A_T 進行相似性比對（如公式(3.13)所示）。

$$\begin{aligned} \text{ADW_Set} &= \{AL_j \mid AL_j \text{ not exist in ADW_Set } \forall j\} \\ \text{LDW_Set}_p &= \{L_{p,k} \mid L_{p,k} \text{ not exist in LDW_Set}_p \forall k\} \end{aligned} \tag{3.13}$$

步驟(B2)—計算目標論壇文章之評閱分數

本步驟首先乃將目標論壇文章與訓練文章之共用詞語作為兩文章之相似度 $\text{Sim}(A_T, L_p)$ ，並同時計算相似程度之標準差 $\text{Sd}(A_T, L_p)$ ，如公式(3.14)所示，以使評閱分數可收斂於評閱範圍中。接著透過公式(3.15)計算訓練文章對於目標文章之語意差異程度 $\text{Sem}(A_T, L_p)$ ，以得知各訓練文章共用詞語之差異性，最後即以公式(3.16)計算目標論壇文章之評閱分數 PGrade ，若評閱分數 PGrade 低於門檻值 Th_PGrade 即表示該目標論壇文章之語意表達程度較低。

$$\text{Sim}(A_T, L_p) = N(\text{ADW_Set} \cap \text{LDW_Set}_p)$$

$$\text{Sd}(A_T, L_p) = \sqrt{\frac{\sum_{\text{all } p} \left(\text{Sim}(A_T, L_p) - \frac{\sum_{\text{all } p} \text{Sim}(A_T, L_p)}{L_p} \right)^2}{L_p - 1}} \quad (3.14)$$

$$\text{Sem}(A_T, L_p) = \frac{\left(\frac{N(\text{LDW_Set}_p) - \sum_{p \neq m} N(\text{LDW_Set}_m)}{L_p} \right)}{\text{Sd}(A_T, L_p)} \quad (3.15)$$

$$\text{PGrade} = \sum_{\text{all } p} \text{Sim}(A_T, L_p) \times \text{Sem}(A_T, L_p) \quad (3.16)$$

此外，本研究乃提出四種門檻值設定方式以提供使用者針對需求篩選論壇文章，其門檻值 Th_PGrade 設定方式分別為平均值、中位數、最小值、門檻值直接定義與四分位數等方式，亦表示若評閱分數 PGrade 小於門檻值 Th_PGrade 即表示將該論壇文章語句及語意通順度不佳，並將該論壇文章放置合適改寫文章集合 CRT_Set 中，其各門檻值設定方式如下：

(a) 以平均值作為門檻值

首先乃針對訓練文章庫中各篇論壇文章重複公式(3.14)至公式(3.16)計算其評閱分數並加總，再除以訓練文章總數求得評閱分數平均值，進而將平均值作為門檻值 Th_PGrade_1 ，若評閱分數低於整體平均值，即表示該網頁文件之語意表達性較低，如公

式(3.17)所示。

$$\text{Th_PGrade}_1 = \frac{\sum_{\text{all } p} \text{LPGrade}_p}{L} \quad (3.17)$$

$$\text{CRT_Set} = \{ A_T \mid \text{PGrade} \leq \text{Th_PGrade}_1 \}$$

(b)以中位數作為門檻值

為避免評閱分數加總最大值與最小值差距過大而影響整體平均值，因此以整體筆數之中位數作為門檻值 Th_PGrade_2 ，如公式(3.18)所示，若目標論壇文章評閱分數低於（或相等）中位數，即表示該論壇文章之內容架構須進行重組程序。

$$\text{Th_PGrade}_2 = \begin{cases} \text{LPGrade}_{\frac{p+1}{2}} & \text{IF } L \text{ is odd} \\ \frac{1}{2} \times \left(\text{LPGrade}_{\frac{p+1}{2}} + \text{LPGrade}_{\frac{p+1}{2}+1} \right) & \text{IF } L \text{ is even} \end{cases} \quad (3.18)$$

$$\text{CRT_Set} = \{ A_T \mid \text{PGrade} \leq \text{Th_PGrade}_2 \}$$

(c)直接定義門檻值

使用者亦可自行制定門檻值，以進行篩選合適語句重組之論壇文章。如公式(3.19)所示，若評閱分數小於自行定義門檻值 Th_PGrade_3 則將該網頁文件放置合適語句重組文章集合 CRT_Set 中。

$$\text{CRT_Set} = \{ A_T \mid \text{PGrade} \leq \text{Th_PGrade}_3 \} \quad (3.19)$$

(d)以最小值作為門檻值

訓練文章庫乃經篩選後之文章，其中文章評閱分數最小者極可能文章語意表達性較低，因此，本研究將評閱分數最低者作為門檻值 Th_PGrade_4 供使用者篩選論壇文章，若評閱分數小於最小值門檻值 Th_PGrade_4 ，即表示該論壇文章相對於訓練文章之語意表達性較為不佳，故文章須進行語句重組之程序，如公式(3.20)所示。

$$\text{Th_PGrade}_4 = \text{Min}(\text{LPGrade}_p) \forall p \quad (3.20)$$

$$\text{CRT_Set} = \{ A_T \mid \text{PGrade} \leq \text{Th_PGrade}_4 \}$$

(e)以四分位數作為門檻值

為能更精確挑選且不受極端值影響篩選論壇文章之結果，於此乃制定整體筆數之四分位數作為門檻值，即以第三位之四分位數進行篩選。首先計算第三位之四分位數指標 Q ，以界定門檻值 Th_PGrade_5 ，進而篩選論壇文章之評閱分數，如公式(3.21)所示。當中，若評閱分數小於最小值門檻值 Th_PGrade_5 則表示該論壇文章語意不佳，並放置合適語句重組文章集合 CRT_Set 中。

$$\begin{aligned}
 Q &= L_{\bullet} \times 75\% \\
 Th_PGrade_5 &= \begin{cases} LPGrade_{P \times 75\%} & \text{IF } Q \notin \{X : |X| \in N\} \\ \frac{1}{2} \times (LPGrade_{P \times 75\%} + LPGrade_{P \times 75\% + 1}) & \text{IF } Q \in \{X : |X| \in N\} \end{cases} \quad (3.21) \\
 CRT_Set &= \{A_T \mid PGrade \leq Th_PGrade_5\}
 \end{aligned}$$

步驟(B3)—去除語意相近之語句

由於中文語句常有不同詞彙組成之句子，但語意相同之情形，故本步驟乃去除論壇文章中語意相同之語句，以更精確針對重要語句進行重組。本研究乃將論壇文章目標語句利用中英翻譯器 (Denisowski's CEDICT) 取得語句中各詞性詞彙之英文詞彙 ASL_{m_i} ，之後利用 Wordnet 取得各種詞性之英文詞彙延伸定義詞彙集合 $ASLE_Set_m$ ，並針對所有語句之英文詞彙計算其關聯程度 $Re[ASL_{m_i}, ASL_{k_i}]$ ，如公式(3.22)所示，最後，將所有目標語句與其他語句之各種詞性詞彙相關程度加總後，即為目標語句與其他語句之歧義相似度 $amp[AS_m, AS_k]$ ，如公式(3.23)所示，並以最大歧義相似度之語句作為目標語句之歧義句，而本研究乃去除目標語句與歧義句中詞性種類最少之語句，以取得論壇文章之重要語句。

$$\begin{aligned}
 CL_Set_i &= ASLE_Set_{m_i} \cap ASLE_Set_{k_i} \quad \forall i \\
 Re[ASL_{m_i}, ASL_{k_i}] &= \text{Max}(-\log(\frac{N(CL_Set_i[EL_u])}{ASL_{\bullet, i}})) \quad \forall u \quad (3.22)
 \end{aligned}$$

$$amp[AS_m, AS_k] = \sum_{\text{all } i} Re[ASL_{m_i}, ASL_{k_i}] \quad (3.23)$$

步驟(B4)–取得重要語句與代表語句之候選填充句

先前步驟中所取得之重要語句及代表語句（具代表事件之語句），本步驟首先依據語句於論壇文章之位置順序，彙集成擴充語句集合 Ex_Set（如公式(3.24)所示），接著將擴充語句集合 Ex_Set 中所有語句與訓練文章庫中所有語句 L_b 進行相似性比對（如公式(3.25)所示），並以相似程度最高語句之前後語句作為候選填充句，此外，為使改寫論壇文章後可與原先論壇文章欲表達之情緒一致，故本步驟將過濾與論壇文章原先情緒不一致之候選填充句，如公式(3.26)所示。

$$Ex_Set = \{AS_1, AS_2, AS_3, \dots, AS_m \mid \text{Max}(\text{amp}[AS_m, AS_k]) \text{ and } \text{Max}(ASL_{m,*})\} \quad (3.24)$$

$$\text{Sim}(Ex_Set [IS_g], L_b) = \frac{Ex_Set[IS_g] \cdot L_b^o}{|Ex_Set[IS_g]| \cdot |L_b^o|} \quad (3.25)$$

$$Con_Set = \{L_{b+1}, L_{b-1} \mid \text{Max}(\text{Sim}(Ex_Set[IS_g], L_b)) \text{ and } LS_{p,b,d} \in AS_Set_v\} \forall g, w \quad (3.26)$$

步驟(B5)–取得候選填充句之多重組合句

考量候選填充句 Con_Set[CS_q] 與重要語句及代表語句組合後可能產生句子不連貫之情況，因此，本步驟乃將候選填充句之關鍵詞（普通名詞、情緒詞彙、地方詞及形容詞）與訓練文章庫中所有文章之語句進行比對，以取得候選填充句之填充詞 LF_{p,b,x}（關鍵詞與關鍵詞間之詞彙），進而彙集候選填充句之多重組合句 CS_MUS_{q,z}，如公式(3.27)所示。

$$CS_MUS_{q,z} = \left\{ \begin{array}{l} LF_{p,b,x} \mid \text{Con_Set}[CS_{q,r}] \text{ exist in } L_{p,b} \\ \text{and Con_Set}[CS_{q,r}] \text{ not exist in } LF_{p,b,x} \end{array} \right\} \forall r, x \quad (3.27)$$

步驟(B6)–重組論壇文章架構

根據前一步驟所取得候選填充句之多重組合句，本步驟乃依據語句之順序，將目標論壇文章候選填充句之多重組合句 CS_MUS_{q,z}、代表語句及重要語句進行組合，如公式(3.28)所示，以將合適重組之論壇文章進行語句重組，此外，因候選填充句具有多個組合句，故重組後之論壇文章將有數個不同之語句架構，為使從中取得架構較完整之論壇文章，本步驟乃將重組後之論壇文章藉由步驟(B1)與步驟(B2)取得評閱分數 Rec_Grade

(彙整如表 3.7 所示)，並由評閱分數最高之語句架構作為目標論壇文章重組後之內容。

$$\text{Rec_A}_n = \begin{bmatrix} \text{CS_MUS}_{1,1} \\ \text{CS_MUS}_{1,2} \\ \vdots \\ \text{CS_MUS}_{1,c} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CS_MUS}_{2,1} \\ \text{CS_MUS}_{2,2} \\ \vdots \\ \text{CS_MUS}_{2,a} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CS_MUS}_{3,1} \\ \text{CS_MUS}_{3,2} \\ \vdots \\ \text{CS_MUS}_{3,v} \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} \text{CS_MUS}_{q,1} \\ \text{CS_MUS}_{q,2} \\ \vdots \\ \text{CS_MUS}_{q,z} \end{bmatrix} \rightarrow \dots \quad (3.28)$$

whereMax(Rec_Grade_n)

表 3.7、論壇文章架構與評閱分數彙整表

論壇文章語句架構	語句架構內容	評閱分數
Rec_A ₁	CS_MUS _{1,1} → CS_MUS _{2,2} → … → CS_MUS _{q,3}	Rec_Grade ₁
Rec_A ₂	CS_MUS _{1,1} → CS_MUS _{2,3} → … → CS_MUS _{q,4}	Rec_Grade ₂
…	…	…
Rec_A ₁₄	CS_MUS _{1,5} → CS_MUS _{2,1} → … → CS_MUS _{q,2}	Rec_Grade ₂
…	…	…
Rec_A _n	CS_MUS _{1,c} → CS_MUS _{2,a} → … → CS_MUS _{q,v}	Rec_Grade _n

針對論壇文章語句重組部分，過去研究大多乃根據論壇文章之語意結構及流暢度進行語句重組，但大多數未加入文章撰寫者欲表達之情緒因素，使得重組後之內容常與文章撰寫者欲表達之情緒不一致，是故，本研究乃先行判斷論壇文章之語意程度，並藉由論壇文章之情緒類別分析重要語句及代表語句，以取得論壇文章重組所需之候選填充句（包含情緒語句），最後透過候選填充句之多重組合句重組論壇文章之語句架構，以期望協助文章撰寫者提升論壇文章內容之流暢度，其內容亦可與原先欲表達之情感一致。

根據本研究所發展之「以論壇文章蘊含情緒為基之文章重組模式」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 3.8。

表 3.8、參考文獻延伸與本研究發展彙整表

探索主題	解析目標	解析演算法	參考文獻	本研究發展與延伸
文章情緒類別判定	文章代表事件解析	-	中研院斷詞系統	<ul style="list-style-type: none"> ● 以中研院斷詞系統拆解文章之詞彙並將區分詞性。 ● 擴展文章代表詞彙之分析為文章內容代表主題 ● 結合斷詞法、詞性分析法、標題加權分析及語句上下文解析法，計算候選事件代表性得分
		語句上下文與標題分析法 (擴展既有研究方法)	Tung 與 Lu (2012)	
	情緒詞彙之情緒類別隸屬係數建立	情緒語句庫建立；	Gregory 及 Daniel (2008)	<ul style="list-style-type: none"> ● 以 482 個基礎英文詞彙與對應之情緒類別，作為基礎情緒詞彙庫
		情緒詞彙隸屬係數推論法； (擴展既有研究方法)	Miao 等人 (2012)	<ul style="list-style-type: none"> ● 整合基礎情緒詞彙對應之情緒語句建立情緒語句庫 ● 原以情緒語句為基礎推論隸屬係數，擴展為語句與情緒詞彙之整合形成隸屬係數推論法
	相似語句分析	向量空間模型之餘弦函數 (改良既有之演算法)	-	<ul style="list-style-type: none"> ● 將向量空間模型應用於代表語句與所有語句，以語句向量形成相似語句分析法
	文章情緒類別分析	情緒機率值分析法 (應用既有研究方法)	Zhao 等人 (2010)	<ul style="list-style-type: none"> ● 結合情緒詞彙隸屬係數與重要詞彙，發展為相似語句之情緒機率值分析法
情緒穩定值分析法 (應用既有研究方法)		<ul style="list-style-type: none"> ● 整合相似語句情緒機率值與代表語句，以解析得最終代表語句之情緒穩定值 		
文章語句重組	文章評閱分數分析法	語句相似度分析 (擴展既有研究方法)	Chen 等人 (2010)	<ul style="list-style-type: none"> ● 整合相異詞語數、共詞法則，將原文章分群法，延伸並擴展為虛擬社群成員資料分群法
		門檻篩選法則 (擴展既有研究方法)		<ul style="list-style-type: none"> ● 結合四分位、中位數、平均值與最大值等統計法，劃分文章評級級別。
	語句語意分析法	語意相似度解析法 (整合既有研究方法)	Kuo 與 Huang (2007)	<ul style="list-style-type: none"> ● 針對語句分析相似度以擴展為語句歧義相似度分析法
		-	WordNet、Denisowski's CEDICT	<ul style="list-style-type: none"> ● 將語句翻譯為英文，並延伸詞彙相關定義
	多重組合句建立法	候選填充句建立法則 (應用既有研究方法)	Liu 等人 (2011)	<ul style="list-style-type: none"> ● 整合重要語句、代表語句與相似語句，並以最高相似度語句之前後語句形成候選填充句
		多重組合句建立法則 (應用既有研究方法)		<ul style="list-style-type: none"> ● 藉由候選填充句與填充句之替換，以建立多重組合句
文章語句重組	語句架構重組	啟發式演算法 (創新研究方法)	<ul style="list-style-type: none"> ● 整合多重組合句並根據語句順序重新組合文章之架構 	
<p>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。</p>				

第四章、系統架構

根據第三章所發展之方法論與模式，本研究乃開發一套以論壇文章蘊含情緒為基之文章重組系統，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「文章表達情緒判定模組」與「發文者文章語句重組模組」兩模組，以分別針對上述兩面向之議題進行解析，於前者中，主要以發文者文章之「文章標題」、「文章內容」、「文章代表語句」及「基礎情緒詞彙庫」等資料做為分析之基礎，以推論「情緒詞彙隸屬係數」後再根據「相似語句」代表語句所含之「情緒詞」等資料，並將上述數據整合與計算後得知「文章表達之情緒類別」；於後者中乃以文章內文之「間接特徵」、「相異詞語數」為基礎，先行以此基礎資料與訓練文章進行比對，以得知文章之評閱分數，並將文章語意相似之語句予以去除，同時從中取得數個填充句，最後，將填充句與文章語句進行替換再重新分析評閱分數並取最大值，以重組評閱分數低落文章之語句。藉由上述兩模組判定之結果，於論壇管理者層面，可協助管理者從大量文章中，過濾違規或具有偏激字詞之文章，以維持論壇文章品質；於論壇發文者層面，可避免發文時因無意寫入之違規字詞因而違反論壇之規範，以提升發文者持續分享文章之意願。本章即針對本研究所提之「以論壇文章蘊含情緒為基之文章重組系統」，分別以系統核心架構、系統功能架構、資料模式定義及系統流程與系統開發工具進行說明。

4.1 以論壇文章蘊含情緒為基之文章重組系統之核心架構

本研究所開發之「以論壇文章蘊含情緒為基之文章重組系統」依其運作流程可分為「論壇文章上傳」、「資料預處理」、「文章表達情緒解析」及「發文章文章語句重組」等四大階段，本系統之運作流程架構如圖 4.1 所示，各功能層次之詳細說明如下。

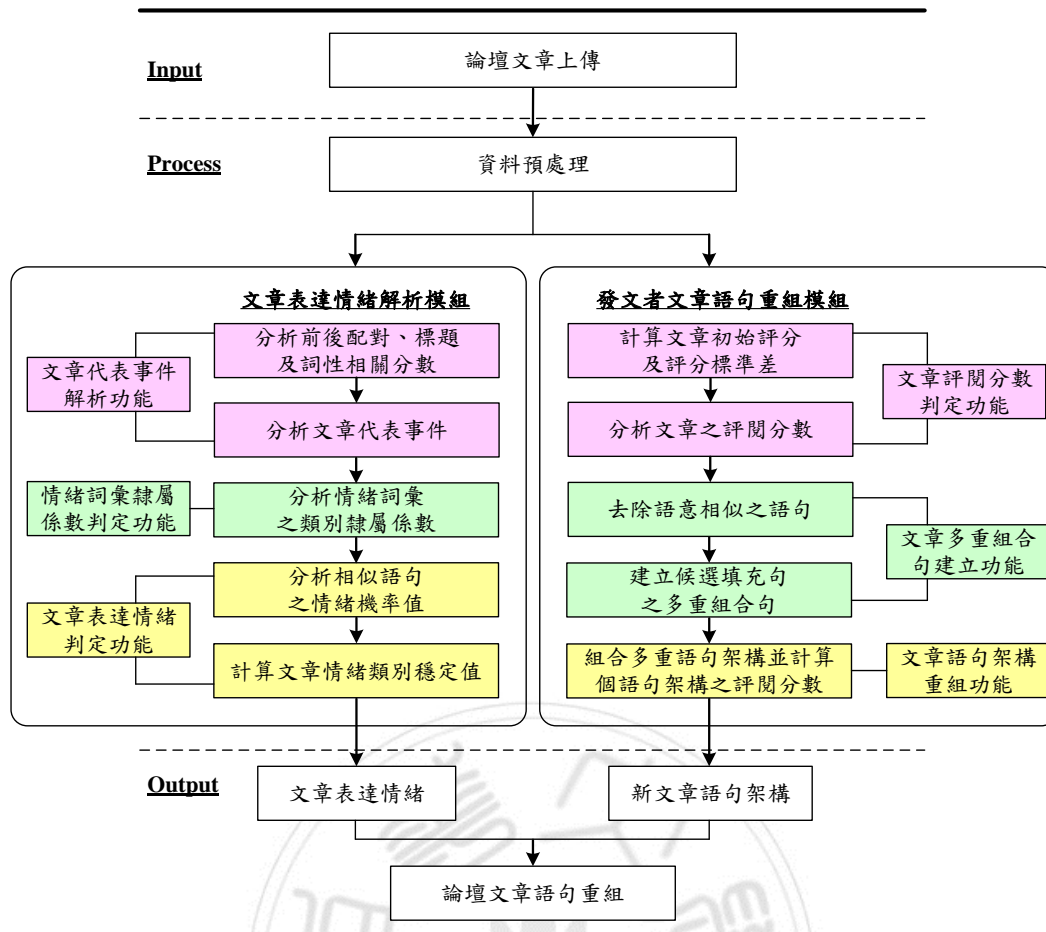


圖 4.1、以論壇文章蘊含情緒為基之文章重組系統之流程架構

- 論壇文章上傳

系統使用者可蒐集論壇文章（如論壇文章），並將文章上傳至系統中，系統即會截取文章之內文並透過「資料預處理」之斷詞、相異詞語數後，將作為「文章表達情緒模組」中文章代表事件解析之依據，並作為「發文者文章語句重組模組」中分析文章評閱分數之依據。

- 文章表達情緒解析

文章表達情緒判定模組可分為「文章代表事件解析」、「情緒詞彙隸屬係數判定」及「文章表達情緒判定」三大功能，於「文章代表事件解析」功能中，透過文章之匯入及入斷詞之預處理，分析文章前後文、標題及詞性相關性分數，以取得文章代表事件；於「情緒詞彙隸屬係數判定功能」中，藉由各情緒類別之情緒語句初步取得情緒詞之關係係數，再將關係係數予以正規化，以建立情緒詞彙之類別隸屬係數；於「文章表達情緒判定功能」中，透過文章之代表語句，並針對代表語句與情緒詞彙之關係取得情緒機率值，以取得文章所表達之情緒類別。

● 發文者文章語句重組

發文者文章語句重組模組可分為「文章評閱分數判定」、「文章多重組合句建立」及「文章語句架構重組」三大功能。於「文章評閱分數判定功能」中，藉由預處理取得之相異詞語數作為初始評分及評分標準差，再根據訓練文章與目標文章內文之比對，以計算得文章之評閱分數；「於文章多重組合句建立功能」中乃先行去除語意相近之語句，並根據文章代表及重要語句計算訓練文章庫各語句之關聯性，以取得候選填充句之多重組合句；最後於「文章語句架構重組功能」中，藉由候選填充句之多重組合句與訓練文章庫之前後語句組合後，取得數個文章多重語句架構，再針對所有語句架構進行「文章評閱分數判定功能」之判定，從中取得最高評閱分數之語句架構。

4.2 系統功能架構

本研究所建置以論壇文章蘊含情緒為基之文章重組系統乃建置於網際網路環境下。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。於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，以下分別針對此兩種身分使用者所能使用之功能加以說明：

一般使用者

1. 可新增與上傳未判定文章表達情緒、語句重組及斷詞預處理之文章資料至系統中。
2. 可瀏覽符合使用者上傳之文章及斷詞內容。
3. 可瀏覽符合使用者所上傳之文章，經分析後之文章表達情緒及文章重組後之結果。

系統管理者

1. 可新增、修改、查詢及刪除論壇文章內容。
2. 可新增、修改、查詢及刪除情緒詞彙。
3. 可瀏覽/編輯系統資料庫內所有訓練文章及分析資料。
4. 可修改/查詢系統參數門檻值
5. 可執行情緒詞彙隸屬係數判定
6. 可執行文章表達情緒解析
7. 可執行文章語句重組解析

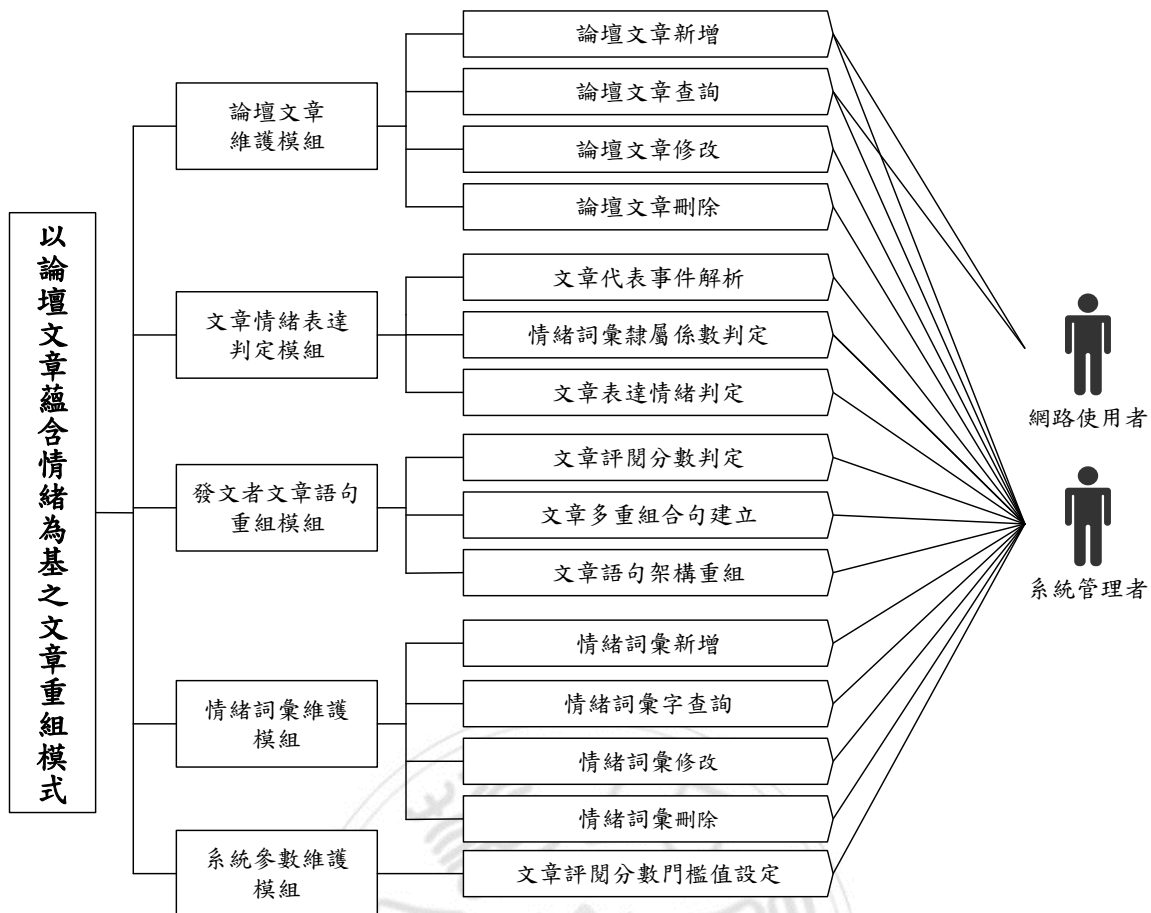


圖 4.2、以論壇文章蘊含情緒為基之文章重組系統之功能架構

針對上述系統架構所包含之基本功能模組說明如下：

(一) 論壇文章維護模組

- 論壇文章新增：提供系統管理者及網路使用者將論壇文章匯入並維護於系統資料表中。
- 論壇文章查詢：提供系統管理者及網路使用者查詢文章內容及解析之結果。
- 論壇文章修改：提供系統管理者修改錯誤論壇文章之內容。
- 論壇文章刪除：提供系統管理者刪除錯誤內容之論壇文章。

(二) 文章表達情緒判定模組

- 文章代表事件解析：系統管理者可輸入查詢字串、選取查詢條件，待查詢完畢後選取欲解析之文章，即可針對文章先行解析前後語句、字詞詞性分佈情況及標題權重分數，並給予候選事件加權分數以取得文章之代表事件。
- 情緒詞彙隸屬係數判定：系統管理者查詢訓練文章數後，系統首先將統計各情緒類別之情緒詞彙數、詞彙於各情緒類別之頻率數，以初步給予情緒詞彙與情緒類別之關係係數，最後再予以正規化以取得情緒詞彙之情緒類別隸屬係數。

- 文章表達情緒判定：系統管理者可輸入查詢字串及查詢條件，待查詢完畢並選擇目標文章後，即可將文章代表事件之語句與訓練文章之語句進行相似性分析，並將相似語句及代表語句之情緒詞彙進行分析，以得知文章之表達情緒。

(三) 發文者文章語句重組模組

- 文章評閱分數判定：系統管理者可輸入查詢字串及查詢條件，待查詢完畢後選擇欲解析之文章，即可針對文章之相異詞語數及評分標準差進行解析，在與訓練文章之語意差異程度進行分析，以取得文章之評閱分數。
- 文章多重組合句建立：系統管理者可輸入查詢字串及查詢條件，待查詢完畢後選擇欲解析之文章，即可針對文章之重要語句、代表詞彙之英文及延伸定，進行兩語句關聯程度之分析，並進行門檻值篩選去除相似語句，並根據候選填充句與訓練文章之語句相似度建立文章之多重組合句。
- 文章語句架構重組：系統管理者可輸入查詢字串及查詢條件，待查詢完畢後選擇欲解析之文章，即可針對文章多重組合句及代表語句進行替換，以取得數個文章語句架構，並重複文章評閱分數判定，分析所有文章語句架構之評閱分數，從中取得最高評閱分數之語句架構作為最後輸出結果。

(四) 系統參數設定模組

- 文章評閱分數門檻值設定：提供權限內使用者設定文章評閱之門檻條件及數值，可依中位數、四分位數、平均值、最大值及自行定義等條件進行設定。

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論壇文章新增及查詢功能（如圖4.3之一般使用者所示），而系統管理者則可進行論壇文章及情緒詞彙之維護（包含新增、查詢、修改與刪除）、系統參數設定以及文章表達情緒判定、發文者文章語句重組等（如圖4.3之系統管理者所示）。

首先，當使用者上傳未解析之論壇文章資料後，系統管理者乃需完成參數設定以及情緒詞彙上傳，待設定與上傳完畢後，系統管理者即可執行文章表達情緒判定模組，系統將根據文章之前後文、標題及詞性相關性分數解析代表事件，再根據情緒詞彙之分佈狀況建置情緒詞彙隸屬系數，最後透過代表事件之相關程度、語句相似度及情緒詞彙關聯程度解析文章之表達情緒；另一方面，系統管理者則透過發文者文章語句重組模組，依據文章之相異詞語數、相似度標準差等數據分析文章之評閱分數，並根據門檻值判斷是否需進行後需語句之重組，接著，系統乃

先行去除相似之語句，並根據語句關聯程度建立候選填充句，再依據填充句之語句組合架構重新計算評閱分數，並將最高評閱分數者視為最終輸出之語句架構；最後，使用者即藉由論壇文章資料維護模組之查詢功能，檢視系統中分析之文章表達情緒判定及文章語句架構重組等結果。綜合上述，本研究將以論壇文章蘊含情緒為基之文章重組系統之運作架構繪製如圖4.3。

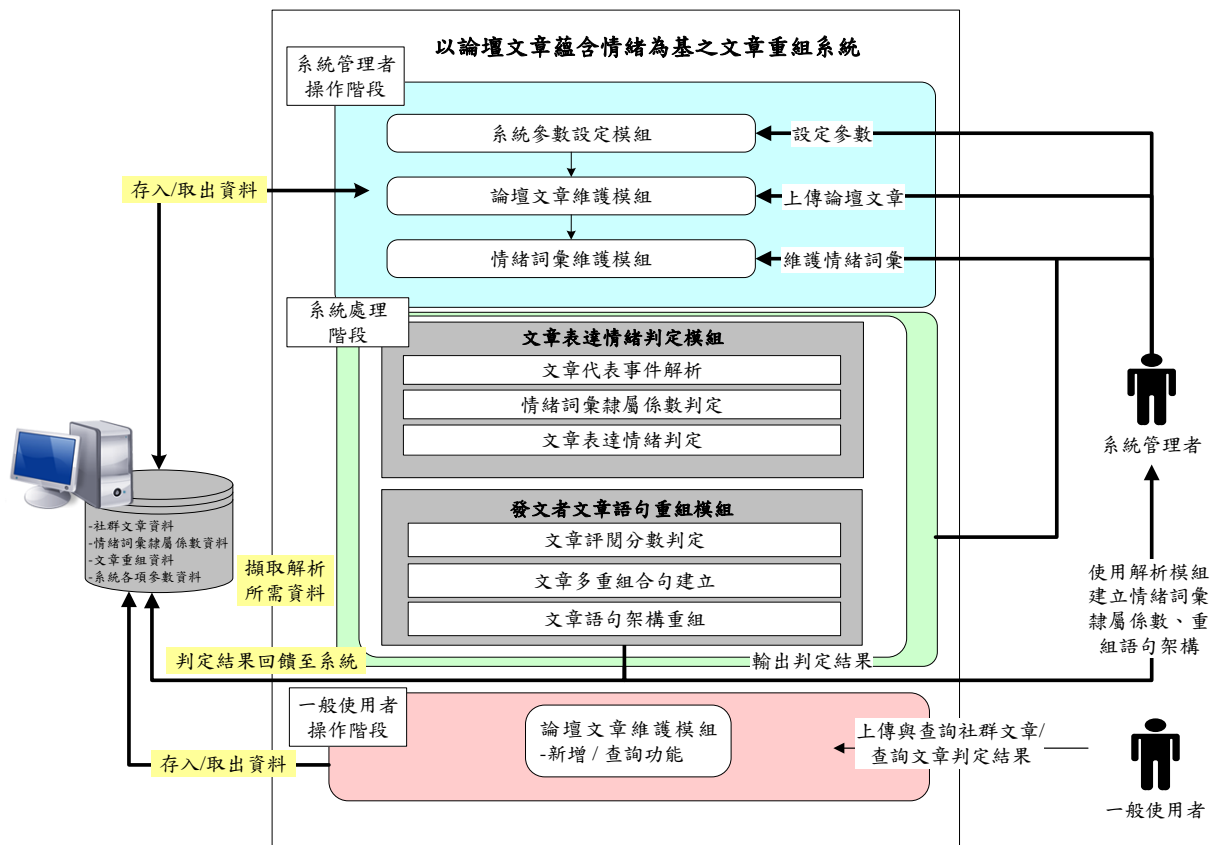


圖 4.3、以論壇文章蘊含情緒為基之文章重組系統運作架構

4.3 資料模式定義

本研究發展之以論壇文章蘊含情緒為基之文章重組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，期使論壇資料管理、文章表達情緒解析與文章語句重組等任務可即時完成。依據系統運作之需要，將以論壇文章蘊含情緒為基之文章重組系統之資料分為「論壇文章維護資料」、「情緒詞彙維護資料」、「文章表達情緒判定資料」、「發文者文章語句重組資料」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) 論壇文章維護資料

此資料之目的乃記錄論壇文章相關資料內容及解析時所需關鍵資訊，包含斷詞內容、文章

標題及詞性等解析資料，其所屬之資料表及其相關定義說明如下：

- 論壇文章基本資料表：記錄論壇文章之基本資料，包含文章標題、內容、斷詞標題與內容、文字字數、作者等資料。
- 候選事件資料表：記錄候選事件內容及詞性種類資料。
- 候選事件/論壇文章資料表：記錄論壇文章與候選事件之關係，以得知論壇文章內所含之候選事件數與內容。
- 語句順序資料表：記錄候選事件於論壇文章內之語句順序資料。
- 詞性種類資料表：記錄候選事件所屬之詞性。

(二) 情緒詞彙維護資料

此資料之目的乃記錄情緒詞彙資料內容及解析時所需關鍵資訊，包含情緒詞彙與情緒類別之隸屬係數資料，其所屬之資料表及其相關定義說明如下：

- 情緒詞彙資料表：記錄論壇情緒詞彙之編號及名稱資料。
- 情緒類別資料表：記錄情緒類別之編號及名稱等資料。
- 情緒詞彙/類別資料表：記錄情緒詞彙與情緒類別之關係，即包含情緒詞彙對應情緒類別之隸屬係數值，以及判斷情緒詞所屬情緒類別。

(三) 文章表達情緒判定資料

此資料之目的乃記錄文章之代表事件及表達情緒相關資料，包含候選事件、前後文、標題、詞性相關性分數、情緒詞彙與情緒類別之隸屬係數、以及當中所需之關鍵資訊，以有效進行分析資料之管理，以及文章表達情緒之解析，其所屬之資料表及其相關定義說明如下：

- 詞性種類資料表：記錄中研院詞類標記集中詞性之名稱、英文代稱等資料。
- 候選事件與論壇文章資料：記錄論壇文章中各候選事件之前後文配對、標題加權及詞性相關性分數。
- 情緒語句資料表：記錄含有情緒詞彙之語句編號及內容。
- 情緒語句/詞彙資料表：記錄情緒語句中所含之詞彙資料。
- 代表語句資料表：記錄文章代表語句之內容及編號。
- 相似語句資料表：記錄與代表語句相似之語句與內容。
- 文章代表/相似語句資料表：記錄代表語句與相似語句之相似關係資料。
- 代表語句/情緒類別資料表：記錄代表語句與情緒類別之關係。

(四) 發文者文章語句重組資料

此資料之目的乃記錄文章之評閱分數及語句架構相關資料，包含相異詞語數、候選填充句、代表語句英文延伸定義、語句重組架構、以及當中所需之關鍵資訊，以有效進行分析資料之管理，以及語句架構之重組，其所屬之資料表及其相關定義說明如下：

- 論壇文章評閱資料表：記錄論壇文章之評閱分數資料，即文章所對應之評閱分數等資料。
- 評閱門檻/批次資料：記錄各批次訓練數與評閱門檻之分數資料，以準確取得評閱分數之門檻。
- 批次資料表：記錄各批次訓練文章數之資料。
- 評閱門檻資料表：記錄各批閱門檻之名稱及所對應之門檻值資料。
- 文章/評閱門檻資料表：記錄文章與各評閱門檻之關係資料。
- 重要語句資料表：記錄語句詞義消歧後所取得之重要語句。
- 文章/重要語句順序資料表：記錄重要語句於文章之順序，其順序乃以標點符號為間隔以區分順序。
- 填充句資料表：記錄候選填充句與訓練文章之關係（如語句順序）。
- 重要語句/詞彙資料表：記錄重要語句與詞彙之關係。
- 詞彙資料表：記錄重要語句中詞彙之中英文內容。

(五) 系統參數設定資料

此資料之目的乃記錄系統參數之資料，即評閱分數門檻值資料，透過設定以有效判斷文章之語意表達程度；其所屬資料表及其相關定義如下：

- 系統參數資料表：記錄門檻值編號、名稱、門檻值說明及當前門檻值，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，使本研究所發展之以論壇文章蘊含情緒為基之文章重組系統可方便地進行論壇文章資料，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、以論壇文章蘊含情緒為基之文章重組系統之資料關聯

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「論壇文章維護模組」、「情緒詞彙維護模組」、「文章表達情緒判定模組」、「發文章文章語句重組模組」及「系統參數設定模組」等五大模組，以下及說明各系統功能之流程規劃。

論壇文章維護模組

此模組可供權限內使用者上傳訓練文章，以作為文章表達情緒判定及發文者文章語句重組模組之分析資料。此外，權限內使用者亦可於系統中維護論壇文章資料，以維持系統內資料之正確性，各項子功能包含新增、查詢、修改與刪除等子功能，以提供權限內使用者進行相關資料之維護動作；其流程設計概念如圖 4.5 所示。

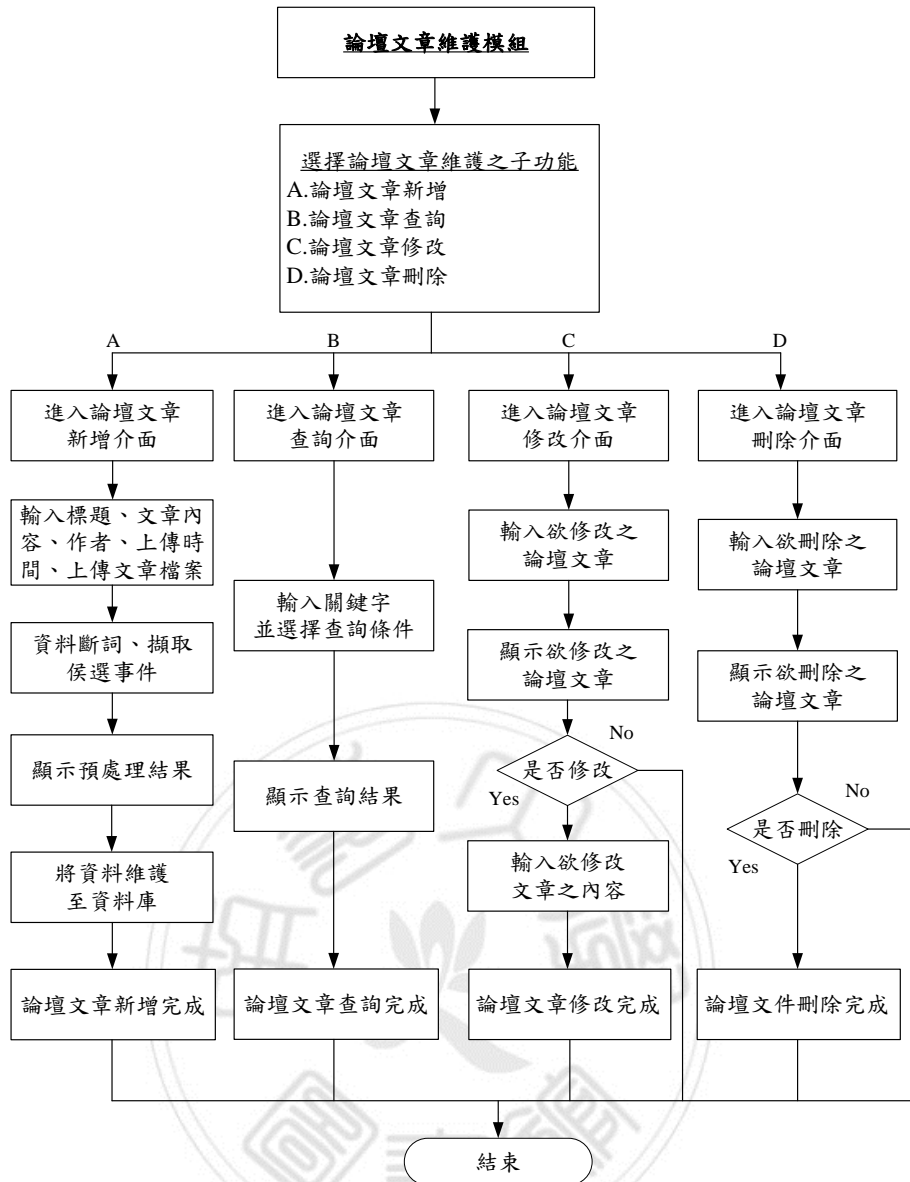


圖 4.5、論壇文章維護模組之功能流程

情緒詞彙維護模組

此模組可供權限內使用者上傳情緒詞彙資料，以作為文章表達情緒判定模組之分析資料，權限內使用者亦可透過情緒詞彙新增、查詢、修改及刪除等功能，維護情緒詞彙之相關資料，以維持詞彙於系統內之準確性與正確性；其流程設計概念如圖 4.6 所示。

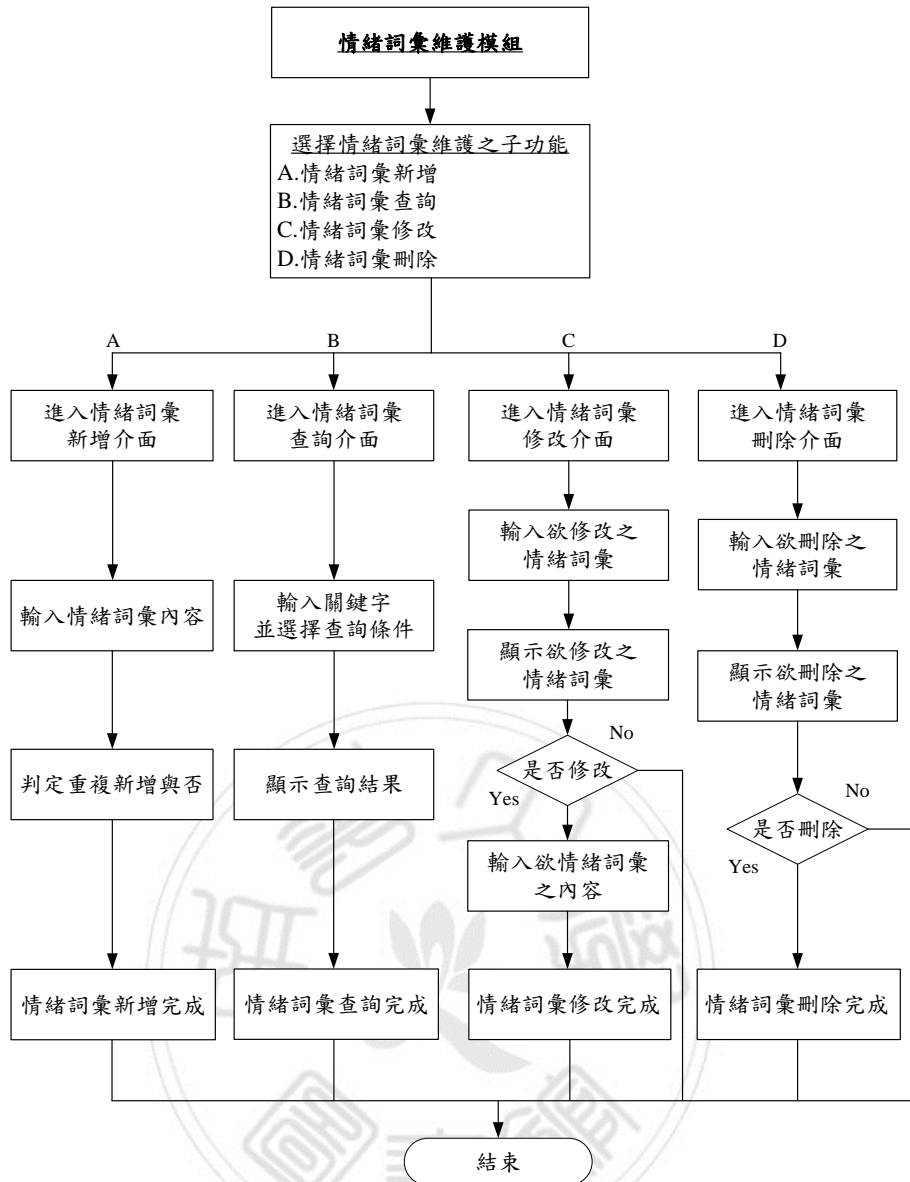


圖 4.6、情緒詞彙維護模組之功能流程

文章表達情緒判定模組

本模組乃劃分為「文章代表事件解析功能」、「情緒詞彙隸屬係數判定功能」及「文章表達情緒判定功能」。首先，於「文章候選事件解析功能」乃藉由斷詞將文章所有詞句進行分詞，並從文章中挑選具專有名詞或地方詞之代表詞彙作為「候選事件」，系統透過文章各候選事件之前後文分佈比例並取平均值以取得候選事件之前後文配對分數，同時根據候選事件與標題之關係計算標題加權分數，最後，依據文章之詞性種類數分析各候選事件所含詞性之分佈比例以獲取詞性相關性分數，系統即可將各候選事件之三項比重分數相加並以最大值者作為文章之代表事件。其整體運作流程如圖 4.7 左半部所示。

於「情緒詞彙隸屬係數判定功能」中，首先，利用基礎情緒詞建構對應之情緒語句，系統將統計情緒詞彙於各情緒語句之詞頻數，以初步取得情緒詞彙與各類別之關係係數

後，系統即會將情緒詞彙予以正規化並取得情緒詞彙與各類別之類別隸屬係數。整體運作流程如圖 4.7 中間部分所示。

於「文章表達情緒判定功能」中，首先，利用向量空間相似法計算代表語句與訓練語句之關係，以獲取與代表語句相似之語句，次之，系統透過相似語句內情緒與重要詞彙所組合之極性項目，取得與文章較相關之相似語句，系統並可針對此相似語句計算情緒偏向機率值，最後，再根據所有代表語句所含之相似語句計算情緒穩定值，即可取得代表語句之情緒類別並以此作為文章之情緒類別；整體運作流程如圖 4.7 右半部所示。

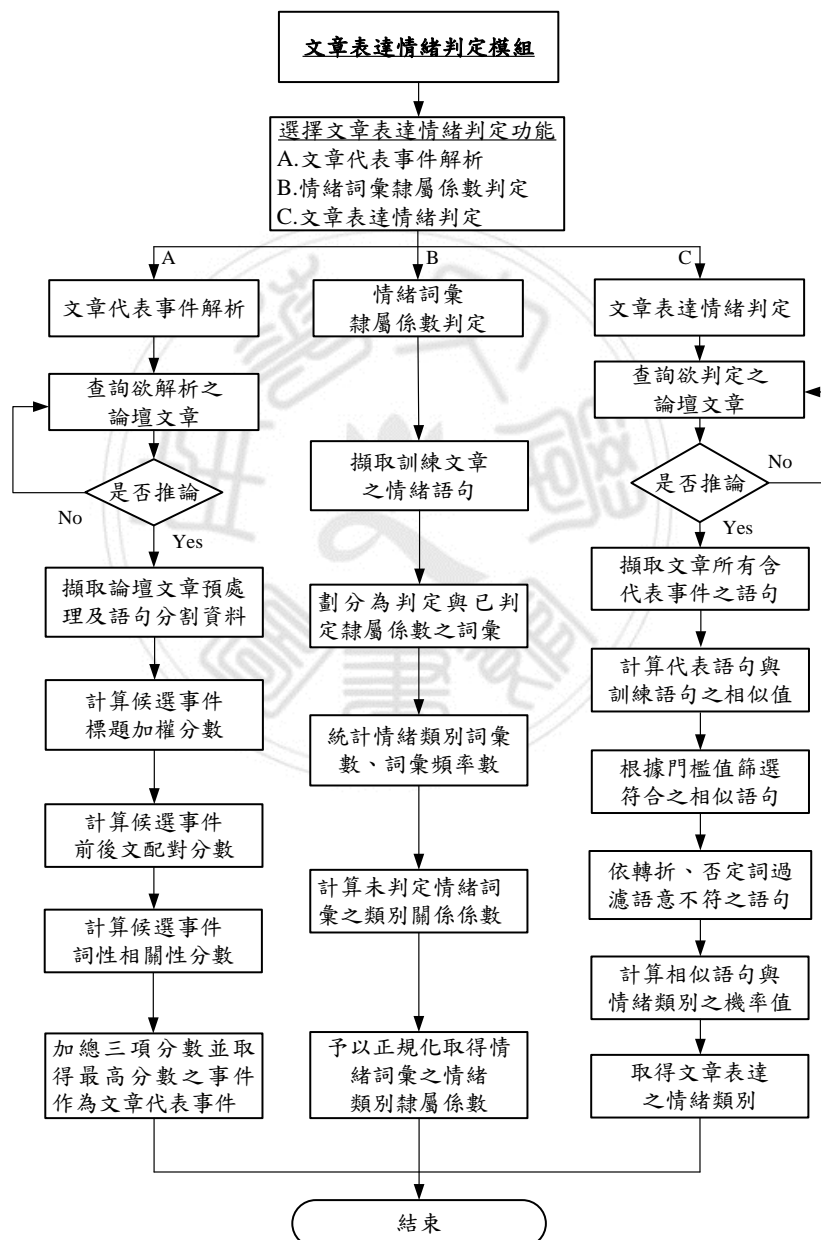


圖 4.7、文章表達情緒判定模組之功能流程

本模組乃劃分為「文章評閱分數判定功能」、「文章多重組合句建立功能」及「文章語句架構重組功能」。首先，於「文章評閱分數判定功能」中，系統乃去除文章重複出現之詞語，系統將相同之詞語視為 1 種詞語種類以取得初始分數，接著，系統根據目標論壇文章與訓練文章之共用詞語分析相似標準差，再依據共用詞語之語意差異性，解析目標論壇文章之評閱分數，並依據五種門檻值判斷目標論壇文章語句重組之合適性；其功能整體運作流程如圖 4.8 左半部所示。

而於「文章語句架構重組功能」中，系統利用中英翻譯器取得語句中各詞性詞彙之英文詞彙，並依據 WordNet 建立英文詞彙之延伸定義，系統即會計算各語句中英文詞彙之歧義相似性，以將相似之語句去除，接著，系統乃標定論壇文章之語句位置，並計算文章語句與訓練語句之相似性，將相似程度最高者作為候選填充句，同時系統將去除不與論壇文章原先表達情緒一致之候選填充句，最後，系統將候選填充句之關鍵詞與訓練文章進行比對，以替換與候選填充句之關鍵詞，進而形成論壇文章之多重組合句；其功能整體運作流程如圖 4.8 中間部分所示。

最後於「文章語句架構重組功能」中，系統將依據文章語句之順序組合多重組合句，以形成數個文章語句架構，為能從中取得語句架構較完整之組合，系統將重新評閱各語句架構之分數，並以最高評閱分數之語句架構，作為論壇文章語句重組後之內容；其功能整體運作流程如圖 4.8 左半部所示。

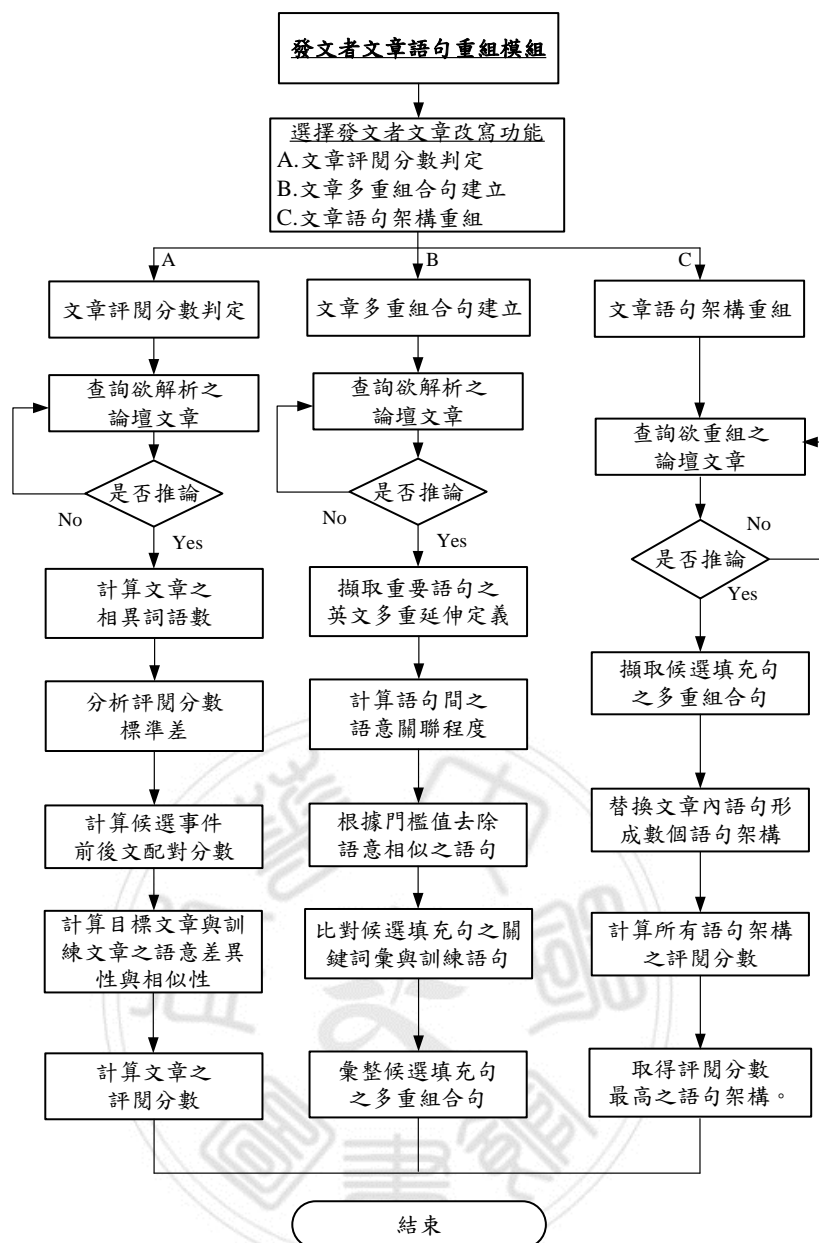


圖 4.8、發文者文章語句重組模組之功能流程

4.4.2 系統資料流程

本系統運作之初，系統管理者與一般使用者將未解析之論壇文章匯入系統，並進行基礎情緒詞彙之建立，待系統管理者完成情緒詞彙庫之建置後，即可執行文章表達情緒判定推論，以取得文章代表事件、情緒詞彙隸屬係數之建立以及文章表達情緒之推論。爾後，系統管理者須先針對文章評閱分數門檻值參數進行設定，方可執行發文者文章語句重組推論，系統會先行根據評閱門檻判斷文章語句重組之必要性，之後透過候選填充句之相似語句建立多重組合句，再根據多重組合句之重組順序形成數個文章語句架構並重新判定評閱分數，最後，將評閱分數最高之語句架構存儲置系統資料庫中。上述步驟完成後，一般使用者即可於論壇文章查詢功能進行查詢文章之最終結果，本系統相關資料之存取與傳遞情

形如圖 4.9 所示。

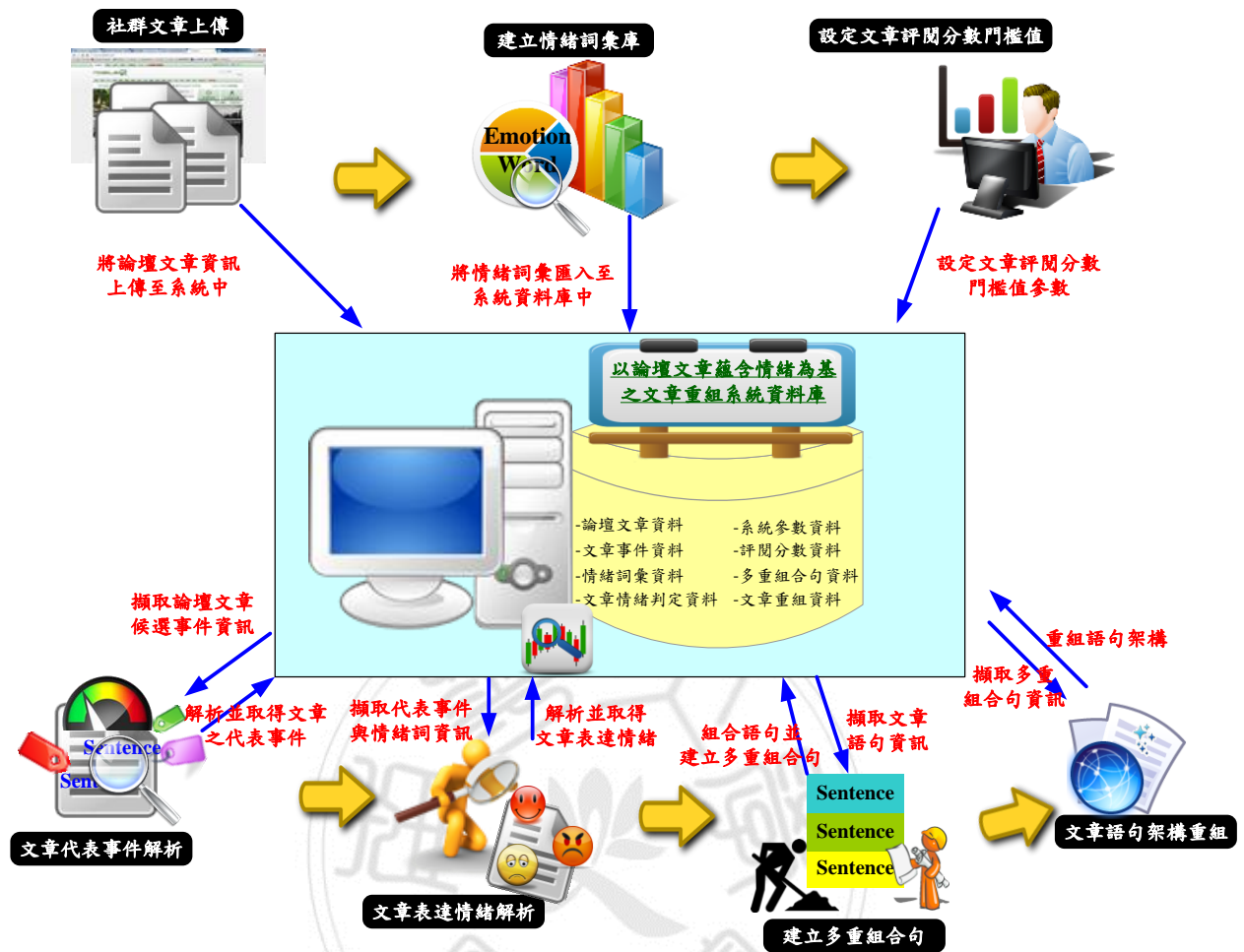


圖 4.9、系統資料流程

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、中研院斷詞系統，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執行後所產生之之是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 7，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，此些表格通稱為資料表。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。

➤ CKIP Client 中研院斷詞系統客戶端

CKIP係由中央研究院詞庫小組所開發之斷詞系統，其乃分析並標記文章未知詞並包含未知

詞之斷詞標記動作，完成分析之文章將呈現未知詞列表以及斷詞標記結果。該系統除了提供線上展示網站外，亦提供客戶端一個簡易、小量的斷詞服務，使使用者較易處理大量文章之斷詞。

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，本研究乃發展一套以論壇文章蘊含情感為基之文章重組系統，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中，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，於「文章表達情緒判定模組」中，本研究乃以「奇摩新聞」文章作為「系統自我績效評估」之驗證與訓練文章，並借鑑 Tung 與 Lu (2012) 所採用之驗證方式於「痞客邦」部落格中蒐集文章，以作為「比較研究」之驗證與訓練文章；於「發文者文章語句重組模組」中，本研究乃以「Mobile01 論壇」實際文章作為驗證與訓練文章之樣本，以驗證本研究發展方法論之可行性與績效，而所有驗證資料來源之可用性與嚴謹度乃闡述於 5.2 節中。

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「文章表達情緒判定」與「發文者文章語句重組」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，本研究分別以「奇摩新聞」、「痞客邦」論壇之真實文章內容，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，藉由本系統文章表達情緒判定與發文者文章語句重組等核心模組之運行，以評估本系統推論文章情緒類別與語句重組之正確性；於訓練文章樣本中，本研究則以上述三個論壇各領域類別之真實文章進行蒐集，透過本系統「情緒類別隸屬係數判定」功能之運行，透過訓練文章中情緒語句、情緒詞彙之關係，自動推論情緒詞彙之情緒類別隸屬係數，以作為判定文章表達情緒判定之解析基礎，此外，本研究亦以蒐集得之文章語句，形成文章之多重填充句以重組文章之語句內容。

首先，系統管理者必須蒐集訓練文章，將訓練文章上傳至系統中，以透過「情緒詞彙隸屬係數判定」功能建立情緒詞彙於情緒類別之隸屬係數。其次，一般使用者乃上傳論壇之真實文章內容至系統中。接著，系統管理者設定系統參數後，將執行第一核心模組，以判定文章之代表事件、推論情緒詞彙之隸屬係數與推論文章之情緒類別，待分析完畢後，系統即輸出相關判定結果輸出予使用者，並同時將此些結果回饋並儲存於系統中，接著系統管理者將可執行第二核心模組，系統將先行計算文章之評閱分數，再根據文章之情緒類別形成候選填充句以重組文章之語句內容，將推論結果回饋並儲存於系統中。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。

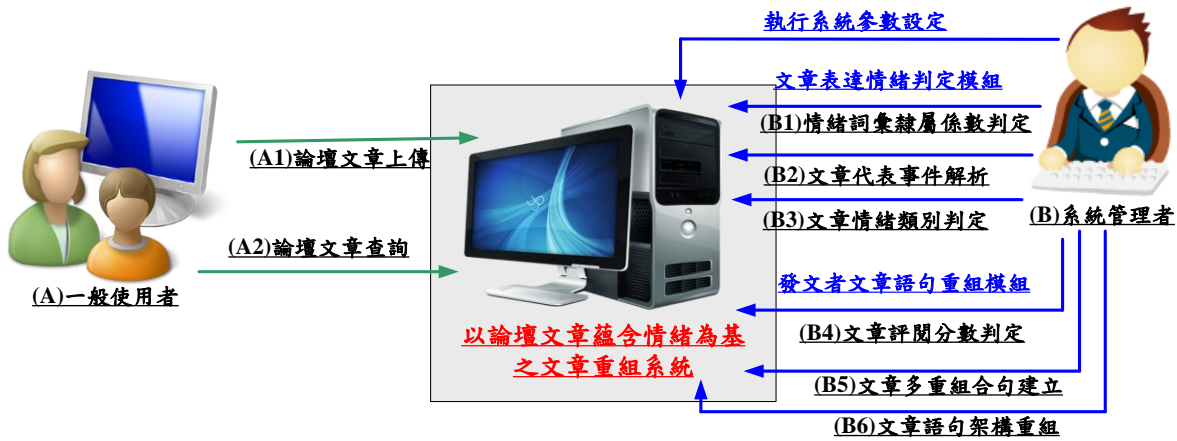


圖 5.1、以論壇文章蘊含情感為基之文章重組系統之應用流程

➤ 系統管理者蒐集訓練文章

開始執行本系統前，系統管理者須蒐集相關訓練文章，於此系統管理者針對「Mobile01」進行訓練文章之蒐集，如圖 5.2 至圖 5.4 所示，並於擷取訓練文章之文章內容後，輸入於系統中以透過「情緒詞彙隸屬係數判定」功能建立情緒詞彙庫，進而作為判定文章情緒類別之解析基礎，並能給於後續發文者文章語句重組模組建立多重組合句。

➤ 一般使用者蒐集論壇文章資料

一般使用者使用本系統前，先行蒐集未解析之論壇文章資料，以透過本系統幫助其推論文章之情緒類別並重組語句內容。於案例驗證中，本研究驗證方式乃蒐集「Mobile01 論壇」、「奇摩新聞文章」與「痞客邦文章」資料作為驗證資料樣本，由於兩模組與「發文者文章語句重組模組」之文章評閱分數功能皆可獨立運行，於此，本研究乃以 Mobile01 論壇之三個真實文章資料作為樣本資料，各別針對兩模組進行應用流程之說明，如圖 5.2、圖 5.3 與圖 5.4 所示。



圖 5.2、MOBILE01 論壇之樣本資料(1)



圖 5.3、MOBILE01 論壇之樣本資料(2)



圖 5.4、MOBILE01 論壇之樣本資料(3)

➤ 系統管理者執行參數設定

於執行系統核心模組之推論前，系統管理者乃需進行評閱分數門檻值設定功能，系統將呈現於畫面下方呈現修改介面並提供 5 種門檻值設定以供使用者選擇，如圖 5.5 所示，同時，以選擇「自行定義」門檻值並於文字欄中輸入門檻值為「0.5」並按下確定設定鈕後，系統即會呈現修改成功之提示，並可透過介面得知修改後之門檻值為「0.5」，如圖 5.6 所示。



圖 5.5、門檻值修改介面



圖 5.6、門檻值修改成功提示

(A1)一般使用者上傳未解析之文章資料

當一般使用者執行論壇文章新增時，可依據介面所提供之填入選項，依序填入論壇文章標題「台灣 FBI 犯罪心理關...等」、文章內容「這個頭銜是個很特別...等」、文章作者「unido」於介面中，同時選擇上傳時間「2014 年 03 月 15 日」，並將選擇欲上傳之「*.mht 檔」論壇文章，一般使用者即可按下確認鈕，如圖 5.7 所示，之後，系統乃自動將文章內容進行斷詞預處理，並顯示文章斷詞內容為「這(Nep) 個(Nf) 頭銜(Na).....等」，系統即會顯示新增成功之提示於介面中，如圖 5.8 所示。



圖 5.7、論壇文章上傳輸入介面



圖 5.8、論壇文章上傳成功介面

➤ 文章表達情緒判定模組

「文章表達情緒判定」功能主要目的乃針對一般使用者所上傳之論壇文章，先行推論情緒

詞彙之隸屬係數，再根據代表事件與情緒語句推論文章之情緒類別，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。

(B1)情緒詞彙隸屬係數判定

於推論文章之情緒類別前，系統管理者需先執行「情緒詞彙隸屬係數判定功能」以能建立情緒詞彙庫。首先，系統將區分已判定情緒類別隸屬係數之詞彙個數為「7」，未判定詞彙數為「8」，並依據情緒類別之詞彙數、詞彙於情緒類別語句之頻率數，計算情緒詞彙「驚詫」於情緒類別「恐懼」、「驚奇」、「厭惡」與「焦慮」之關係係數，分別為「0.1」、「0.77」、「0.3」、「0.2」，而剩餘3個情緒類別之關係係數皆為「0」，如圖 5.9 所示，最後，系統乃將詞彙之關係係數正規化後即可得知情緒詞彙「驚詫」最大值類別隸屬係數為「0.56」，所對應之情緒類別為「驚奇」，使用者亦可點選「查看」鈕得知詞彙「驚詫」於各情緒類別之隸屬分佈情況，如圖 5.10 所示。



圖 5.9、各情緒詞彙之關係係數解析結果



圖 5.10、各情緒詞彙與類別隸屬係數解析結果

(B2)文章代表事件解析

待使用者完成論壇文章上傳動作，系統管理者進入文章事件解析功能後，系統即計算目標文章相異候選事件個數為「4」個，同時得知候選事件「台灣」於標題中出現「1」次後，系統即會計算候選事件「台灣」於語句順序「3」、「7」、「9」、「17」、「26」、「29」中標題加權分數皆為「0.16」，如圖 5.11 所示，接著，計算候選事件「台灣」、語句順序為「3」中前後語句詞彙之分佈值，分別取得「0.18」與「0.22」並相加兩分佈值形成此候選事件之前後文配對分數「0.4」，如圖 5.12 所示，爾後，系統自動取得詞性參數值為 6，並擷取語句順序「3」、候選事件為「台灣」之前 6 個詞彙分別為「對」、「一樣」、「職業」、「個」、「這」、「婦」，同時系統將會計算此 6 個詞彙於所有候選事件前文之分佈平均值為「0.24」，如圖 5.13 所示，待完成三項分數之解析後，系統即會將候選事件「台灣」依據語句順序所取得之三項總分分別為「0.61」、「0.59」、「0.91」、「0.87」、「0.69」、「0.72」，同時將此分數加總平均取得「0.73」，並將此值予以正規化後即可得知候選事件「台灣」之代表分數為「0.196」，其中候選事件「犯罪」之代表分數「0.459」為最大值，故可得知「犯罪」為文章之代表事件，如圖 5.14 所示。



圖 5.11、候選事件標題加權分數解析介面



圖 5.12、候選事件前後文配對分數解析介面



圖 5.13、候選事件詞性相關性分數解析介面



圖 5.14、文章代表事件解析介面

(B3)文章表達情緒判定

待完成代表事件解析後，系統乃分析文章中具有代表事件「台灣」之語句有「對台灣來講都很陌生」、「可喜可賀也就是說台灣變態殺人狂很罕見」等六個語句，如圖 5.15 所示，接著，系統利用向量空間模型之餘弦函數，計算代表語句「對台灣來講都很陌生」與相似語句「而且被害人有特定連結但不認識的陌生人」之相似值為「0.31」，如圖 5.16 所示，之後，系統首先判斷相似語句「可喜可賀也就是說台灣變態殺人狂很罕見」中是否存在轉折詞與否定詞，並依據否定值參數值「2」及「轉折詞存在與否」等篩選條件保留相似語句「可喜可賀..」，同時根據相似語句「可喜可賀也就是說」中情緒詞與類別隸屬係數關係，分析該語句於「高興」情緒類別之情緒機率值為「0.53」、於「憤怒」之情緒機率值為「0.28」、於「恐懼」之情緒機率值為「0.19」，而其餘之情緒類別之情緒機率值皆為「0」，如圖 5.17 所示，最後，系統將依據代表語句「陳進興算是台灣「治安史」上唯一...等」對應之相似語句，計算「恐懼」情緒類別之情緒穩定值為「-0.726」、「驚奇」情緒類別為「-0.202」、「焦慮」情緒類別為「-0.152」，並以最趨近於 0 之穩定值「-0.152」得知代表語句「陳進興算是台灣「治安史」上唯一...等」之情緒類別為「焦慮」，系統亦透過目標文章所有代表語句所對應之情緒類別，判定文章內容「這個頭銜或者說職業很特別吧...等」所表達之情緒類別為「焦慮」及「高興」，如圖 5.18 所示。

文章表達情緒判定
—取得文章代表語句之相似語句

步驟一 步驟二 步驟三

文章資訊

代表事件	語句總數
【台灣】	31

文章內容
這個領域或者說職業很特別吧！跟家政類這個職業一樣，對台灣來講都很陌生，也不存在的行業。但這2個職業在日本都有，職業需求成正比，可當可買也就是說台灣變態殺人狂很罕見，如果說有，陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人，只是一種職業的延伸，也就是去救贖心，為了自保才有後續的犯罪……「犯罪」心理分析～是針對連續殺人，而且被害人有特定連結但不認識的陌生人，足以影響一般民眾的恐懼，所以這個人一定更想快刑，不僅會不斷有受害者增加，台灣的殺人犯都是有特定目標的行為，基於衝動或預謀，一般安份守己的老百姓不會成為被害者或在生活中形成一種恐懼，如：殺父母、[繼續閱讀](#)

取得代表語句

代表語句資訊

語句順訊	語句內容 (藍色字體為代表事件)
3	對台灣來講都很陌生
7	可當可買也就是說台灣變態殺人狂很罕見
9	陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人
17	台灣的殺人犯都是有特定目標的行為
26	我們台灣社會整體還算健全
29	不然台灣有一天也會發生像美國

取得代表語句之相似語句 上一頁

圖 5.15、取得文章語句中具有代表事件之語句

文章表達情緒判定
—取得文章代表語句之相似語句

步驟一 步驟二 步驟三

代表事件

代表事件	語句總數
【台灣】	31

系統相似參數值 0.2

參數說明
「相似門檻值」為代表語句與相似語句相似程度之門檻參數，用以篩選取得與代表語句相似之語句。

分析相似語句

相似語句解析結果 (僅列出高於門檻值之相似語句)

代表語句	相似語句	相似值
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	可當可買也就是說台灣變態殺人狂很罕見	0.25
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	對台灣來講都很陌生	0.25
對台灣來講都很陌生	而且被害人有特定連結但不認識的陌生人	0.31

圖 5.16、取得文章代表語句之相似語句

文章表達情緒判定
—解析相似語句之情緒機率值

步驟一 步驟二 步驟三

篩選相似語句

代表事件	相似語句數
【台灣】	3

否定詞參數值 2

參數說明
「否定詞參數值」為篩選與代表語句語意相反之相似語句，當否定詞出現數大於參數值則去除該相似語句。

相似語句	否定詞數	是否有轉折詞	是否保留
可當可買也就是說台灣變態殺人狂很罕見	0	無	✓
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	1	無	✓
而且被害人有特定連結但不認識的陌生人	1	無	✓

分析情緒機率值

相似語句情緒機率值解析結果 (藍色字體為最大值)

相似語句	高興	悲傷	憤怒	恐懼	驚奇	厭惡	焦慮
可當可買也就是說台灣變態殺人狂很罕見	0.53	0	0.28	0.19	0	0	0
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	0	0	0.19	0.1	0	0	0.71
而且被害人有特定連結但不認識的陌生人	0	0	0	0	0	0	0

圖 5.17、取得相似語句之情緒機率值



圖 5.18、取得文章表達之情緒類別

➤ 發文者文章語句重組

「發文者文章語句重組」功能主要目的乃針對一般使用者所上傳之論壇文章，先行計算評閱分數並去除相似語句，再建立多重組合句以重組文章之語句架構。以下乃分別以步驟方式說明本功能之應用流程。

(B4)文章評閱分數判定

系統管理者執行「文章評閱分數判定」，於此本研究乃以文字數較長之文章為例，並選定標題「台灣這些球員還嫌看球的觀眾太多嗎...等」、內容「2012 觀護盃籃球賽是第 10 季 SBL 熱身賽，雖然戰績...等」之文章，系統即分析文章中不同之詞語數有「172」個，同時根據訓練文章之相異詞語數計算評分標準差為「10」，如圖 5.19 所示，之後權限內使用者即可點選分析文章評閱分數鈕後，系統將依據評分標準差計算目標論壇文章與訓練文章內容「最近親人都在問說...等」之語意差異程度為「-12.7」、共同詞語數為「29」、與訓練文章內容「他就說不高興那就要不要!分手阿...等」之語意差異程度為「12.7」、共同詞語數為「49」，最後，系統乃依據語意差異程度與共同詞語數計算目標論壇文章評閱分數為「264」，如圖 5.20 所示。



圖 5.19、文章初始分數計算介面



圖 5.20、文章評閱分數計算介面

(B5)文章多重組合句建立

系統管理者執行「文章多重組合句建立」後，系統乃分析目標語句之英文詞彙延伸定義集合與其他語句「小弟住的紐約天氣還算穩定」之交集詞彙為「Snowfall」，詞彙於文章出現頻率為3次，且文章英文代表名詞詞彙總數為21後，系統即根據上述數據分析得知目標語句「但這個禮拜下了一場雪」與重要語句「小弟住的紐約天氣還算穩定」之關聯程度為「0.65」，如圖 5.21 所示，接著，系統即會根據門檻條件將語意關聯程度「0.85」、語句內容為「但還是有餘雪結冰的狀況」予以去除，如圖 5.22 所示，之後系統將重要語句「但這個禮拜下了一場雪」與訓練文章庫所有語句進行相似性比對得知語句「但上個禮拜下了一場大雪」之相似度為「0.707」，該語句之相似度為所有語句中最高值，因此系統將判定該語句為重要語句「但這個禮拜下了

一場雪」之候選填充句，如圖 5.23 所示，待取得候選填充句後，系統乃依據候選填充句「小弟住的紐約天氣還算穩定」之關鍵詞彙「天氣、紐約、穩定」，與訓練文章庫比對相符之語句，將具有關鍵詞彙之語句「最主要的原因就是紐約當天的天氣不穩定」、「紐約天氣狀況不穩定」等五個組合句，彙集成候選填充句之多重組合句，如圖 5.24 所示，即完成目標文章標題為「下雪天開車，可怕的經驗」多重組合句之建立。

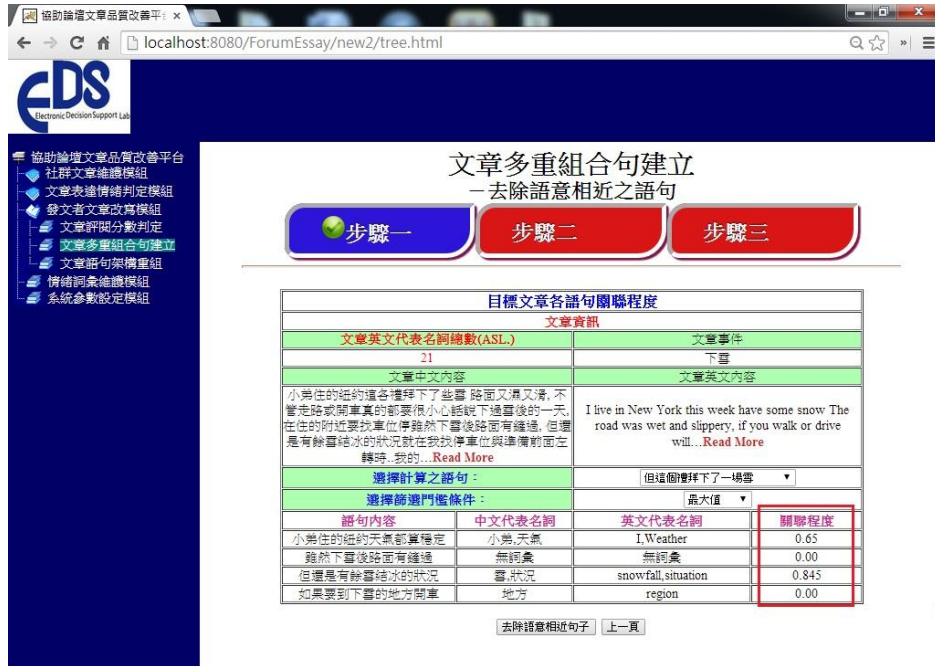


圖 5.21、取得文章各語句之關聯程度介面



圖 5.22、去除語意相似之語句介面

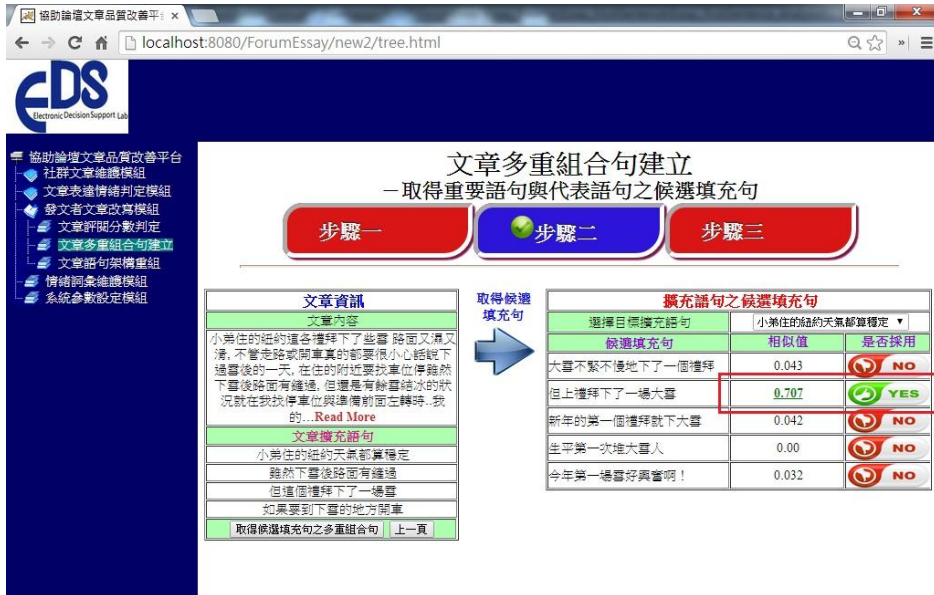


圖 5.23、取得代表及重要語句之候選填充句介面

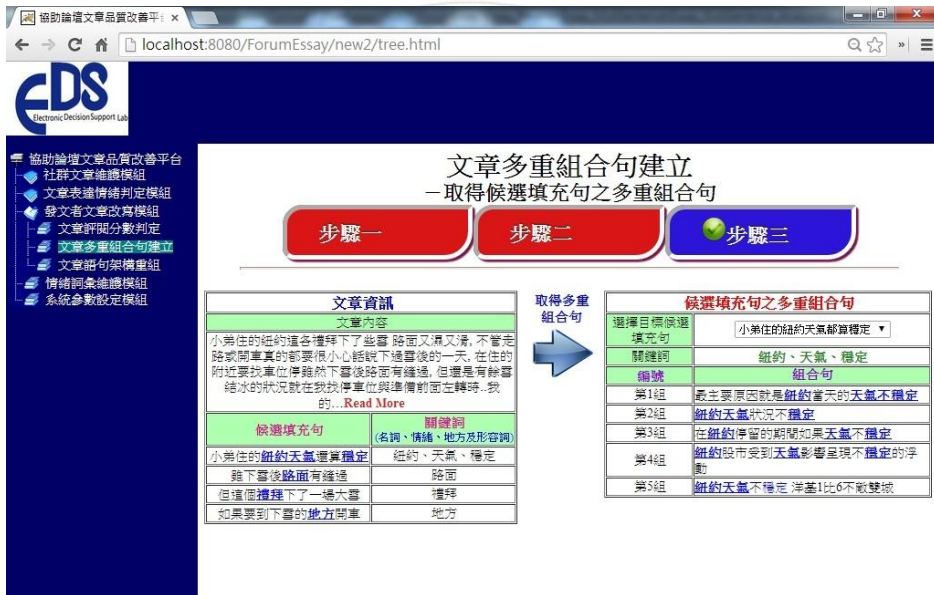


圖 5.24、取得候選填充句之多重組合句

(B6) 文章語句架構重組

完成文章多重組合句之建立後，系統首先取得目標文章內容為「小弟住的紐約天氣都算穩定，但這個..等」之欲替換語句共四句，同時根據系統於前一功能所取得之多重組合句，當中包含「但這個禮拜下了一場雪」有其他 5 組組合句、「如果要到下雪的地方開車」有其他 1 組組合句，得知文章語句共有「12」種不同之語句架構，當權選擇第 1 組組合時，系統即會將「但這個禮拜下了一場雪」替換為「但上禮拜下了場大雪」、「如果要到下雪的地方開車」替換為「如果要再下雪天開車」並將新的文章架構呈現於介面中，如圖 5.25 所示，而第 2 組新的文章架構內容如圖 5.26 所示，之後，權限內使用者即可執行最後步驟，並點選取得最高評閱分數之文章

鈕後，系統乃根據相異詞語數計算得知第 1 組文章架構相似程度標準差為「1138.01」、第 2 組為「1143.81」，並根據語意差異性、共用詞語數，計算得知所有語句組合中最高評閱分數為第 4 組「1183.62」，如圖 5.27 所示，最後，系統乃顯示第 4 組之語句架構，其語句架構乃為原文章重組後之最終輸出結果，其中，語句架構乃替換「下雪不緊不慢下了一個禮拜」及「如果要在下雪天開車」兩語句，而替換之語句乃與情緒不相關，因此系統判定後得知新文章架構之情緒類別為「恐懼」，並與原文章之情緒類相符，如圖 5.28 所示。

圖 5.25 顯示了系統的第一步操作。在「取得文章語句架構」面板中，顯示了目標文章的資訊：

文章字數	文章作者
406	henman

下方是「原始文章內容」的摘要，以及一個表格，列出了可供替換的語句及其數量：

語句順序	欲替換語句	可替換語句總數
1	小弟住的紐約天氣都算穩定	1
2	但這個禮拜下了一場雪	6
7	雖然下雪後路面有積雪	1
38	如果要在下雪的地方開車	2

在「新的文章語句架構」面板中，顯示了新的架構組合：

文章語句架構組合數	選擇組合
12組	第1組

新的文章語句架構（藍色語句為替換之語句）：

小弟住的紐約天氣都算穩定，**新年的第一個禮拜就下大雪**，路面又濕又滑，不管走路或開車真的要很小心，**話說下過雪後的一天**，在住的附近要找車位停，雖然下雪後路面有積雪，但還是有餘雪結冰的狀況，就在我找停車位與準備前面左轉時，我的車竟然打滑了...衝突，忘了當時的時速，感覺可能有 20 mph 左右吧...繼續閱讀

圖 5.25、取得文章多重語句架構-第 1 組

圖 5.26 顯示了系統的第二步操作。在「新的文章語句架構」面板中，顯示了新的架構組合：

文章語句架構組合數	選擇組合
12組	第2組

新的文章語句架構（藍色語句為替換之語句）：

小弟住的紐約天氣都算穩定，**但上禮拜下了一場大雪**，路面又濕又滑，不管走路或開車真的要很小心，**話說下過雪後的一天**，在住的附近要找車位停，雖然下雪後路面有積雪，但還是有餘雪結冰的狀況，就在我找停車位與準備前面左轉時，我的車竟然打滑了...衝突，忘了當時的時速，感覺可能有 20 mph 左右吧...繼續閱讀

圖 5.26、取得文章多重語句架構-第 2 組



圖 5.27、取得評閱分數最高之文章組合



圖 5.28、評閱分數最高之文章資訊與內容

5.2 系統案例驗證與評估

本研究乃以「以論壇文章蘊含情感為基之文章重組」為基，開發一套「以論壇文章蘊含情感為基之文章重組」系統，於「以論壇文章蘊含情感為基之文章重組」主題中乃包含「文章表達情緒判定模組」與「發文者文章改模組」兩子議題，故本研究將分別針對上述兩議題進行獨立之驗證，以確保兩議題之驗證結果不相互影響，以獲致更具參考性與之系統驗證與評估結果。

整體而言，本研究於 5.2 節之整體規劃如圖 5.29 所示。

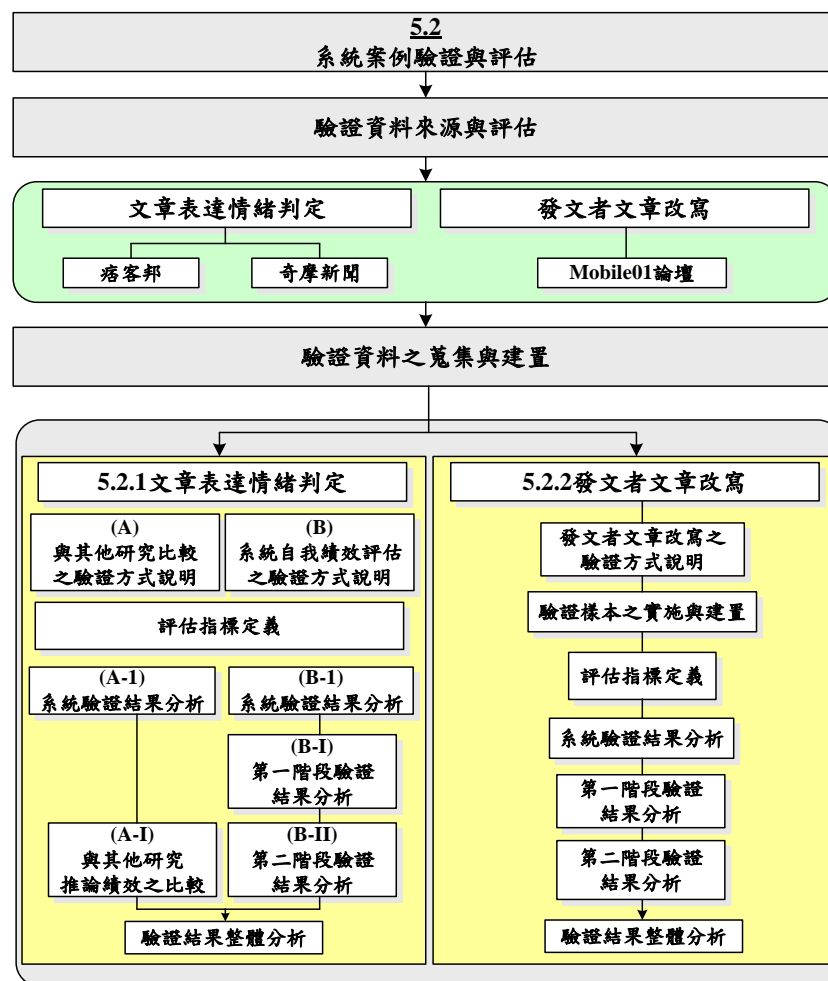


圖 5.29、系統驗證與評估之架構

驗證資料來源與評估

於「文章表達情緒判定」之驗證中，本研究乃分為「比較研究」與「系統自我績效評估」兩部分進行驗證。於「比較研究」中，本研究乃參考 Tung 與 Lu (2012) 所設計之驗證方式，以「痞客邦」之文章作為驗證資料與訓練文章；於「系統自我績效評估」中，本研究乃以「奇摩新聞」之文章作為驗證資料與訓練文章。於「發文者文章語句重組」之驗證中，本研究乃以「Mobile01 論壇」之文章作為驗證與訓練文章之樣本，以下乃針對「痞客邦」、「奇摩新聞」與「Mobile01 論壇」驗證與訓練文章之嚴謹度進行說明。

➤ 「痞客邦」資料之可信度與可用度：

本研究乃以台灣較大之部落格「Xuite 日誌」、「yam 天空部落」、「PChome 個人新聞台」與「痞客邦」部落格進行質化之比較，以評估「痞客邦」驗證資料之嚴謹度；此外，由於無名小

站已被奇摩所併購，而統計數據之日期為 2010 年 7 月、8 月與 10 月，故於此將排除「奇摩部落格」與「無名小站」部落格之比較，相關數據之彙整如表 5.1 所示。

表 5.1、各大部落格之質化比較（數據為 2010 年 7 與 8 月之統計結果）

比較項目	痞客邦	Xuite 日誌	yam 天空部落	PChome 個人新聞台
受訪 2721 人中部落格經營率	10.01%	9.8%	9.3%	4.7%
平均使用時間	15 分鐘	8.9 分鐘	10.3 分鐘	4.6 分鐘
具有站長制度	●	●	-	-
Alexa 統計台灣網站排名 (2014/11/03)	第 4 名	第 10 名	第 22 名	第 16 名
以關鍵字「台灣」搜尋文章數	17795623 筆	22655438 筆	665,000 筆	無統計數據

針對於上述質化比較言之，可得知「痞客邦」與「Xuite 日誌」係當中文章數較多之部落格，且兩者皆具站長制度，故作為驗證資料之來源，相較於其他部落格更為適切，此外，「Alexa」公司乃以網站每天平均使用人數、人均存取頁面數、與其他網站連結或曝光數、網友所留言討論的訊息篇數等資訊為基礎進行網站排名，雖然以關鍵字搜尋文章數之結果「痞客邦」乃小於「Xuite 日誌」，然而，「痞客邦」不論於 Alexa 統計、平均使用時間、部落格經營率等數據皆高於「Xuite 日誌」，是固，Tung 與 Lu (2012) 所使用之驗證資料「痞客邦」乃具備較高程度之嚴謹性與可用度，而本研究最終亦以「痞客邦」作為驗證資料，並與 Tung 與 Lu (2012) 於相同之驗證情境下比較推論結果。

➤ 「奇摩新聞」資料之可信度與可用度：

為了能釐清奇摩新聞之可用度與可信度，本研究乃以台灣具代表性之新聞網站進行質化之比較，如表 5.2 所示，就比較結果而言，於瀏覽人數之層面，「奇摩新聞」與「聯合新聞網」乃唯二瀏覽人數超過 500 萬人次之新聞網站；就可觀看投票情緒感受總人數之比較項目，雖然「奇摩新聞」與「聯合新聞網」兩者皆具備此制度，但「聯合新聞網」尚未具備防範「重複投票情緒感受」之機制，其他之新聞網站皆具備此機制；最後於「每位使用者平均瀏覽次數」項目中，奇摩新聞使用者平均瀏覽次數約高於其他 3 至 4 倍。針對上述質化比較言之，可發現「聯合新聞網」與「奇摩新聞」乃具備「投票情緒感受」之機制，但「聯合新聞網」尚未具備「重複投票情緒感受」機制之網站，故作為驗證資料之來源相較於其他新聞網，其統計之文章實際情緒感受將可能備受質疑並缺乏公正性，此外，「蘋果新聞」與「今日新聞」尚未具備「投票情緒感受」之機制，於此以兩此新聞網站作為驗證資料之來源，將無

法取得實際之情緒感受，而本研究乃考量驗證資料乃需具有「實際情緒感受」之文章，且所統計實際情緒感受數據乃具有公平性、透明化，是故，本研究乃於最終乃選定「奇摩新聞」作為驗證資料之來源。

表 5.2、各大新聞網站之質化比較（統計數據為 2013 年 5 月）

新聞網站 比較項目	奇摩新聞	聯合新聞網	蘋果新聞	今日新聞
瀏覽人數	1069 萬	534.7 萬	361.3 萬	328 萬
總瀏覽花費時間	1025 萬分鐘	約 73 萬分鐘	約 84 萬分鐘	約 88 萬分鐘
每位使用者平均瀏覽次數	約 22 次	4 至 5 次	6 至 7 次	4 至 5 次
每次瀏覽停留時間	3 至 5.5 分鐘	2 至 4 分鐘	3 至 4 分鐘	5 至 6 分鐘
可觀看投票情緒感受總人數	●	●	-	-
可觀看各情緒類別投票人數	●	-	-	-
可重複投票	-	●	-	-
為繁體中文	●	●	●	●

➤ 「Mobile01 論壇」資料之可信度與可用度：

本研究乃以台灣各大同性質之論壇網站如奇摩 3C 科技、T 客邦、癮科技等，與 Mobile01 論壇進行質化之比較（比較表如表 5.3 所示），當中，「Mobile01 論壇」與「T 客邦」乃具備管理者制度且一般使用者皆可於網站內自由發表文章；於文章類型而言，「Mobile01 論壇」與「T 客邦」乃具有使用者發表與新聞兩類型之文章，「奇摩 3C 科技」僅具有新聞文章，「癮科技」乃包含新聞文章與評測文章；而於「到達率」與「平均瀏覽時間」層面，「Mobile01 論壇」於兩比較項目皆高於其他論壇 3 至 5 倍。

針對上述質化比較而言，可得知「Mobile01 論壇」與「T 客邦」乃具備管理者審核制度，且文章乃兼具客觀（新聞文章）、主觀（使用者發表）之角度，因此，文章所闡述之內容較為全面性之觀點，故作為驗證資料相較於其他論壇較為適切，此外，針對上述兩網站數據之比較可發現，「Mobile01 論壇」不論於「到達率」、平均瀏覽時間皆高於「T 客邦」，且 Mobile01 論壇於所有討論主題、討論版內皆設有「管理者」機制，反之，T 客邦所有討論主題乃集中於一管理者團隊共同管理，較無法針對性逐一管理各大主題，因此，本研究最終決定以「Mobile01 論壇」作為「發文者文章語句重組模組」之驗證資料來源。

表 5.3、各大科技與 3C 產品討論論壇之質化比較

論壇名稱 比較項目	Mobile01	奇摩 3C 科技	T 客邦	癮科技
到達率	19.1%	6%	5.1%	3.8%
每位使用者平均瀏覽時間(分鐘)	27.1	3.9	4.8	5.3
具管理者制度	●	-	●	-
文章類型	使用者發表、 新聞文章	新聞文章	使用者發表、 新聞文章	新聞文章、評測文章
使用者可自由發表文章	●	-	●	-

驗證資料之蒐集與建置

有鑑於上述比較之結果，本研究於「文章表達情緒判定」中，比較研究之驗證資料乃選定「痞客邦」，而系統自我績效評估則選定「奇摩新聞」作為驗證資料之來源，此外，於「發文者文章語句重組」中，本研究乃選定「Mobile01 論壇」作為驗證與測試資料之樣本，以驗證本系統之可行性，而於驗證「文章表達情緒判定」系統績效前，本研究乃參考 Gregory 及 Daniel(2008) 所建構之 482 個情緒詞彙，對應至各情緒類別以形成「基礎情緒詞彙」，以使後續之判定能順利運行。

本研究乃於「痞客邦」中隨機蒐集部落客所發表之文章(如圖 5.30 所示)以作為訓練文章，於蒐集之過程中乃篩選字數過少之文章以使訓練文章更為確實；「奇摩新聞」乃存有數量龐大之新聞文章，以及各篇新聞皆可投票閱讀後之情緒感受等特性，是故，本研究乃隨機於各領域中蒐集投票情緒感受高於 30 人次之「新聞文章」資料(如圖 5.31 所示)，以使訓練文章之實際情緒感受更為準確；最後，本研究乃於「Mobile01 論壇」前文章數前五之主題「小惡魔的動力研究室」、「生活娛樂」、「相機與攝影」、「智慧型與傳統手機」、「個人電腦」中隨機蒐集訓練文章，實際文章如圖 5.32 所示，於蒐集過程中亦篩選字數過少之文章，以建構語句較為完整之「候選填充句集合」，進而於後續取得更具準確性推論與驗證之結果。



圖 5.30、痞客邦部落格之文章資料



圖 5.31、奇摩新聞之新聞文章



圖 5.32、MOBILE01 論壇之文章資料

5.2.1 文章表達情緒判定之驗證

針對「文章表達情緒判定」本研究乃分為「比較研究」與「系統自我績效評估」兩方式驗證系統之績效，並非別設計驗證指標以進行系統驗證分析。以下乃針對各驗證方式細述與說明驗證設計、指標定義及驗證績效與結果。

(A) 與其他研究比較之驗證方式說明

與其他研究之比較部分，本研究乃參考 **Tung 與 Lu (2012)** 之驗證方式，於「痞客邦」部落格中隨機蒐集 1300 筆驗證資料（如表 5.4 以 5 份實際文章為例），同時額外蒐集 500 筆訓練文章，並以人工方式判斷 1300 筆驗證資料之實際情緒類別，而情緒類別與比較研究相同分為「生氣」、「快樂」、「思考」、「害怕」、「悲傷」、「擔心」、「驚嚇」7 大情緒類別，爾後，將 500 筆訓練文章逐一匯入系統中，以使系統能自動建置與推論情緒詞彙之情緒類別係數；完成情緒詞彙隸屬係數之推論後，再逐一推論 1300 筆測試資料之情緒類別，最後，將推論結果與實際結果進行統計藉此得知系統推論之績效，再與 **Tung 與 Lu (2012)** 比較驗證之績效。

表 5.4、痞客邦實際測試文章（以 5 份為例）

編號	新聞文章標題	新聞文章內容	實際情緒類別
1	回到令人害怕的工作崗位	6 點就起床 中壢 陰冷的天氣 外面飄著毛毛雨捨不得離開溫暖的被窩~~還有溫暖的人工暖爐~~~再怎麼不捨得 還是要振作!在車上吃了我家大頭買的愛心飯團和熱豆漿	害怕
2	不說才最難過傷心	自己也不清楚到底怎麼回事，許多解釋都很像藉口累積了許多情緒，又會害怕以文章的方式面對，太清楚了，即使很希望也很喜歡這個出口，因為真的太清楚了	悲傷
3	嚇死媽媽的耳朵發炎	Buffy 病了一個禮拜整整從禮拜天 2/16 一路發燒到今天都禮拜五了還是微微燒最高燒的時候我以為是不是我的耳溫槍壞了??!居然是 41.7C	驚嚇
4	令人擔憂的老弟	這週我回家渡假沒有趕上預定的火車有點失望但是真正讓我想扁人的事老弟跑步跑到暈倒了我印象中之前醫生明明交代過不能跑的丫雖然他只跑了兩百	擔心
5	猶豫	今天仔細思考考試這件事本來很想考都市計畫類的考試但因為沒上過相關方面的課想利用在學校工作的時間再回去學校上點課	思考

(B)系統自我評估之驗證方式說明

首先，本研究乃於「奇摩新聞」中隨機蒐集 1000 筆訓練文章，並隨機挑選 20 份奇摩新聞文章作為測試資料（如表 5.5 所示），當中測試資料皆高於 30 人次投票閱讀後情緒感受之文章，此外，本研究乃依據奇摩新聞網站所分類之情緒，將情緒類別分類為「火大」、「害怕」、「無聊」、「超扯」、「開心」、「感人」、「難過」等 7 大類別，當中情緒類別與情緒代號如表 5.6 所示。接著，本研究乃將系統驗證過程規劃為兩階段，於系統驗證第一階段乃於 1000 筆訓練文章中隨機挑選 200 筆匯入至系統中，並針對 20 份測試資料進行推論，以觀察系統初期之判定績效，進而觀察本研究所提方法論之正確性。待完成上述之第一階段系統績效驗證後，即進行系統測試第二階段之驗證，於此階段乃分為 8 個週期，每週期持續匯入 100 筆訓練文章資料（共計 800 筆），藉由持續匯入以分析系統於不同訓練測試資料下之推論結果，於各個週期乃利用前述 20 份測試資料重新進行推論，以分析系統之長期學習趨勢。

表 5.5、奇摩新聞之 20 份實際測試文章

編號	文章標題	文章內容	實際情緒類別
1	子孫為領 18 趴 老父變活死人	(中央社記者陳清芳台北 6 日電)陽明大學附設醫院醫師陳秀丹今天說,有位老校長 7、8 年來靠呼吸器飽	難過
2	《阿嬤 搵巾》情深無怨	「人有兩撇好寫,人卻難做!」64 歲婦人郭秀酌為了照顧患有多重身障的 12 歲孫女「奕妍」,從出生第一天	感人
3	8 年殺 3 女 竟要母別賠償	[自由時報記者楊國文/台北報導]有思覺失調症病史的陳姓男子屢稱「心中有惡魔叫我去殺人」,先殺害國	火大
4	女遭男友割五官 轉送林口長庚	黃姓老翁狠心割下女友祝姓女子五官,傳出祝女是自願被割的說法,不過檢警對此不願證實	害怕
5	3 身分 馬國男混跡台灣 28 年	另類《神鬼交鋒》電影版!馬來西亞籍夏姓男子,32 年前來台讀大學,畢業後逾期居留被驅逐出境;隔年以	超扯
6	38 公斤還減肥 30 歲女骨質疏鬆	30 歲張姓女子體重才 38 公斤,她長期防曬、減肥,一次穿高跟鞋下樓梯時扭傷,就醫發現竟骨折,才知自己	超扯
7	1 台幣爽換 3.7 日圓 遊日樂透	你換日圓了嗎?日本央行擴大寬鬆政策壓垮日圓,昨日貶破 1 美元對 114 日圓,對新台幣交叉匯率下殺至	開心
8	大贏韓國 中華隊 3 連勝晉級	21U 世界盃棒球賽昨上演傳統台韓戰戲碼,中華隊在「亞運雙雄」郭俊麟、王柏融投打都有好表現帶領下,	開心
9	加油別急~汽、柴油 周一起每公升降 0.5 元、0.6 元	中油公司今(16)日宣布,依浮動油價機制計算,將自明(17)日凌晨零時起調降國內油價,汽油每公升調降 0.5	開心
10	艾怡良整眼鼻下巴 大方認再進化	艾怡良出道 4 年,最近推出第 2 張專輯《大人情歌》,長相和當年參賽《超級偶像 5》判若兩人,和 2012 年	無聊
11	10 歲女童獨自顧阿祖 3 人窩 1.5 坪閣樓	高雄這位 10 歲的鄭姓女童,因為媽媽生病沒有同住,她從 2 歲就和兩位阿祖一起生活,住在 1.5 坪大的空間,	感人
12	中山高自小客翻覆 3 人送醫	一輛自小客車今天下午在中山高台南新營路段疑因爆胎翻覆,車中 3 人送醫,所幸多是輕傷無大礙。	難過
13	內灣老街大火 飲料店老闆葬身火窟	凌晨新竹內灣老街發生火警,一間古早味飲料店,突然冒出火舌,不到 5 分鐘的時間	難過
14	大漢板豆腐曾用永昌石膏粉 已預防下架	大漢傳統板豆腐煎、煮、滷、炸,還能做成麻婆豆腐、或加進火鍋裡,現在卻傳出製作大漢板豆腐的川武食品	火大
15	Makiyo 又喝酒 臉頰腫如麵龜	藝人 Makiyo 在 2 年前,因酒醉毆運將一事成為全民公敵,演藝工作也跌落谷底,她承諾遠離酒類、常做公益	無聊
16	男當街掀女衣還狠踹肚 路人經過沒阻止	街頭上演全武行!網友 15 日在臉書貼出一段影片,其中,疑似一對情侶當街爭吵,男子不但粗魯得將女子衣	火大
17	SJ 晟敏喜帖曝光 喊話「請務必祝福我們」	韓國夯團 Super Junior 成員晟敏,9 月被爆料戀愛 ING,緊接著他馬上在 10 月承認,12 月 13 日將迎娶女友金思	無聊
18	抗癌騎士環島 神采奕奕抵終站	廿六名來自全台各地及香港的抗癌勇士,昨天完成十天環島行程,抵達終點站新北市政府市民廣場,新北市副	感人
19	女友父要求買房 提親演出全武行	桃園縣游姓男子與古姓女友相戀論及婚嫁,去年 12 月間游男前往準岳父家討論婚事,卻因古父要求游男得買	超扯
20	天壽! 3 種貢丸、3 家烏骨雞吃不得	食品藥物管理署昨公布市售禽畜產品動物用藥殘留檢驗結果,共有 6 件違規,包括 3 件烏骨雞及 3 件貢丸,	火大

表 5.6、系統自我績效評估之情緒類別與情緒代號

情緒類別	情緒代號
火大	K1
害怕	K2
無聊	K3
超扯	K4
開心	K5
感人	K6
難過	K7

兩驗證方式之評估指標定義

藉由系統推論文章之情緒類別與實際情緒類別驗證系統之準確性，於此階段績效指標乃包含「文章表達情緒推論召回率」、「文章表達情緒推論正確率」與「整體推論正確率與召回率之協調值」，其參數定義與計算方式如公式(5.1)至公式(5.3)所示

文章表達情緒判定之召回率

文章表達情緒解析之召回率為一相對比例值，為「實際文章情緒類別與推論文章情緒類別之相符個數」與「實際文章情緒類別數」之比例，其相關變數定義如下：

- RE_i 系統推論文章情緒時，第 i 份文章之情緒類別推論召回率
 ES_i 第 i 份文章實際所屬情緒類別與推論情緒類別相符之個數
 AN_i 第 i 份文章實際所屬之情緒類別

$$RE_i = \frac{ES_i}{AN_i} \quad (5.1)$$

文章表達情緒判定之正確率

文章表達情緒判定之正確率為一相對比例值，為「實際文章情緒類別與推論文章情緒類別之相符個數」與「推論文章情緒類別數」之比例，期望藉由此項指標評估系統推論情緒類別與實際類別之差異程度。相關變數定義如下：

- PE_i 系統推論文章情緒類別時，第 i 份文章之情緒類別推論正確率
 ES_i 第 i 份文章實際所屬情緒類別與推論情緒類別相符之個數
 DN_i 系統推論第 i 份文章之情緒類別個數

$$PE_i = \frac{ES_i}{DN_i} \quad (5.2)$$

文章表達情緒解析推論正確率與召回率之協調值 (F-Measure)

文章表達情緒解析推論協調值為評估系統推論召回率與正確率之成效指標，為「系統正確率乘以召回率再乘以二」與「系統正確率加召回率」之比例，期望透過此項指標評估系統整體績效，其相關變數定義如下：

FM_i 系統推論文章情緒類別時，第 i 份文章之情緒類別協調值

$$FM_i = \frac{2 \times RE_i \times PE_i}{RE_i + PE_i} \quad (5.3)$$

(A-1) 與其他研究之比較-驗證結果分析

本研究乃與 **Tung 與 Lu (2012)** 於相同驗證情境與驗證資料來源之情況下，比較推論之績效，其驗證結果如下。

(A-I) 推論績效之比較

本研究於 500 份訓練文章之數量下，系統乃針對 1300 份測試資料（如表 5.4）進行情緒類別之判定，其中，本研究系統判定結果於 1300 筆測試資料中，推論正確實際情緒類別之個數為 1068，其召回率、正確率與 F 值皆為「82.1%」，而 **Tung 與 Lu (2012)** 推論之召回率、正確率與 F 值皆為「72.5%」，因此，本研究於相同驗證方式下，正確率、召回率與 F 值之推論數據皆略高於 **Tung 與 Lu (2012)** 近 10% 之績效，然而，本研究與 **Tung 與 Lu (2012)** 所使用之驗證資料、訓練資料與測試資料雖不盡相同，但本研究於驗證中所有之文章資料皆於部落格中以「隨機」方式所蒐集之，以此方式蒐集之資料，其系統推論績效與 **Tung 與 Lu (2012)** 比較時較符合公平性，藉由上述數據之比較可得知，本研究所提出之方法論與開發之系統，推論成效將略優於 **Tung 與 Lu (2012)**。本研究與 **Tung 與 Lu (2012)** 之推論數據之比較如圖 5.33 所示。

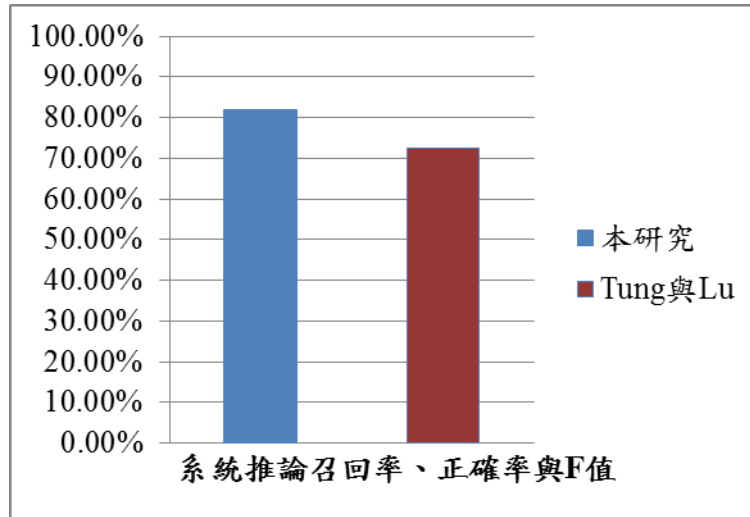


圖 5.33、本研究與 TUNG 與 LU (2012) 推論數據之比較

(B-1)系統自我績效評估-驗證結果分析

本研究乃將系統自我績效評估分為「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。於第一階段中乃先行匯入 200 筆訓練文章後，先行推論情緒詞彙之類別隸屬係數，再推論 20 筆驗證資料之結果，藉此觀察系統初期之推論績效；接著於第二階段中，本研究乃劃分為 8 個週期，於每週期匯入 100 筆訓練文章並重新推論 20 筆驗證資料之結果，以測試不同訓練量下系統進行「文章表達情緒判定」之績效。以下即針對各階段說明系統驗證過程與結果。

(B-1)第一階段驗證結果分析

於第一階段系統驗證中，於 200 筆訓練文章基礎之下，系統針對 20 筆實際測試文章進行判定，獲得情緒類別推論平均召回率、正確率與 F 值為 45%，當中，實際文章情緒類別數為 20 個，系統推論之情緒類別數為 20 個，推論之正確數為 9 個，此階段詳細推論結果與三項指標之分佈趨勢如表 5.7、圖 5.34 所示。

整體而言，於此階段中，召回率、正確率與 F 值平均分佈趨勢呈現兩極化之狀態，但目前階段尚有過半之驗證資料無法判定正確，是故，根據本階段驗證結果中所獲致之三項指標而言，文章情緒類別推論之準確率與績效欠佳。

表 5.7、系統自我績效評估之第一階段驗證結果（共 200 筆訓練文章）

文章編號	實際情緒類別	推論情緒類別	實際情緒類別數	推論情緒類別數	正確情緒類別數	系統評估指標			
						召回率	正確率	F-值	
1	K7	K7	1	1	1	100%	100%	100%	
2	K6	K5	1	1	0	0%	0%	0	
3	K1	K1	1	1	1	100%	100%	100%	
4	K2	K1	1	1	0	0%	0%	0%	
5	K4	K4	1	1	1	100%	100%	100%	
6	K4	K6	1	1	0	0%	0%	0	
7	K5	K2	1	1	0	0%	0%	0	
8	K5	K5	1	1	1	100%	100%	100%	
9	K5	K2	1	1	0	0%	0%	0	
10	K3	K3	1	1	1	100%	100%	100%	
11	K6	K5	1	1	0	0%	0%	0	
12	K7	K7	1	1	1	100%	100%	100%	
13	K7	K4	1	1	0	0%	0%	0	
14	K1	K2	1	1	0	0%	0%	0	
15	K3	K3	1	1	1	100%	100%	100%	
16	K1	K4	1	1	0	0%	0%	0	
17	K3	K3	1	1	1	100%	100%	100%	
18	K6	K7	1	1	0	0%	0%	0	
19	K4	K1	1	1	0	0%	0%	0	
20	K1	K1	1	1	1	100%	100%	100%	
平均值							45%	45%	45%

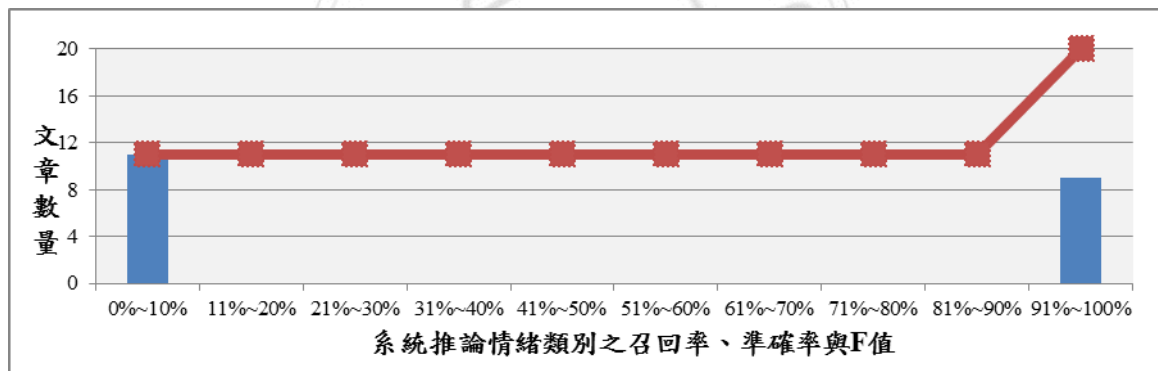


圖 5.34、第一階段情緒類別推論三項指標推論之分佈趨勢

(B-II) 第二階段驗證結果分析

本研究將第二階段分為 8 個週期，並於各週期匯入 100 份訓練文章，以觀察隨訓練文章之增加，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 5.8 所示，而各驗證週期文章情緒類別三項指標推論之績效分佈趨勢如圖 5.35 所示。

表 5.8、系統自我績效評估之推論結果彙整

文章表達情緒判定-系統自我績效評估		各週期情緒類別推論之驗證-訓練文章匯入數量									平均	
		第一階段	第二階段(各週期)									
		第一週期 200份	第二週期 300份	第三週期 400份	第四週期 500份	第五週期 600份	第六週期 700份	第七週期 800份	第八週期 900份	第九週期 1000份		
召回率	平均值	45%	50%	55%	60%	65%	70%	75%	80%	80%	64%	
	成長率	-	5%	5%	5%	5%	5%	5%	5%	0%	4.4%	
正確率	平均值	45%	50%	55%	60%	65%	70%	75%	80%	80%	64%	
	成長率	-	5%	5%	5%	5%	5%	5%	5%	0%	4.4%	
F 值	平均值	45%	50%	55%	60%	65%	70%	75%	80%	80%	64%	
	成長率	-	5%	5%	5%	5%	5%	5%	5%	0%	4.4%	

藉由對表 5.8 之觀察可得知，以每週期增加 100 份訓練文章為單位，平均每週期三項驗證指標之整體成長率約為 4.4%，而針對最後第九週期（共匯入 1000 份訓練文章）之驗證結果言之，三項指標分別由第一週期之 45% 提升至 80%，因此，綜合上述之觀察，本研究所開發之文章表達情緒判定乃具備學習能力與相當程度之正確性。

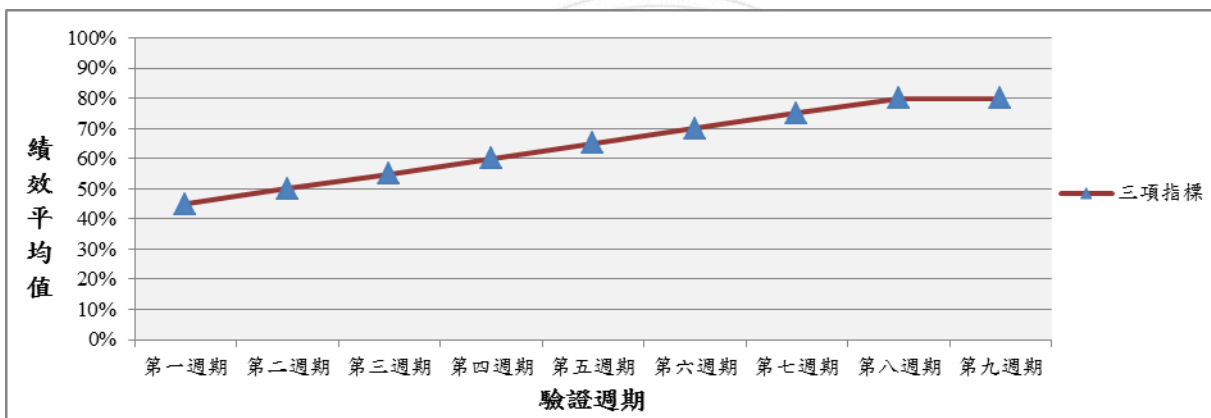


圖 5.35、系統自我績效評估-各驗證週期推論績效之分佈趨勢

文章表達情緒判定之整體驗證結果分析

於文章表達情緒判定之驗證中，首先與 Tung 與 Lu (2012) 之比較中，本研究於相同驗證情境與驗證資料數量下（1300 筆），本研究各項指標皆略優於比較研究；而於系統自我評估驗證中，於「第一階段驗證」時，當訓練文件之資料量為 200 份時，其正確率、召回率與 F 值整體平均值為 45%，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，於第八週期時，系統推論之三項指標可提升至 80% 之水準，此外，由驗證結果中亦可得知，平均每週期系統推論之召回率、正確率與 F 值，其整體成長率皆為 4.4%。承接於上述，本研究於不同訓練文章量下執行「文章表達情緒判定」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文章之增加，持續且穩定地成長與提升，而本研究所建置之「文章表達情緒判定模組」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 5.9 所示。由表 5.9 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八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 900 筆訓練知識文章時，可使系統各項推論指標績效提升至 80% 之水準。

表 5.9、三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	收斂週期累計訓練文章數	收斂前每週期平均成長率	整體每週期平均成長率
文章表達情緒判定召回率、正確率與 F 值	64%	第八週期	900 份	5%	4.4%

5.2.2 發文者文章語句重組之驗證

本研究所提之「發文者文章語句重組模組」乃針對違規文章之語句進行重組，以避免違反發文規範。故於驗證前，本研究乃先行建置驗證樣本，並將驗證樣本作為系統驗證之對照樣本，並說明系統驗證方式與驗證指標定義，以下乃針對「驗證方式說明」、「驗證樣本之實施與建置」、「系統驗證指標定義」與「系統驗證結果」。

發文者文章語句重組驗證方式說明

本研究乃於「Mobile01」論壇中隨機蒐集 600 筆訓練文章，而於驗證資料部分乃於論壇中隨機收集 20 份違規文章（如表 5.12 與表 5.13 所示），將違規文章依標點符號分割語句並隨機排列，以形成 20 份測試文章，接著，本研究乃將系統驗證規劃為兩階段，於系統驗證第一階段乃於 600 筆訓練文章中隨機挑選 200 筆匯入至系統中，並針對 20 份測試資料進行推論，以觀察系統初期之判定績效，進而觀察本研究所提方法論之正確性。待完成上述之第一階段系統績效驗證後，即進行系統測試第二階段之驗證，於此階段乃分為 4 個週期，每週期持續匯入 100 筆訓練文章資料（共計 600 筆），藉由持續匯入以分析系統於不同訓練測試資料下之推論結果，於各個週期乃利用前述 20 份測試資料重新進行推論，以分析系統之長期學習趨勢。

於驗證前本研究乃需建置測試資料之「實際需被改寫語句數」以作為系統推論之對照樣本，因此，首先乃邀請 30 位長期使用資訊論壇且擁有資訊背景、熟悉論壇發文規範之學生（皆屬於資工系與資管系）作為初期受測者，針對 20 份測試文章挑選將可能被歸類為違規文章之語句，以形成測試文章之實際違規語句（即需改寫之偏激語句），然而，受測者可能受不同之時間、環

境、身心狀況等，而影響主觀判斷違規語句之差異性，因此，為了驗證受測者對於閱讀違規文章語句之主觀差異性與一致性，本研究乃參考黃信華（2013）所設計之網路問卷並加以改良，以 20 份違規文章內所有之語句為依據，形成各測試文章之違規語句判斷問卷（如表 5.10 與表 5.11 測試文章編號 1 所示），同時隨機組合 1 份重複性測試文章（語句數為 20 份測試文章語句數之平均值），將重複性測試文章隨機安排至 20 份測試文章後，再請受測者針對 20 份測試文章進行「實際違規語句」之評估，並參考李旻芳（2010）以均方根（Root Mean Square；RMS）測試受測者主觀差異性，以「受測者重複性」、「受測者正確性」與「整體主觀差異值」三項指標，評估 30 位初期受測者閱讀語句之「主觀一致性與差異性」，最後，本研究乃以「整體主觀差異值」排名最好之前 20 位作為最終受測者，並以此些受測者所挑選之違規語句，作為 20 份驗證文章之實際違規語句，三項受測者主觀評估指標定義與說明如下：

受測者重複性指標為「受測者第一次挑選之違規語句數」與「受測者第二次挑選之違規語句，與第一次挑選之違規語句相符個數」之差異平均，相關符號定義如下：

OTS 受測者受測次數

ORF_i 第 i 位受測者於所有測試文章中，第一次挑選之違規語句個數

ORS_i 第 i 位受測者於所有測試文章，第二次挑選之違規語句與第一次之相符個數

ORE_i 第 i 位受測者之受測重複性

$$ORE_i = \sqrt{\frac{(ORF_i - ORS_i)^2}{OTS}} \quad (5.4)$$

受測者正確性指標為「受測者第二次挑選之違規語句個數」與「受測者第一次挑選之違規語句相符合個數」之差異平均，相關符號定義如下：

OAF_i 第 i 位受測者於所有測試文章中，受測者第二次挑選之違規語句個數

OAS_i 第 i 位受測者於所有測試文章，第二次挑選之違規語句與第一次之相符個數

OAC_i 第 i 位受測者之正確性

$$ORe_i = \sqrt{\frac{(ORF_i - ORS_i)^2}{OTS}} \quad (5.5)$$

受測者整體主觀差異指標為「受測者重複性乘以受測者正確性再乘以二」與「受測者重複值加受測者正確性」之比例，期望透過此項指標評估受測者主觀程度，當中，主觀程度愈低（即愈趨近於 0），即標示受測者具備良好的主觀感受，相關符號定義如下：

OSU_i 第 i 位受測者之整體主觀差異值

$$OSU_i = \frac{2 \times ORE_i \times ORA_i}{ORE_i + ORA_i} \quad (5.6)$$

表 5.10、測試文章編號 1 之偏激語句判斷問卷

語句編號	語句內容	請給予一個情緒
1	這樣對岸打過來要怎麼辦	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
2	不過軍方在接受媒體詢問時卻表示	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
3	真是沒有包庇嗎?	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
4	我男朋友現在跟一堆軍官在 KTV 玩脫衣陪酒	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
5	你們快去抓!	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
6	其中還包括一名高階將領	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
7	不過根據了解	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
8	屏東驚傳有軍官集體在三軍演訓期間跑去喝花酒被女友抓包告發	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
9	根據媒體消息指出	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
10	涉嫌在 13 日深夜前往有脫衣陪酒的 KTV 飲酒尋歡	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
11	軍方表示, 初步查證並無高階將領涉案	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
12	一群正在參與恆春半島例行性三軍演訓的軍官	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
13	其中還包括一名高階將領	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
14	直到其中一名上尉軍官的女友來電「查勤」得知男友竟在演訓期間喝花酒	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
15	而立即打電話向各單位報案	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒

表 5.11、測試文章編號 1 之偏激語句判斷問卷(續)

語句編號	語句內容	請給予一個情緒
16	才讓這起荒唐的花酒事件曝光	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
17	目前初步查證	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
18	一定會依法嚴辦	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒
19	表示「我男朋友現在跟一堆軍官在 KTV 玩脫衣陪酒, 你們快去抓!	<input type="checkbox"/> 偏激情緒 <input type="checkbox"/> 非偏激情緒

驗證樣本之實施與建置

本研究乃根據黃信華(2013)所設計之問卷加以改良，並邀請30位具資訊背景之學生，以得知20份測試文章之「實際需被改寫語句數」，其調查結果如表5.12與表5.13所示，最後，本研究乃以主觀一致性最佳之前20位受測者作為最終受測者(主觀差異指標愈趨近於0愈佳)，如表5.14、表5.15與表5.16所示，並以此20位受測者挑選20份測試文章「實際需被改寫語句數」之平均值作為最終之驗證樣本資料，如表5.13所示，以能與本系統推論之結果相互對照，進而得知系統推論之績效。

表 5.12、20 份測試文章之實際內容與偏激語句數

文章編號	實際文章標題	實際文章內容	實際語句數	受測者調查需改寫語句數
1	三軍演訓 將軍被爆到 KTV 玩脫衣陪酒	這樣對岸打過來要怎麼辦.真是沒有包庇嗎?「我男朋友現在跟一堆軍官在 KTV 玩脫衣陪酒,你們快..	23	5
2	王丹:男人不好色就是有人格缺陷	太陽花學運真的會被這些人搞爛掉 很多大學生美好的回憶 瞬間會變成恐怖記憶..	16	3
3	可笑又腦殘的規定	如果是這樣標準一致,我就沒有意見。但是,身分驗證應該要有彈性,一間醫院,初診時要出示身分證和健保卡,之後每次看診只..	30	6
4	台積電中科擴廠案環評初審過關	台中有台中港,大肚油庫,太平營區,新社花海,太平大里違章工業聚落,這麼多地方可以給台積電,都在台中市,沒有人要...	41	8
5	台灣女權快淪落到跟印度一樣	每次看到新聞上報導印度又有小女生被人輪姦後殺害, 內心都在想,這是什麼爛國家,幸好我們台灣還沒淪落到這樣..	16	2
6	性騷擾與嫖妓 哪一個比較不可原諒?	鬼打牆~你怎麼做我不知道~但是爺從來不摸也不對良家婦女開黃腔 嫖妓是自己私德 就不用拿召妓來護航了..	17	2
8	原受理性的少數也不過如此	每天拿太陽花女生,陳為亭的八卦來嬉笑怒罵,看來跟那些易被煽動的綠民有何分別 自居理性,群體表現卻只關心不同意見者..	7	1
9	國一南下約 41 公里處車禍	如果油門卡住往前暴衝,理當往路肩靠,想辦法用檔煞、護欄...嘗試減慢速度 而不是打故障燈,一昧地往內線衝吧~~想...	9	1
10	慈雲路興築興世代開賣!(有附圖)	你就只會扣帽子反正這版上 還是有不少人知道我是自住客,非什麼房仲跟投資客懷疑 有自住客 就不能在..	29	3
11	辦公室文員的提神妙方???	是哪家公司?建議可以在上班時間,開放貴公司股東,或有購買貴公司股票投資人,以及有購買貴公司產品的民眾參觀..	19	4
12	奴隸世代的年輕人如何自處?	革命! 推翻現今既得利益者 讓自己成為新的既得利益者 就連一個色狼慣犯癡漢陳韋廷 都懂得把握大腸花學..	7	1

表 5.13、20 份測試文章之實際內容與偏激語句數(續)

文章編號	實際文章標題	實際文章內容	實際語句數	受測者調查需改寫語句數
13	如何分辨非台北人	1. 視交通號誌為無物 2. 對於幾乎每條路都禁止左轉感到非常訝異 3. 電扶梯左線當路隊長放空..	10	2
14	柯 P:請國稅局夠大的人來解釋	父母遭查稅 柯 P:請國稅局夠大的人出來解釋 2015 年 01 月 12 日 15:50 前台南市議員王定宇在臉書指台北市長柯文哲父母..	47	9
15	Makiyo 母親癌逝台大醫院 享壽 64 歲	我深信一句話 狗是改不了吃屎的 MA 要怎麼喝都無所謂 喝死也好..洪幹也好 別再打司機或是酒駕撞死無辜的人就好..	6	1
16	M8 用不到 16 個小時就爆了	這篇居然可以活那麼久噢!?!三星版上的三星客服說只要在 play 商店或者是 samsung Apps 都可能導致刷機 而且客服..	15	2
17	千萬不要買 HTC RE 慘品,尤其在電腦展 HTC!	不要沒有受害就說受害者是工讀生 想護航請先自己買一個 RE 來看看 記得喔 要買白色的 而且 請觀察一個月..	10	1
18	好人好事(這沒置頂我還算是人嗎?)	幹麻自己對號入座 您不是神旁邊的小侍者而已嗎跟屁蟲不是在說你 不要自以..	4	1
19	中國萬噸 055 驅逐艦 可納百枚飛彈 -- 不曉得單船戰力是否超越美國柏克艦?	何苦跟著中國網民搖旗吶喊?他們吹噓 055 能多強,充其量頂多打憑柏克二,美國有破百艘,中國能維持幾艘..	5	1
20	香港房價真不是普通的可怕	我就是香港人 在此先抱歉,香港人都給大陸的人氣壞了,才好像冷漠無禮 20 到 30 歲的我們其實都在想可以去那裡..	19	3
平均語句數			17.4	2.95

表 5.14、30 位受測者對於 20 份測試文章之偏激語句數調查結果

受測者 文章編號	受測 次數	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
		1	第一次	3	3	3	6	7	6	4	2	8	3	7	7	4	7	7	2	5	6	7	3	3	4	8	8	3	4	8	7
	第二次	6	4	5	8	3	5	5	6	4	7	8	8	6	5	2	3	5	3	2	4	5	5	6	6	6	5	2	3	8	3
2	第一次	3	5	2	2	1	2	0	1	2	3	2	5	0	1	5	3	5	5	2	2	3	4	5	2	0	4	2	4	4	3
	第二次	1	3	2	3	1	1	2	1	0	3	5	3	0	3	3	3	3	5	3	1	2	4	3	3	3	4	3	2	2	2
3	第一次	5	4	3	8	3	8	3	2	6	8	3	7	6	1	7	7	7	6	6	9	5	7	6	8	6	6	8	6	0	5
	第二次	3	3	5	5	4	7	7	5	4	4	8	9	7	2	5	7	6	6	8	4	4	5	3	5	6	6	9	9	1	5
4	第一次	5	1	4	9	8	3	4	14	9	13	6	11	5	12	8	13	8	8	8	7	4	14	5	6	8	6	11	9	5	10
	第二次	3	2	5	8	12	3	7	9	6	8	8	12	8	11	14	10	3	6	4	10	2	10	10	4	14	14	2	9	12	6
5	第一次	1	3	1	0	0	1	1	1	4	2	0	2	2	1	3	3	1	3	3	1	0	0	0	2	0	1	2	2	0	2
	第二次	1	2	4	1	2	3	1	3	2	1	2	1	0	2	0	4	1	2	3	1	2	3	2	2	4	2	4	2	0	1
6	第一次	3	1	1	3	0	3	5	1	2	2	1	3	1	2	2	0	4	2	1	1	1	5	5	1	0	5	5	4	0	0
	第二次	3	3	0	4	1	4	3	4	2	0	4	5	4	3	2	0	4	4	4	1	3	2	2	1	0	1	5	1	1	1
7	第一次	5	1	3	1	5	1	4	2	2	4	5	2	4	0	3	4	5	1	0	5	3	5	0	0	2	3	2	3	5	3
	第二次	3	1	4	2	0	1	2	2	3	2	3	3	3	5	2	2	5	2	3	4	0	5	1	5	4	2	1	0	2	3
8	第一次	1	0	1	1	1	1	0	2	1	0	2	2	1	2	1	0	0	2	2	0	2	0	1	1	0	0	2	1	1	2
	第二次	1	2	1	1	2	0	0	0	2	2	1	1	0	1	0	1	0	0	0	0	0	0	0	0	2	0	2	2	0	1
9	第一次	2	0	1	0	3	0	1	2	0	3	2	2	1	1	0	1	1	2	3	0	1	3	3	1	2	1	3	2	3	3
	第二次	2	1	0	1	3	2	1	3	2	0	1	1	3	0	2	0	2	1	0	2	0	1	3	2	2	3	2	1	1	1
10(重複性文章共 17 個語句)	第一次	3	3	2	0	3	3	6	1	3	4	0	1	4	4	4	3	4	1	1	0	0	0	4	3	6	3	3	1	0	4
	第二次	1	4	1	2	1	4	5	5	4	6	2	5	3	5	0	5	2	2	4	5	1	0	2	4	4	3	6	2	6	3
11	第一次	4	3	0	1	5	5	5	6	2	3	7	3	4	1	3	7	3	3	5	1	3	7	0	2	3	1	2	5	0	2
	第二次	3	5	2	6	7	5	0	3	4	3	1	5	6	6	0	5	4	3	3	0	0	7	5	4	7	6	6	0	1	2
12	第一次	2	1	1	2	2	2	0	2	0	1	0	0	0	0	0	2	0	2	2	0	1	0	1	0	0	0	1	0	0	0
	第二次	2	0	2	2	2	1	0	0	2	2	0	0	2	2	0	0	1	1	1	1	0	2	0	1	0	1	1	2	0	0

表 5.15、30 位受測者對於 20 份測試文章之偏激語句數調查結果(續)

受測者 文章編號	受測 次數	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
		13	第一次	3	0	3	2	0	0	3	0	1	1	0	2	1	0	3	3	1	3	1	3	3	0	1	3	2	1	1	3
	第二次	3	3	0	1	2	3	2	0	2	3	3	2	3	1	1	1	1	2	0	0	3	3	0	3	0	0	0	1	0	2
14	第一次	13	5	8	9	9	9	6	12	15	6	7	10	11	6	13	12	14	12	8	9	11	13	6	6	13	10	8	8	7	5
	第二次	8	10	9	10	8	13	5	9	9	9	7	14	12	12	11	9	11	11	7	5	7	9	10	6	15	12	12	15	14	5
15	第一次	2	0	0	0	0	2	1	1	2	1	1	2	1	2	0	1	0	1	2	0	1	0	2	0	2	2	1	2	1	0
	第二次	2	1	2	2	2	0	1	2	1	2	1	0	2	2	0	0	0	0	0	0	0	1	0	2	1	2	0	0	2	2
16	第一次	2	3	0	1	4	2	4	1	4	1	2	3	3	3	4	2	3	4	3	2	1	4	4	0	2	0	4	0	1	0
	第二次	1	4	4	1	1	1	4	1	4	1	2	2	2	0	3	2	3	3	2	0	4	1	3	2	2	0	1	3	0	4
17	第一次	0	0	1	1	1	0	1	1	0	2	0	0	2	0	1	2	1	2	1	2	2	0	0	0	1	0	1	0	2	1
	第二次	0	2	1	1	0	1	0	2	0	0	0	1	0	1	2	1	2	0	2	2	1	1	0	0	2	2	2	2	0	2
18	第一次	0	1	0	1	0	1	1	0	0	0	1	0	1	1	1	0	1	0	0	1	0	1	0	1	0	0	0	0	1	1
	第二次	1	0	1	1	1	1	1	0	1	0	1	0	1	1	0	1	1	1	0	0	0	0	0	1	0	1	1	0	1	0
19	第一次	0	0	0	1	1	1	2	1	2	1	0	0	1	0	2	1	1	0	2	1	2	2	2	0	1	0	2	1	2	1
	第二次	2	1	1	2	2	1	2	2	1	1	1	2	2	0	2	0	2	2	2	1	2	1	1	0	2	1	0	2	2	0
20	第一次	7	2	7	6	7	0	1	5	0	3	7	2	2	7	7	2	7	1	3	3	4	0	0	0	2	7	3	5	1	6
	第二次	1	7	5	0	4	6	6	4	7	5	4	0	0	0	2	5	2	2	6	5	1	3	5	7	3	1	5	3	0	7

表 5.16、30 位受測者主觀差異性調查結果

受測者	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
第一次挑選語句總數	64	36	41	54	60	50	52	57	63	61	53	64	54	51	74	68	71	64	60	50	50	69	53	44	53	54	69	63	38	53
第二次挑選語句總數	47	58	54	61	58	62	54	61	60	59	62	74	64	62	51	59	58	56	54	46	37	63	56	58	77	66	64	59	53	50
第二次挑選語句與第一次之相符總個數	30	30	37	45	50	39	47	39	56	48	37	51	35	42	46	55	48	51	42	45	34	56	49	38	42	52	66	55	35	49
受測者重複性	24.0	4.2	2.8	6.4	7.1	7.8	3.5	12.7	4.9	9.2	11.3	9.2	13.4	6.4	19.8	9.2	16.3	9.2	12.7	3.5	11.3	9.2	2.8	4.2	7.8	1.4	2.1	5.7	2.1	2.8
受測者正確性	12.0	19.8	12.0	11.3	5.7	16.3	4.9	15.6	2.8	7.8	17.7	16.3	20.5	14.1	3.5	2.8	7.1	3.5	8.5	0.7	2.1	4.9	4.9	14.1	24.7	9.9	1.4	2.8	12.7	0.7
整體主觀差異指標(愈趨近於 0 愈好)	16.03	6.99	4.58	8.15	6.29	10.52	4.12	14.00	3.60	8.43	13.80	11.75	16.23	8.78	6.00	4.33	9.86	5.11	10.18	1.18	3.57	6.43	3.60	6.53	11.84	2.47	1.70	3.77	3.64	1.13
排名	2	13	19	12	16	7	21	3	24	11	4	6	1	10	17	20	9	18	8	29	26	15	24	14	5	27	28	22	23	30
不採納之使用者	不採納					不採納		不採納			不採納	不採納	不採納	不採納			不採納		不採納							不採納				

發文者文章語句重組驗證之評估指標定義

藉由系統推論文章之改寫語句數與實際需被改寫語句數驗證系統之準確性，於此階段績效指標乃包含「發文者文章語句重組推論召回率」、「發文者文章語句重組推論正確率」與「整體推論正確率與召回率之協調值」，其參數定義與計算方式如公式(5.7)至公式(5.9)所示

發文者文章語句重組推論之召回率

發文者文章語句重組推論之召回率為一相對比例值，為「實際文章需被改寫語句數與推論文章需被改寫語句數之相符個數」與「實際文章需被改寫語句數」之比例，其相關變數定義如下：

$RSen_j$ 系統推論文章語句重組語句時，第 j 份文章之需改寫語句推論召回率

Con_j 第 j 份文章實際需被改寫語句數與推論需被改寫語句數相符之個數

$ASen_j$ 第 j 份文章實際需被改寫語句數

$$RSen_j = \frac{Con_j}{ASen_j} \quad (5.7)$$

發文者文章語句重組推論之正確率

發文者文章語句重組推論之正確率為一相對比例值，為「實際文章需被改寫語句數與推論文章需被改寫語句數之相符個數」與「推論文章需被改寫語句數」之比例，期望藉由此項指標評估系統推論需被改寫語句數與實際需被改寫語句數之差異程度。相關變數定義如下：

$PSen_j$ 系統推論文章需被改寫語句數時，第 j 份文章之需改寫語句推論正確率

Con_j 第 j 份文章實際需被改寫語句數與推論需被改寫語句數相符之個數

$DSen_j$ 系統推論第 j 份文章之需被改寫語句數

$$PSen_j = \frac{Con_j}{DSen_j} \quad (5.8)$$

發文者文章語句重組推論正確率與召回率之協調值 (F-Measure)

發文者文章語句重組推論協調值為評估系統推論召回率與正確率之成效指標，為

「系統正確率乘以召回率再乘以二」與「系統正確率加召回率」之比例，期望透過此項指標評估系統整體績效，其相關變數定義如下：

$FSen_i$ 系統推論文章需被改寫語句數時，第 j 份文章之改寫語句數協調值

$$FSen_j = \frac{2 \times RSen_j \times PSen_j}{RSen_j + PSen_j} \quad (5.9)$$

驗證結果分析

於發文者文章語句重組模組之驗證結果中，本研究乃分為「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。於第一階段中乃先行匯入 200 筆訓練文章後，推論 20 筆驗證資料之結果，藉此觀察系統初期之推論績效；接著於第二階段中，本研究乃劃分為 4 個週期，於每週期匯入 100 筆訓練文章並重新推論 20 筆驗證資料之結果，以測試不同訓練量下系統進行「發文者文章語句重組」之績效。以下即針對各階段說明系統驗證過程與結果。

第一階段驗證結果分析

於第一階段系統驗證中，於 200 筆訓練文章基礎之下，系統針對 20 筆實際測試文章進行推論，獲得需被改寫語句推論平均召回率為 46.61%、正確率為 51.08% 與 F 值為 47.22%，當中，此階段詳細推論結果與三項指標之分佈趨勢如圖 5.36、圖 5.37、圖 5.38 與表 5.17 所示。整體而言，於此階段中，召回率、正確率與 F 值平均分佈趨勢大多落於 50% 以下之狀態，但目前階段尚有過半之驗證資料無法判定正確，是故，根據本階段驗證結果中所獲致之三項指標而言，發文者文章語句重組推論之準確率與績效欠佳。

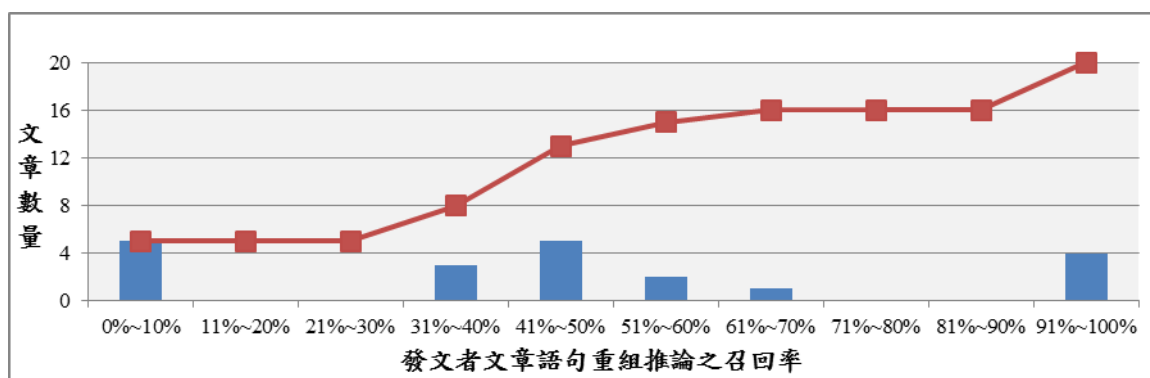


圖 5.36、發文者文章語句重組推論召回率

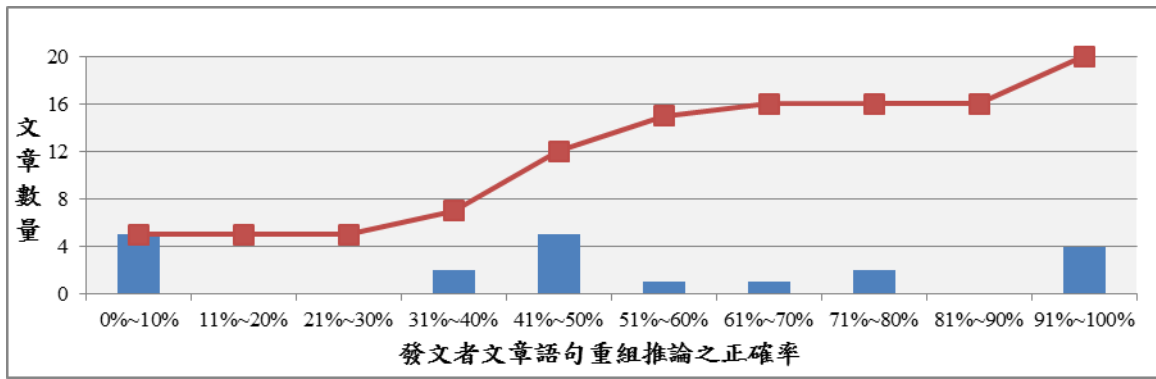


圖 5.37、發文者文章語句重組推論正確率

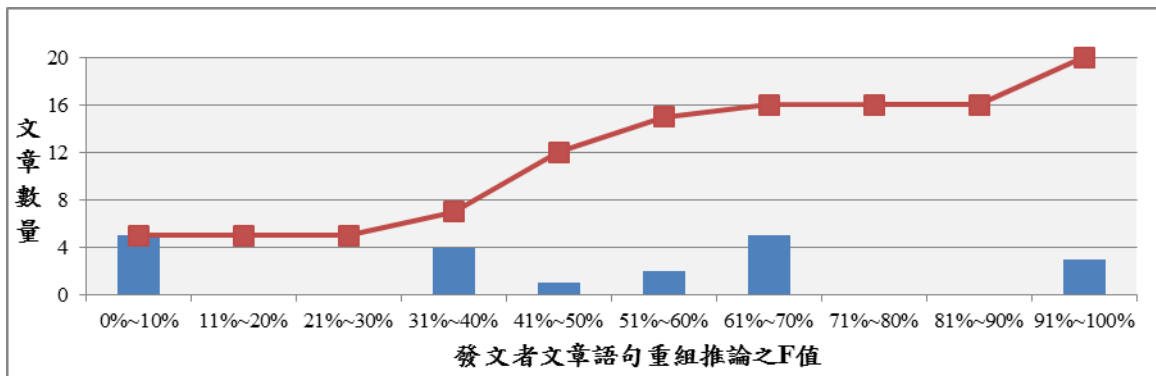


圖 5.38、發文者文章語句重組推論 F 值

表 5.17、系統自我績效評估之第一階段驗證結果（共 200 筆訓練文章）

文章編號	文章實際語句數	實際需改寫語句數	系統推論需改寫語句數	系統推論需改寫語句正確數	召回率	正確率	F-值
1	23	5	4	3	60.00%	75.00%	66.67%
2	16	3	4	2	66.67%	50.00%	57.14%
3	30	6	5	3	50.00%	60.00%	54.55%
4	41	8	5	4	50.00%	80.00%	61.54%
5	16	2	1	0	0.00%	0.00%	0.00%
6	17	2	1	0	0.00%	0.00%	0.00%
7	18	3	2	1	33.33%	50.00%	40.00%
8	7	1	1	0	0.00%	0.00%	0.00%
9	9	1	0	0	0.00%	0.00%	0.00%
10	29	3	2	1	33.33%	50.00%	40.00%
11	19	4	5	2	50.00%	40.00%	44.44%
12	7	1	1	1	100.00%	100.00%	100.00%
13	10	2	1	1	50.00%	100.00%	66.67%
14	47	9	6	5	55.56%	83.33%	66.67%
15	6	1	1	1	100.00%	100.00%	100.00%
16	15	2	3	1	50.00%	33.33%	40.00%
17	10	1	2	1	100.00%	50.00%	66.67%
18	4	1	0	0	0.00%	0.00%	0.00%
19	5	1	1	1	100.00%	100.00%	100.00%
20	19	3	2	1	33.33%	50.00%	40.00%
總平均					46.61%	51.08%	47.22%

第二階段驗證結果分析

本研究將第二階段分為 4 個週期，並於各週期匯入 100 份訓練文章，以觀察隨訓練文章之增加，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 5.18 所示，而各驗證週期文章情緒類別三項指標推論之績效分佈趨勢如圖 5.39 所示。

表 5.18、發文者文章語句重組之推論結果彙整

發文者文章語句重組推論績效		各週期文章語句重組推論之驗證-訓練文章匯入數量					平均
		第一階段		第二階段			
		第一週期 200 份	第二週期 300 份	第三週期 400 份	第四週期 500 份	第五週期 600 份	
召回率	平均值	46.61%	58.28%	66.22%	77.89%	78.51%	65.50%
	成長率	-	11.67%	7.94%	11.67%	0.62%	7.98%
正確率	平均值	51.08%	60.63%	68.88%	78.49%	78.58%	67.53%
	成長率	-	9.55%	8.25%	9.61%	0.09%	6.88%
F 值	平均值	47.22%	57.88%	65.97%	77.15%	77.53%	65.15%
	成長率	-	10.66%	8.09%	11.18%	0.38%	7.58%

藉由對表 5.19 之觀察可得知，以每週期增加 100 份訓練文章為單位，平均每週期召回率之整體成長率約為 7.98%、正確率為 6.88%、F 值為 7.58%，而針對最後第五週期（共匯入 600 份訓練文章）之驗證結果言之，召回率由第一週期 46.61% 提升至 78.51%，正確率由 51.08% 提升至 78.58%，F 值由 47.22% 提升至 77.53%，因此，綜合上述之觀察，本研究所開發之發文者文章語句重組推論乃具備學習能力與相當程度之正確性。

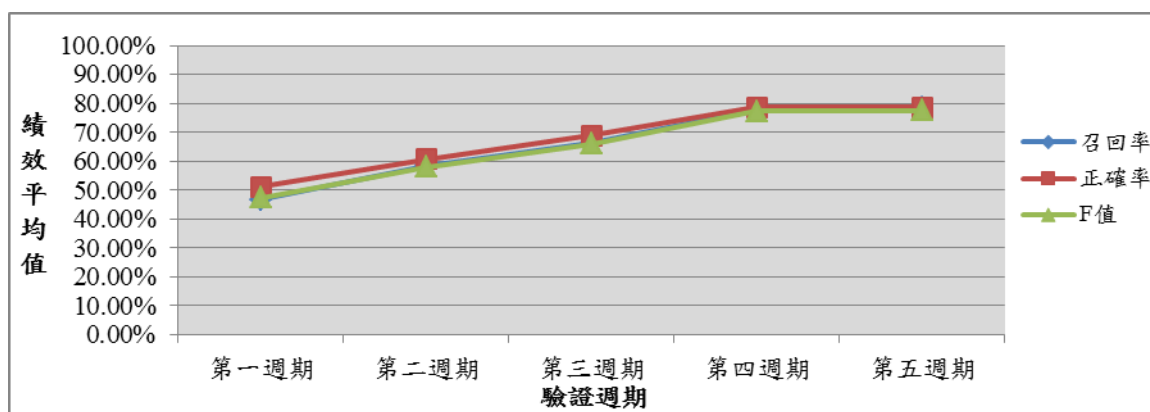


圖 5.39、發文者文章語句重組推論-各驗證週期推論績效之分佈趨勢

發文者文章語句重組推論之整體驗證結果分析

整體驗證結果而言，於「第一階段驗證」時，當訓練文件之資料量為 200 份時，其召回率、正確率與 F 值整體平均值分別為 46.62%、51.08% 與 47.22%，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，於第五週期時，系統推論之三項指標可分別提升至 78.51%、78.58% 與 77.53% 之水準，此外，由驗證結果中亦可得知，平均每週期系統推論之召回率、正確率與 F 值，其整體成長率分別為 7.98%、6.88% 與 7.58%。承接於上述，本研究於不同訓練文章量下執行「發文者文章語句重組推論」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文章之增加，持續且穩定地成長與提升，而本研究所建置之「發文者文章語句重組模組」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 5.19 所示。由表 5.20 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八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 900 筆訓練知識文章時，可使系統各項推論指標績效提升至 80% 之水準。

表 5.19、三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	收斂週期累計 訓練文章數	收斂前每週期 平均成長率	整體每週期 平均成長率
文章語句重組 推論召回率	65.50%	第四週期	500 筆	10.43%	7.98%
文章語句重組 推論正確率	67.53%	第四週期	500 筆	9.14%	6.88%
文章語句重組 推論 F 值	65.15%	第四週期	500 筆	9.98%	7.58%

5.3 小結

為了驗證本研究所提方法論與系統之可行性與正確率，本研究乃將「文章表達情緒判定模組」與「發文者文章語句重組模組」個別進行案例驗證。於「文章表達情緒判定模組」之驗證中，本研究乃分為「與其他研究之比較」與「系統自我績效評估」兩方式進行驗證，於「與其他研究之比較」中，本研究乃借鑑 Tung 與 Lu (2012) 之驗證方式並以「痞客邦」部落格文章作為驗證資料，以能與該研究進行績效之比較並藉由驗證結果得知，本研究推論之績效指標略優於 Tung 與 Lu (2012) 近 10% 之水準；而於「系統自我績效評估」中，本研究乃以「奇摩新聞」作為驗證資料之來源，同時透過驗證結果得知，系統乃隨著不同週期與訓練文章數之增加下，於最終週期（第九周期）時三項指標（正確率、召回率與 F 值）皆可達到 80% 之水準。

於「發文者文章語句重組模組」驗證中，本研究乃以「Mobile01」論壇作為驗證資料之來源，同時藉由受測者主觀性調查與最終受測者之評估，以形成驗證資料之實際需改寫語句並與系統推論之數據進行對照，最後，藉由驗證結果可得知於第五週期時，系統推論召回率、正確率與 F 值分別可達 78.51%、78.58% 與 77.53% 之水準。整體而言，透過上述驗證之結果可得知，本研究所提之方法論與開發之系統乃實具可行性與正確性。

第六章、結論與未來發展

近年因網路之蓬勃發展及論壇之崛起，使用者可透過發言規範與社群管理員之審核，即可輕易地發表各項文章並討論其內容。然而，文章撰寫者所發表之文章篇幅與詞語用法不盡相同，為維持論壇文章之品質，論壇乃以現有之文字比對技術及發文規範，審核文章之內容，藉此過濾違規文章，但因網路創造甚多奇異詞彙，當文章具有獨特用詞時，將降低社群之文字比對方法過濾違規文章之成效，因此，論壇管理者需逐筆審視所有之文章，方可從中過濾違規之文章或給予修正，但此舉將花費大量人力成本與時間；對於文章撰寫者而言，可能因個人疏忽之關係，而於無意識中將帶有偏激之詞語寫入文章內，因而違規而遭致論壇平台或論壇管理者移除，將降低再發文之意願。

為改善既有論壇審核違規文章機制所面臨之問題，並可立即修正違規文章之內容，藉由對近期相關文獻之借鑑與改良，本研究乃發展一套「以論壇文章蘊含情感為基之文章重組模式」，藉由論壇文章之基礎資訊，結合語句相似度分析技術、自動化情緒隸屬係數建立技術、語句評閱技術、多重組合句之建立等，推論文章隱含之情緒資訊，並根據偏激情緒之語句重組文章語句之內容以協助論壇使用者與管理者避免文章違反論壇規範之可能性。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「以論壇文章蘊含情感為基之文章重組模系統」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「以論壇文章蘊含情感為基之文章重組模式」方法論

本研究發展之模式主要針對「文章表達情緒判定」，以及「發文者文章語句重組」等兩面向之議題進行解析，於「文章表達情緒判定模組」中，本研究乃結合「自動化情緒詞彙隸屬係數推論」、「文章代表事件解析」、「代表語句相似度解析」等技術，以判定發論壇文章之情緒類別。另一方面，於「發文者文章語句重組」中，本研究乃透過「文章評閱技術」、「文章語句相似度」、「多重組合句建立」及「文章語句架構重組」等技術，

重組文章中帶有偏激字詞之語句。

2. 建置「以論壇文章蘊含情感為基之文章重組模式」

本研究乃以「以論壇文章蘊含情感為基之文章重組模式」方法論為依據，運用 JSP (Java Server Pages) 語法進行開發，並利用 SQL Server 2008 存取資料庫，以建置一套「以論壇文章蘊含情感為基之文章重組系統」，令使用者能進行相關欲分析資料與訓練文章之上傳與維護，並透過對系統兩核心模組之運用，藉由自動化方式分析得文章情緒類別與重組後之文章內容，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，本研究乃以「痞客邦」與「奇摩新聞」作為「文章表達情緒判定模組」驗證資料與訓練文章之來源；此外，本研究以「Mobile01 論壇」作為「發文者文章語句重組模組」之驗證資料與訓練文章資料來源，同時以質化之比較探討驗證資料來源之可行性與可用性，並能個別進行獨立之驗證，以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得知，「文章表達情緒判定模組」與其他研究比較後，其三項指標皆高於 10% 之水準，同時於系統自我績效評估部分，三項指標皆獲致 80% 之水平；於「發文者文章語句重組模組」中，於最終績效驗證指標中，可獲致整體平均召回率 78.51%、正確率 78.58%、F 值為 77.53%。

綜合於各小節之系統驗證結果，包含「文章表達情緒判定模組」、「發文者文章語句重組」，本研究所提出之「以論壇文章蘊含情感為基之文章重組」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

本研究乃以論壇文章為基礎，發展一套「以論壇文章蘊含情感為基之文章重組」模式，以論壇文章之「文章標題」、「文章語句」、「情緒詞彙」等資料作為分析之基礎，解析得「文章情緒類別」與「文章重組後之內容」，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、Miao 等人 (2012) 以及為鑒借，建立自動化情緒詞彙庫、文章代表事件之解析、語句相似度之解析、文章情緒類別之解析、文

章評閱分數解析，分析得知目標論壇文章之情緒類別與語句重組之內容，以協助論壇管理者快速取得違規之文章。

- 藉由Chen等人(2010)與向量空間模型之鑒借，經本研究之改良後，整合並發展出一套適用於發文者文章語句重組模組，藉以分析文章評閱分數，並自動多重組合句以重組文章語句之內容，以能針對違規但具有獨特看法之文章重組其內容，進而保留有意義之文章。

技術開發層面

本研究乃以「以論壇文章蘊含情感為基之文章重組」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「以論壇文章蘊含情感為基之文章重組」系統，其系統具體成效乃歸納如下：

- 本系統所開發之「以論壇文章蘊含情感為基之文章重組」乃建置於網際網路中，並可執行「文章表達情緒判定」與「發文者文章語句重組」，藉此取得目標文章代表事件、情緒類別以及文章語句重組之內容。
- 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，當使用一定訓練文章量時，系統即可有效推論文章情緒類別與重組文章之語句，且具備相當程度之可行性。

實務應用層面

本研究可有效針對論壇之文章，分析文章之情緒類別並重組語句內容，以避免發文者違反論壇規範，並可協助論壇管理者保留有意義且未違規之文章，因此，本研究於實務應用層面，可應用於線上論壇中，具體貢獻與成效乃歸納如下：

- 於論壇管理者層面，可針對特定情緒之文章可快速取得違規文章並將改寫違規之語句，以節省逐筆審核違規文章之時間。
- 於文章撰寫者層面，可自動將違規文章之語句進行修正，以節省修改違規文章之時間與人力，並可立即得知文章違規之可能性。
- 於整體論壇之層面，可協助論壇管理者快速過濾違規之文章，並避免文章撰寫者撰寫之文章觸犯社群規範，進而促使使用者持續使用之意願。

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「以論壇文章蘊含情緒為基之文章重組」方法論與系統理論層面、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. 發展一套整合文章圖片、符號與文字情緒為基之文章重組模式：本研究主要乃以文章文字內容為解析之對象，但尚未解析文字搭配圖片之蘊含情緒深入探討，故期望未來將圖片與文字之搭配進行解析，以能更精確得知文章所蘊含之情緒類別，並發展更具精確性與多方性之判定模式。
2. 發展一套多語言以論壇文章蘊含情緒為基之文章重組模式：現階段本研究主要解析中文論壇中之文字內容，然而中文文字之詞彙用法與各國語言不同，故期望發展跨語言文章情緒解析與語句重組技術，以協助系統進行跨國語言論壇資料之解析。

參考文獻

1. Afraz, Z. S., Muhammad, A. and Ana M. M. E., 2010, "Lexicon based sentiment analysis of Urdu text using SentiUnits," *Advances in Artificial Intelligence*, Vol. 6437, pp. 32-43.
2. Alavi, M. and Leidner, D. E., 2012, "Review knowledge management and knowledge management systems Conceptual foundations and research issues," *MIS Quarterly*, Vol. 25, No. 1, pp. 107-136.
3. Bai, X., 2011, "Predicting consumer sentiments from online text," *Decision Support Systems*, Vol. 50, No. 4, pp. 732-742.
4. Chan, S. W. K., 2006, "Beyond keyword and cue-phrase matching: A sentence-based abstraction technique for information extraction," *Decision Support Systems*, Vol. 42, No. 2, pp. 759-777.
5. Chen, Y. Y., Liu, C. L., Chang, T. H. and Lee, C. H., 2010, "An unsupervised automated essay scoring system," *Intelligent Systems*, Vol. 25, No. 5, pp. 61-67.
6. Danushka, B., Naoali, O. and Mitsuru, I., 2012, "A preference learning approach to sentence ordering for multi-document summarization," *Information Sciences*, Vol. 217, pp. 78-95.
7. Das, Dipankar. and Bandyopadhyay, S., 2010, "Sentence level emotion tagging on blog and news corpora," *Advances in Pattern Recognition*, Vol. 6256, pp. 332-341.
8. David, D. R., Fernando, M., Ramon, L. H., Stephan, M., Ricardo, G, Cerferino, M. and Francisco, D. P., 2011, "Conflict and cognitive control during sentence comprehension Recruitment of a frontal network during the processing of Spanish object-first sentences," *Neuropsychologia*, Vol. 49, No. 3, pp. 382-391.
9. Eikvil, L., 1999, "Information extraction from world wide web: A survey," *Norwegian Computing Center*, No. 945.
10. Ercan, G. and Cicekli, I., 2007, "Using lexical chains for keyword extraction," *Information Processing and Management*, Vol. 43, No. 6, pp. 1705-1714.
11. Fan, N., Cai, W. D., Zhao, Y. and Li, H. X., 2009, "Extraction of sentiment topic sentences of Chinese texts," *Journal of Computer Applications*, Vol. 29, No. 4, pp. 1171-1174.

12. 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.
13. Fu, X., Liu, G., Guo, Y. and Wang, Z., 2013, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowledge-Based Systems*, Vol. 37, pp 186-195.
14. Ge, S. L. and Chen, X. X., 2009, "Cluster analysis of college English writing in automated essay scoring," *Computer Engineering and Applications*, Vol. 45, No. 6, pp. 145-148.
15. Gregory, P. S. and Daniel, N. A., 2007, "Emotion intensity and categorization ratings for emotional and nonemotional words," *Cognition and Emotion*, Vol. 22, No. 1, pp. 114-133.
16. Gu, Y. and Grossman, R. L., 2010, "Sector: A high performance wide area community data storage and sharing system," *Future Generation Computer Systems*, Vol. 26, No. 5, pp. 720-728.
17. Guido, B. and Leendert, V. D. T., 2006, "Security policies for sharing knowledge in virtual communities," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 36, No. 3, pp. 439-450.
18. Hamouda, A., Marei, M. and Rohaim, M., 2011, "Building machine learning based senti-word lexicon for sentiment analysis," *Journal of Advances in Information Technology*, Vol. 2, No. 4, pp. 199-203.
19. Hao, X. Y., Li, J. H., You, L. P. and Liu, K. Y., 2007, "A research on building of Chinese reading comprehension corpus," *Journal of Chinese Information Processing*, Vol. 21, No. 6, pp. 29-35.
20. Hollink, L., Schreiber, A. T., Wielinga, B. J. and Worrying, M., 2004, "Classification of user image descriptions," *International Journal of Human-Computer Studies*, Vol. 61, No. 5, pp. 601-626.
21. Hsu, C. L. and Lin, C. C., 2007, "Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation," *Information & Management*, Vol. 45, No. 1, pp. 65-74.

22. Huang, J., Tian, S. W., Yu, L. and Feng, G. J., 2012, "Sentence sentiment analysis based on Uyghur sentiment word," *Computer Engineering*, Vol. 38, No. 9, pp. 182-185.
23. Ichifuji, Y., Konno S. and Sone, H., 2010, "An advisory method for BBS users and evaluation of BBS comments," *Procedia Social and Behavioral Sciences*, Vol. 2, No. 1, pp. 218-224.
24. Jiang, H., Huang, G. Q. and Liu, J. G., 2011, "The research on CET automated essay scoring based on data mining," *Advances in Computer Science and Education Applications*, Vol. 202, pp. 100-105.
25. Jonathan, K. and Yiu, K. N., 2006, "Using word clusters to detect similar web documents," *Knowledge Science, Engineering and Management*, Vol. 4092, pp. 215-228.
26. Jung, H., Yi, E., Kim, D. and Lee, G. G., 2005, "Information extraction with automatic knowledge expansion," *Information Processing and Management*, Vol. 41, No. 2, pp. 217-242.
27. Kaila, S., Huang, H., Wu, X. and Zhang, S., 2006, "A logical framework for identifying quality knowledge from different data sources," *Decision Support Systems*, Vol. 42, No. 3, pp. 1673-1683.
28. Kakkonen, T., Myller, N., Sutinen, E. and Timonen, J., 2008, "Comparison of dimension reduction methods for automated essay grading," *Educational Technology & Society*, Vol. 11, No. 3, pp. 275-288.
29. Ko, Y. and Seo, J., 2008, "An effective sentence-extraction technique using contextual information and statistical approaches for text summarization," *Pattern Recognition Letters*, Vol. 29, No. 9, pp. 1366-1371.
30. Kuo, Y. H. and Huang, H. H., 2007, "Automatic extraction of key sentences via word sense identification for Chinese text summarization," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 11, No. 4, pp. 416-422.
31. Lai, H. M. and Chen, T. T., 2014, "Knowledge sharing in interest online communities: A comparison of posters and lurkers," *Computers in Human Behavior*, Vol. 35, pp. 295-306.
32. Lai, L. F., 2007, "A knowledge engineering approach to knowledge management," *Information Sciences*, Vol. 177, No. 19, pp. 4072-4094.

33. Lee, K. J., Choi, Y. S. and Kim, J. E., 2011, "Building an automated English sentence evaluation system for students learning English as a second language," *Computer Speech and Language*, Vol. 25, No, 2 pp. 246-260.
34. Li, C. L., Zhu, Y. H. and Xu, Y. Q., 2011, "Research of attribute word extraction method in Chinese product comment," *Computer Engineering*, Vol. 37, No. 12, pp. 26-29.
35. Li, G. and Liu, F., 2012, "Application of a clustering method on sentiment analysis," *Journal of Information Science*, Vol. 38, No. 2, pp. 127-139.
36. Li, N. and Wu, D. D., 2010, "Using text mining and sentiment analysis for online forums hotspot detection and forecast," *Decision Support System*, Vol. 48, No. 2, pp. 354-368.
37. Li, S., Ye, Q., Li, Y. j. and Luo, S. Q., 2008, "Information inference based sentiment classification for online news comments," *Journal of Chinese Information Processing*, Vol. 23, No. 5, pp. 75-79.
38. Li, Y. H., David, M., Zuhair, A. B., James, D. O. and Keeley, C., 2006, "Sentence similarity based on semantic nets and corpus statistics," *Transactions on Knowledge and Data Engineering*, Vol. 18, No, 8 pp. 1138-1150.
39. Li, Y. M., Liao, T. F. and Lai, C. Y., 2012, "A social recommender mechanism for improving knowledge sharing in online forums," *Information Processing and Management*, Vol. 48, No. 5, pp. 978-994.
40. Liao, C. H., Kuo, B. C. and Pai, K. C., 2012, "Effectiveness of automated Chinese sentence scoring with latent semantic analysis," *The Turkish Online Journal of Educational Technology*, Vol. 11, No, 2 pp. 80-87.
41. Lin, F. R. and Liang, C. H., 2008, "Storyline-based summarization for news topic retrospection," *Decision Support Systems*, Vol. 45, No. 3, pp. 473-490.
42. Lin, Y., Ye, Q., Li, J., Zhang, Z. and Wang, T., 2011, "Snippet-based unsupervised approach for sentiment classification of Chinese online reviews," *International Journal of Information Technology & Decision Making*, Vol. 10, No. 6, pp. 1097-1110.
43. Liu, J., Wang, B., Lu, H. and Ma, S., 2008, "A graph-based image annotation framework," *Pattern Recognition Letters*, Vol. 29, No. 4, pp. 407-415.
44. Liu, W. P., Zhu, Y. H., Li, C. L., Xiang, H. Z. and Wen, Z. Q., 2009, "Research on building Chinese basic semantic lexicon," *Journal of Computer Application*, Vol. 29, No.

- 10, pp. 2875-2877.
45. Liu, W., Yan, H. L., Xiao, J. G. and Zeng, J. X., 2010, "Solution for automatic web review extraction," *Journal of Software*, Vol. 21, No. 12, pp. 3220-3236.
 46. Liu, Z, M. and Liu, L., 2012, "Empirical study of sentiment classification for Chinese microblog based on machine learning," *Computer Engineering and Applications*, Vol. 40, No. 1, pp. 1-4.
 47. Lu, J., Liang, Y. H., Lu, Y. Q., Li, B. and Yao, J. M., 2009, "The application of classifier combining in automated essay scoring," *Micro Electronics & Computer*, Vol. 26, No. 10, pp. 69-73.
 48. Matsubara, D., Miki, K., Inouchi, H. and Hoshi, T., 2006, "File management using virtual directory architecture for central managed P2P information sharing system (NRBS)," *Information and Media Technologies*, Vol. 1, No. 1 pp.514-523.
 49. Matsumoto, K., Konishi, Y., Sayama, H. and Ren, F., 2011, "Analysis of Wakamono Kotoba emotion corpus and its application in emotion estimation," *International Journal of Advanced Intelligence*, Vol. 3, No. 1, pp. 1-24.
 50. Miao, Y., Su, J., Liu, S. and Zhang, J., 2012, "Bootstrapping-based method for chinese sentiment lexicon construction," *International Conference on Information Engineering Lecture Notes in Information Technology*, Vol. 25, pp. 248-253.
 51. Oh, H., Fiorito, S. S., Cho, H. and Hofacker, C. F., 2008, "Effects of design factors on store image and expectation of merchandise quality in web-based stores," *Journal of Retailing and Consumer Services*, Vol. 15, No. 4, pp. 237-249.
 52. Okamoto, J. and Ishizaki, S., 2011, "Important sentence extraction using contextual semantic network," *Procedia-Social and Behavioral Sciences*, Vol. 27, pp 86-94.
 53. Pang, H. J., 2010, "Text sentiment analysis-oriented commodity review detection," *Journal of Computer Applications*, Vol. 32, No. 7, pp. 2038-2041.
 54. Peng, H., Shi, Z. Z., Qiu, L. L. and Chang, L., 2008, "Matching algorithm of semantic web service based on similarity of ontology concepts," *Computer Engineering*, Vol. 34, No. 15, pp. 51-53.
 55. Peng, J., Yang D. Q., Tang, S. W., Fu, J. and Jiang, H. K., 2007, "A novel text clustering algorithm based on inner product space model of semantic," *Chinese Journal of*

- Computers*, Vol. 30, No. 8, pp. 1354-1363.
56. Pérez, M. S., Sánchez, A., Robles, V. and Herrero, P., 2007, "Design and implementation of a data mining grid-aware architecture," *Future Generation Computer Systems*, Vol. 23, No. 1, pp. 42-47.
 57. Rorissa, A., 2008, "User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory," *Information Processing and Management*, Vol. 44, No. 5, pp. 1741-1753.
 58. Teo, H., Chan, H. C., Wei, K. K. and Zhang, Z., 2003, "Evaluating information accessibility and community adaptivity features for sustaining virtual learning communities," *International Journal of Human-Computer Studies*, Vol. 59, No. 5, pp. 671-697.
 59. Tokuhisa, R., Inui, K., Liu, S. and Matsumoto, Y., 2008, "Emotion classification using massive examples extracted from the web," *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 881-888.
 60. Tractinsky, N., Cokhavi, A., Kirschenbaum, M. and Sharfi, T., 2006, "Evaluating the consistency of immediate aesthetic perceptions of web pages," *International Journal of Human-Computer Studies*, Vol. 64, No. 11, pp. 1071-1083.
 61. 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*.
 62. Wang, C. Y., Yang, H. Y. and Chou S. C. T., 2008, "Using peer-to-peer technology for knowledge sharing in communities of practices," *Decision Support Systems*, Vol. 45, No. 3, pp. 528-540.
 63. Wang, S. and Noe, R. A., 2010, "Knowledge sharing: A review and directions for future research," *Human Resource Management Review*, Vol. 20, No. 2, pp. 115-131.
 64. Wang, S. G. and Li, W., 2010, "Research on sentiment classification problem of sino-Japanese relations forum," *Computer Engineering and Applications*, Vol. 43, No. 32, pp. 174-177.
 65. Wang, S. G., Li, D. Y., Wei, Y. J. and Song, X. L., 2009, "A synonyms based word sentiment orientation discriminating," *Journal of Chinese Information Processing*, Vol. 23, No. 5, pp. 68-74.

66. Wang, X. D., Liu, Q. and Tao, X. J., 2010, "Sentiment ontology construction and text orientation analysis," *Computer Engineering and Applications*, Vol. 46, No. 30, pp. 117-120.
67. Wang, X. D., Liu, Q. and Zhang, Z., 2011, "Topic semantic orientation compute based on sentiment words ontology," *Computer Engineering and Applications*, Vol. 47, No. 27, pp. 147-151.
68. Wang, Z. H. and Jiang, W., 2012, "Online reviews sentiment analysis model based on rough sets," *Computer Engineering*, Vol. 38, No. 16, pp. 1-4.
69. Win, K. T. and Zhang, M., 2006, "Enhancing information quality through a semantic approach in health information retrieval," *International Journal of Electronic Business Management*, Vol. 4, No. 1, pp. 8-15.
70. Xie, L., Zhou, M. and Sun, M., 2012, "Hierarchical structure based Hybrid approach to sentiment analysis of Chinese micro blog and its feature extraction," *Journal of Chinese Information Processing*, Vol. 26, No. 1 pp. 73-83.
71. Xu, C., Wang, M., He, T. T. and Zhang, Y., 2008, "Automatic summarization method based on extracting sentences from local topics," *Computer Engineering*, Vol. 34, No. 22, pp. 49-51.
72. Yang, F, Peng, Q. K. and Xu, T., 2010, "Sentiment classification for online comments based on random network theory," *Acta Automatica Sinica*, Vol. 36, No. 6, pp. 837-844.
73. Yang, J., Peng, S. Y. and Hou, M., 2011, "Recognizing sentiment polarity in Chinese reviews based on top sentiment sentences," *Application Research of Computers*, Vol. 28, No. 2, pp. 569-572.
74. Yin, C. X. and Peng, Q. K., 2012, "Identifying word sentiment orientation for free comments via complex network," *Acta Automatica Sinica*, Vol. 38, No. 3, pp. 389-390.
75. Yu, Z. T., Fan, X. Z., Guo, J. Y. and Geng, Z. M., 2006, "Answer extraction for Chinese question-answer system based on latent semantic analysis," *Chinese journal of computers*, Vol. 29, No. 10, pp. 1888-1893.
76. Zan, H. Y., Zuo, W. S., Zhang, K. L. and Wu, Y. F., 2011, "Sentiment analysis based on rules and statistics," *Computer Engineering and Science*, Vol. 33, No. 5, pp. 146-150.
77. Zhan, P., Zhao, Z. W. and Zhuo, J. W., 2011, "Method of sentence semantic orientation

- distinction based on semantic weighted sentiment word,” *Computer Engineering and Applications*, Vol. 47, No. 35, pp. 161-164.
78. Zhang, C., Zeng, D., Li, J., Wang, F. Y. and Zuo, W., 2009, “Sentiment analysis of Chinese documents: From sentence to document level,” *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 12, pp. 2474-2487.
79. Zhang, J., 2009, “An overview of factors influencing knowledge sharing in virtual communities,” *Journal of Modern Information*, Vol. 29, No. 7, pp. 222-225.
80. Zhang, W., Li, P. and Zhu, Q., 2011, “Research on tree kernel-based sentiment classification in English sentences,” *Computer Applications and Software*, Vol. 28, No. 4, pp. 30-33.
81. Zhang, Y. X., Zhang, M. and Deng, Z. H., 2009, “Feature-driven summarization of customer reviews,” *Journal of Computer Research and Development*, Vol. 46, No. 2, pp. 520-525.
82. Zhao, Y. Y., Bing, Q. and Ting, L., 2010 “Integrating intra- and inter-document evidences for improving sentence sentiment classification,” *Acta Automatica Sinica*, Vol. 36, No. 10, pp. 1417-1425.
83. Zhou, J., Lin, C. and Li, B. C., 2010, “Research of sentiment classification for Netnews comments by machine learning,” *Journal of Computer Application*, Vol. 30, No. 4, pp. 1011-1014.
84. Zhou, L., Tao, Y., Cimino, J. J., Chen, E. S., Liu, H., Lussier, Y. A., Hripesak, G. and Friedman, C., 2006, “Terminology model discovery using natural language processing and visualization techniques,” *Journal of Biomedical Informatics*, Vol. 39, No. 6, pp. 626-636.
85. 李旻芳，2010，「不同質地之組合對情感感受之影響-以木皮與不鏽鋼質地為例」，碩士論文（指導教授：李文淵），大同大學工業設計研究所。
86. 高郁倫，2010，「人格特質對於網路社群知識分享及創造影響情形之研究-以程式設計論壇為例」，碩士論文（指導教授：謝佩璇），國立成功大學資訊管理研究所。
87. 創市際市場研究顧問，「創市際月刊報告書 2010 年 7、8 與 10 月」，http://news.ixresearch.com/?page_id=1534
88. 創市際市場研究顧問，「創市際月刊報告書 2013 年 5 月」，

http://news.ixresearch.com/?page_id=1534

89. 黃信華，2013，「以 Facebook 塗鴉牆文本分析情緒文字的關係」，碩士論文（指導教授：林豪鏘），國立臺南大學數位學習科技學系碩士班。
90. 廖顯忠，2010，「分析隱含於部落格文章的事件、情緒與需求」，碩士論文（指導教授：盧文祥），國立成功大學資訊工程研究所。



附錄(1)、系統功能操作說明

本研究所發展之以論壇文章蘊含情緒為基之文章重組系統乃以本文中第 4.4.1 所提出系統功能流程為依據，並開發「(A)論壇文章維護模組」、「(B)情緒詞彙維護模組」、「(C)文章表達情緒判定模組」及「(D)發文者文章語句重組模組」等四大模組、以下分別針對此四大模組中各功能進行詳細之說明。

(A)論壇文章維護模組

為使權限內使用者方便維護各論壇文章相關資料，本研究乃開發「論壇文章維護模組功能」，並包含論壇文章「新增」、「查詢」、「修改」、「刪除」功能以提供權限內使用者一所需而維護論壇文章資料。其中，「論壇文章新增功能」乃提供權限內使用者新增論壇文章資訊（如文章標題、內容等），同時系統將會自動斷詞文章之內容，並將文章資訊提供於後續核心模組作為計算之基礎資料；「論壇文章查詢功能」乃提供權限內使用者查詢所有論壇文章之資訊，並了解資料庫內所有文章之維護結果與狀況；此外，本模組亦提供「論壇文章修改功能」與「論壇文章刪除功能」功能，以提供權限內使用者進行修改與維護錯誤或缺乏時效性之文章資料，進而保持論壇文章資料之正確性與即時性。

(A-1)論壇文章新增功能一程式說明

權限內使用者可透過「論壇文章新增」功能將文章資訊匯入至系統資料庫中。當權限內使用者執行論壇文章新增功能時，系統乃提供文章資訊輸入介面予權限內使用者，權限內使用者可於該介面中輸入論壇文章標題、內容及作者等資訊，並可透過下拉式選單選擇上傳時間，此外，權限內使用者亦可使用檔案上傳鈕，將論壇文章檔案（如*.mht、*.html 檔）上傳至系統中，如圖 A.1 所示。待完成檔案上傳並新增文章資訊後，權限內使用者即可點選確認鈕，之後系統乃將文章進行斷詞預處理，並根據斷詞內容、斷詞詞性，擷取目標論壇文章之候選事件及語句順序，以供後續核心模組作為計算之基礎資料，如圖 A.2 所示，最後，待權限內使用者確認論壇文章之資訊後即可點選確認新增鈕，系統即會將所有文章資訊維護至資料庫內，並於畫面中顯示新增成功之提示，如圖 A.3 所示。

舉例而言，當權限內使用者點選論壇文章新增時，可依據介面所提供之填入選項，

依序填入論壇文章標題「Camry 03 年避震器...等」、文章內容「3 月份換 4 支 YKYB 避震器...等」、文章作者「tommywoei1」於介面中，同時選擇上傳時間「2014 年 07 月 26 日」，並將選擇欲上傳之「*.mht 檔」論壇文章，權限內使用者即可按下確認鈕，如圖 A.1 所示，之後，系統乃自動將文章內容進行斷詞預處理，並顯示文章斷詞內容為「3 月(Nd) 份(Na) 換(VC) 4(FW) 支(Nf) YKYB(FW)..等」，而系統亦根據詞彙之詞性擷取文章之候選事件為「避震器」，語句順序分佈於第「1」、「2」、「3」、「11」語句中；「原廠」之語句順序分佈於第「7」及「10」語句中；「摟摟」語句順序分佈於第「2」及「8」語句中，如圖 A.2 所示，最後，待權限內使用者確認文章斷詞及候選事件資訊後即可點選確認新增鈕，系統即會顯示新增成功之提示於介面中，如圖 A.3 所示。



圖 A.1、論壇文章資訊上傳

協助論壇文章品質改善平台

社群文章總模組

- 社群文章新增
- 社群文章查詢
- 社群文章修改
- 社群文章刪除
- 文章表達情緒判定模組
- 發文者文章改寫模組
- 情緒詞彙維護模組
- 系統參數設定模組

社群文章資料新增

— 文章斷詞內容與候選事件資訊

文章標題	文章作者	文章字數
Camry 03年避震器異音困擾求助	tommywoei1	175

文章原始內容

3月份換4支YKYB避震器總成，2個月後因右後避震器在時速約10-20公里慢速時會有撲撲的聲音，回原本換避震器的保養廠，他們馬上換1組新的，但是沒多久又有同樣聲音，也再回去，這次換非原廠上座試試看，絕...[繼續閱讀](#)

文章斷詞內容

標題斷詞內容
Camry(FW) 0 3年(Nd) 避震器(Na) 異音(A) 困擾(Na) 求助(VC)

文章斷詞內容

3月(Nd) 份(Na) 換(VC) 4(FW) 支(Nf) YKYB(FW) 避震器(Na) 總(D) 成(VG) ，(COMMACATEGORY) 2(FW) 個(Nf) 月(Na) 後(Ng) 因(Cbb) 右(Ncd) 後(Ng) 避震器(Na) 在(P) 時速(Na) 約(Da) 1 0(Neu) -(FW) 2 0(Neu) 公里(Nf) 慢速(A) 時(Ng) 會(D) 有(V_2) 撲撲(VC) 的(DE) 聲音(Na) ，(COMMACATEGORY) 回(VCL) 原本(D) 換(VC) 避震器(Na) 的(DE) 保養廠(Nc) ，(COMMACATEGORY) 他們(Nh) 馬上(D) 換(VC) 1(FW) 組(Nc) 新(VH) 的(DE) ，(COMMACAT. [繼續閱讀](#)

候選事件列表

候選事件名稱	語句順序	詞性種類
避震器	1	普通名詞
避震器	2	普通名詞
避震器	3	普通名詞
避震器	11	普通名詞
原廠	7	普通名詞
原廠	10	普通名詞
撲撲	2	專有名詞
撲撲	8	專有名詞

圖 A.2、論壇文章資訊擷取

協助論壇文章品質改善平台

社群文章總模組

- 社群文章新增
- 社群文章查詢
- 社群文章修改
- 社群文章刪除
- 文章表達情緒判定模組
- 發文者文章改寫模組
- 情緒詞彙維護模組
- 系統參數設定模組

社群文章資料新增

— 文章資訊新增成功

資訊如下

文章標題	文章作者	文章字數	候選事件數
Camry 03年避震器異音困擾求助	tommywoei1	175	3

文章完整內容

3月份換4支YKYB避震器總成，2個月後因右後避震器在時速約10-20公里慢速時會有撲撲的聲音，回原本換避震器的保養廠，他們馬上換1組新的，但是沒多久又有同樣聲音，也再回去，這次換非原廠上座試試看，結果還是有撲撲的聲音，主要是慢速有點不平的路就有，且也已換過原廠李仔車及平衡桿橡皮，請問是避震器芯，上座或底盤其他橡皮問題嗎？其他3支好像沒有聽到異音。

文章完整段詞內容

3月(Nd) 份(Na) 換(VC) 4(FW) 支(Nf) YKYB(FW) 避震器(Na) 總(D) 成(VG) ，(COMMACATEGORY) 2(FW) 個(Nf) 月(Na) 後(Ng) 因(Cbb) 右(Ncd) 後(Ng) 避震器(Na) 在(P) 時速(Na) 約(Da) 1 0(Neu) -(FW) 2 0(Neu) 公里(Nf) 慢速(A) 時(Ng) 會(D) 有(V_2) 撲撲(VC) 的(DE) 聲音(Na) ，(COMMACATEGORY) 回(VCL) 原本(D) 換(VC) 避震器(Na) 的(DE) 保養廠(Nc) ，(COMMACATEGORY) 他們(Nh) 馬上(D) 換(VC) 1(FW) 組(Nc) 新(VH) 的(DE) ，(COMMACATEGORY) 但是(Cbb) 沒(VJ) 多久(Nd) 又(D) 有(V_2) 同樣(VH) 聲音(Na) ，(COMMACATEGORY) 也(D) 再(D) 回去(VA) ，(COMMACATEGORY) 這(Nep) 次(Nf) 換(VC) 非(A) 原廠(Nc) 上座(Na) 試試看(VE) ，(COMMACATEGORY) 結果(Dk) 還是(D) 有(V_2) 撲撲(VC) 的(DE) 聲音(Na) ，(COMMACATEGORY) 主要(D) 是(SHI) 慢速(A) 有點(Dfa) 不平(VH) 的(DE) 路(Na) 就(D) 有(V_2) ，(COMMACATEGORY) 且(Cbb) 也(D) 已(D) 換(VC) 過(Di) 原廠(Nc) 李仔車(Nb) 及(Caa) 平衡桿(Na) 橡皮(Na) ，(COMMACATEGORY) 請問(VE) 是(SHI) 避震器(Na) 芯(Na) ，(COMMACATEGORY) 上座(Na) 或(Caa) 底盤(Na) 其他(Neqa) 橡皮(Na) 問題(Na) 嗎(T) ?(QUESTIONCATEGORY) 其他(Neqa) 3(FW) 支(Nf) 好像(D) 沒有(D) 聽到(VE) 異音(A) 。(PERIODCATEGORY)

圖 A.3、論壇文章新增成功

(A-2)論壇文章查詢功能一程式說明

為方便使用者查詢論壇文章資料，本研究乃開發「論壇文章查詢」功能，以提供權限內使用者查詢已上傳至系統之論壇文章資訊。當權限內使用者執行論壇文章查詢功能時，系統介面將提供查詢條件、查詢關鍵字及查詢運算式(And)及(Or)等篩選條件，供權限內使用者輸入並選擇所需之論壇文章，待權限內使用者輸入關鍵字、選擇查詢條件及運算式，並按下查詢鈕後，如圖 A.4 所示，系統即會呈現查詢結果於介面下方，如圖 A.5 所示，此外，權限內使用者亦可點選「查詢更多」之連結，系統會以彈跳使窗方式呈現目標論壇文章之詳細內容，以使權限內使用者更了解目標論壇文章之資訊，如圖 A.6 所示。

舉例而言，當權限內使用者執行「論壇文章查詢」功能時，系統乃提供查詢條件及關鍵字欄位供使用者輸入以取得所需之文章。當權限內使用者輸入論壇文章搜尋關鍵字為「Camry」及「紐約」，同時選擇查詢條件分別為「文章標題」及「文章內文」、運算式條件為「Or」條件式(如圖 A.4 所示)，並按下查詢鈕後，系統即會呈現文章編號「E01」及「E02」之文章資訊於介面下方中，其中文章標題分別為「Camry 03 年避震器..等」及「下雪天開車，可怕的經驗」；文章內容分別為「3 月份換 4 支 YKYB 避震器總..等」與「小弟住的紐約天氣都算穩定...等」，如圖 A.5 所示，此外，權限內使用者亦可點擊「查詢更多」之連結，以得知目標文章編號「E01」之文章斷詞內容為「3 月(Nd) 份(Na) 換(VC) 4(FW) 支(Nf)..等」，如圖 A.6 色框處所示。



圖 A.4、論壇文章查詢條件



圖 A.5、論壇文章查詢結果



圖 A.6、論壇文章詳細資訊

(A-3)論壇文章修改功能－程式說明

論壇文章修改功能乃提供權限內使用者修改錯誤之論壇文章資訊，以維持資料庫內文章之正確性。當權限內使用者執行「論壇文章修改」功能時，可於搜尋條件中輸入關鍵字，同時以下拉式選單方式選擇查詢條件（如圖 A.7 所示），並點選查詢鈕後，系統將依照搜尋條件，將查詢結果呈現於介面下方，如圖 A.8 所示，之後，權限內使用者可選取欲修改文章之「資料修改」連結，系統會以彈跳式視窗方式呈現修改介面，其中，資訊修改之內容僅可修改文章標題及內容（即註解**之內容），如圖 A.9 所示，文章其他資訊（如文章字數及斷詞內容）乃由系統自動抓取並維護至系統中，待權限內使用者修改文章標題或內容完畢並按下確認修改鈕後，系統會呈現修改成功之提示於介面中，即完成論壇文章修改之維護動作，如圖 A.10 所示。

舉例而言，權限內使用者執行「論壇文章修改」功能後，依照搜尋條件分別輸入搜尋關鍵字「Camry」、查詢條件為「文章標題」並按下查詢鈕後（如圖 A.7 所示），系統乃依據搜尋條件取得文章編號「E01」、文章標題「Camry 03 年避震器...等」、文章內容「3 月份換 4 支 YKYB 避震器總成...等」之文章資訊呈現於介面下方（如圖 A.8 所示），接著，權限內使用者可點擊目標文章之資料修改連結，系統會以彈跳視窗方式提供使用者修改文章資訊之介面，如圖 A.9 所示，當權限內使用者將文章標題「Camry 03 年避震器」修改為「Camry 2008 年避震器」並按下確認修改鈕後，系統即會呈現修改成功之提示，並將修改後之文章內容呈現於介面中，如圖 A.10 所示，即完成論壇文章修改之維護。



圖 A.7、論壇文章資料查詢



圖 A.8、選取欲修改之論壇文章



圖 A.9、修改目標文章內容

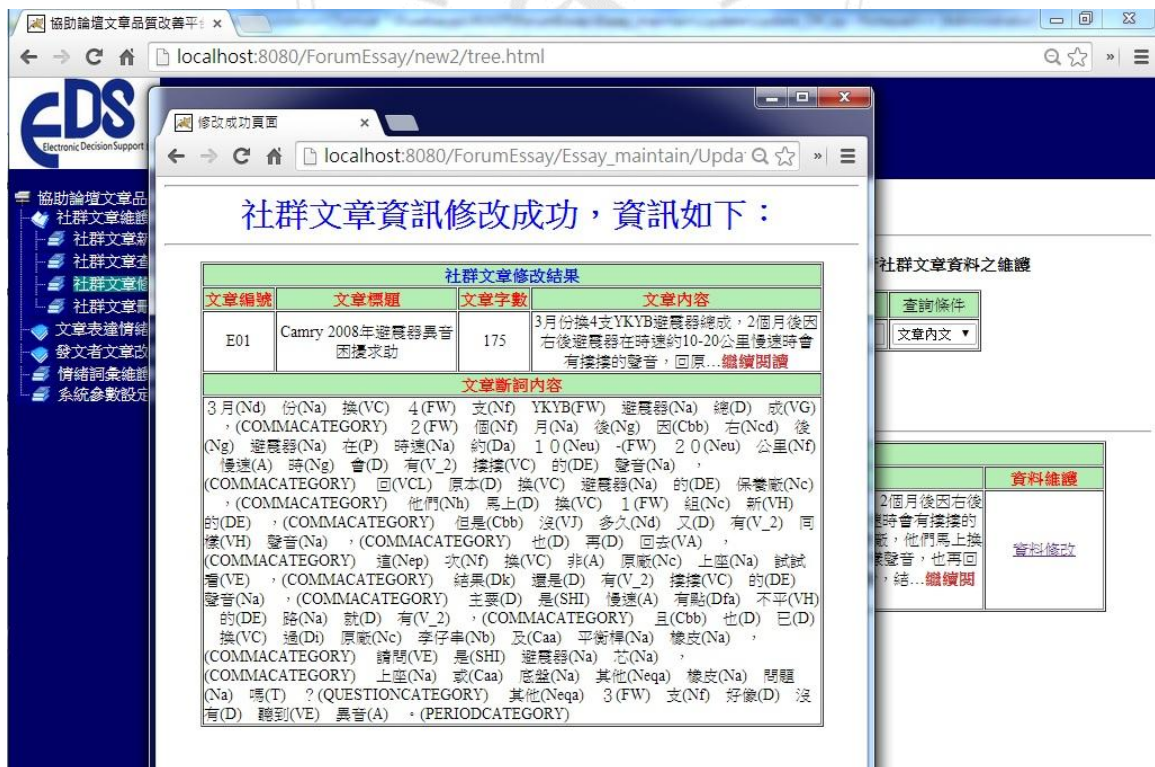


圖 A.10、論壇文章修改成功介面

(A-4)論壇文章刪除功能一程式說明

論壇文章刪除功能乃提供權限內使用者刪除錯誤資訊或缺乏實質性之論壇文章，以維持系統資料庫內文章之正確性與品質。當權限內使用者執行論壇文章刪除功能後，首先乃依據查詢條件依序輸入搜尋關鍵字、查詢條件或運算式等搜尋條件，如圖 A.11 所示，並按下查詢鈕後，系統將依照搜尋條件取得符合使用者所需之文章資訊，並將文章資訊呈現於介面下方，如圖 A.12 所示，之後，權限內使用者可點擊「資料刪除」連結，系統乃呈現目標論壇文章詳細資訊，以供權限內使用者再確認文章之詳細資訊，以避免使用者誤刪除正確之文章，如圖 A.13 所示。待確認完畢後，使用者即可點擊確認刪除鈕，系統會立即刪除目標論壇文章於系統資料庫之相關資訊，並以刪除成功之提示呈現於介面中，即完成論壇文章刪除之維護，如圖 A.14 所示。

舉例而言，當權限內使用者執行論壇文章刪除功能時，依照搜尋條件輸入關鍵字「MRT」、查詢條件「文章內文」，並按下查詢鈕後（如圖 A.11 所示），系統乃依照搜尋條件取得符合使用者所需文章為編號「E03」、標題「為什麼一定要 MRT」等文章資訊呈現於介面下方，如圖 A.12 所示，之後，權限內使用者可點擊目標論壇文章之「資料刪除」連結，系統乃以彈跳式視窗呈現目標論壇文章之詳細資訊，以供使用者確認目標論壇文章之內容，進而避免誤刪除非錯誤之論壇文章，如圖 A.13 所示。待文章資訊確認完畢後，權限內使用者即可點即確認刪除鈕後，系統即會刪除目標論壇文章編號「E03」、標題「為什麼一定要 MRT」之相關資訊，即可完成論壇文章刪除之維護，如圖 A.14 所示。



圖 A.11、輸入論壇文章查詢條件



圖 A.12、論壇文章查詢結果



圖 A.13、論壇文章確認刪除介面



圖 A.14、論壇文章資料刪除成功介面

(B)情緒詞彙維護模組

本研究所開發之情緒詞彙維護模組乃提供權限內使用者維護情緒詞彙之相關資料，其中包含情緒詞彙「新增」、「查詢」、「修改」及「刪除」功能。情緒詞彙「新增」功能乃提供權限內使用者新增情緒詞彙，並將情緒詞彙維護至資料庫中；「查詢」功能乃提供權限內使用者藉由關鍵字及條件式查詢目標情緒詞彙之資訊；「修改」及「刪除」功能乃提供權限內使用者將錯誤或不正確之情緒詞彙資訊予以修改及刪除，以保持資料庫內情緒詞彙資料之完整性與正確性。

(B-1)情緒詞彙新增功能一程式說明

當權限內使用者執行情緒詞彙新增功能時，系統即會提供情緒詞彙輸入欄位供使用者輸入欲新增之情緒詞彙，同時可透過新增欄位鈕擴增情緒詞之輸入欄位數，如圖 B.1 所示，接著點選確認鈕後，系統會自動判斷情緒詞彙於資料庫內重複與否，並於畫面下方顯示新增狀況，如圖 B.2 所示。

舉例而言，當權限內使用者執行情緒詞彙新增功能時，分別輸入「煩人」、「可恨」、

「快樂」及「傷心」4 個情緒詞彙，如圖 B.1 所示，並按下確認鈕後，系統會自動將新增之情緒詞與資料庫之情緒詞彙進行比對，並得知情緒詞彙「煩人」及「可恨」為資料庫內尚未維護之情緒詞彙，而「快樂」及「傷心」為重複新增之情緒詞彙，如圖 B.2 所示



圖 B.1、情緒詞彙新增介面



圖 B.2、情緒詞彙新增狀況

(B-2)情緒詞彙查詢功能一程式說明

當權限內使用者執行查詢功能時，系統即呈現條件式與關鍵字搜尋介面供使用者查

詢欲得知之情緒詞彙，如圖 B.3 所示，待權限內使用者輸入完畢後，系統會於介面下方呈現查詢結果，並可得知情緒詞彙判定情緒類別隸屬係數與否（類別隸屬係數乃由核心模組之功能所判定之結果），如圖 B.4 所示，同時，權限內使用者亦可點選詳細資料查看類別隸屬係數結果，系統會以彈跳視窗方式呈現情緒詞彙對應情緒類別之隸屬係數，如圖 B.5 所示。

舉例而言，當權限內使用者於查詢介面中以空白字串之條件進行查詢，如圖 B.3 所示，系統即會呈現資料庫所有之情緒詞彙，其中已判定類別隸屬係數之情緒詞為「高興」，其餘情緒詞彙皆為判定情緒類別隸屬係數，如圖 B.4 所示，同時，權限內使用者可點選「詳細資料」連結，系統將會呈現目標情緒詞彙對應情緒類別之係數，其中，高興情緒詞對於「快樂」情緒類別之隸屬係數為「1.0」，其餘情緒類別之隸屬細數皆為「0.0」，如圖 B.5 所示。



圖 B.3、情緒詞彙查詢介面



圖 B.4、情緒詞彙查詢結果

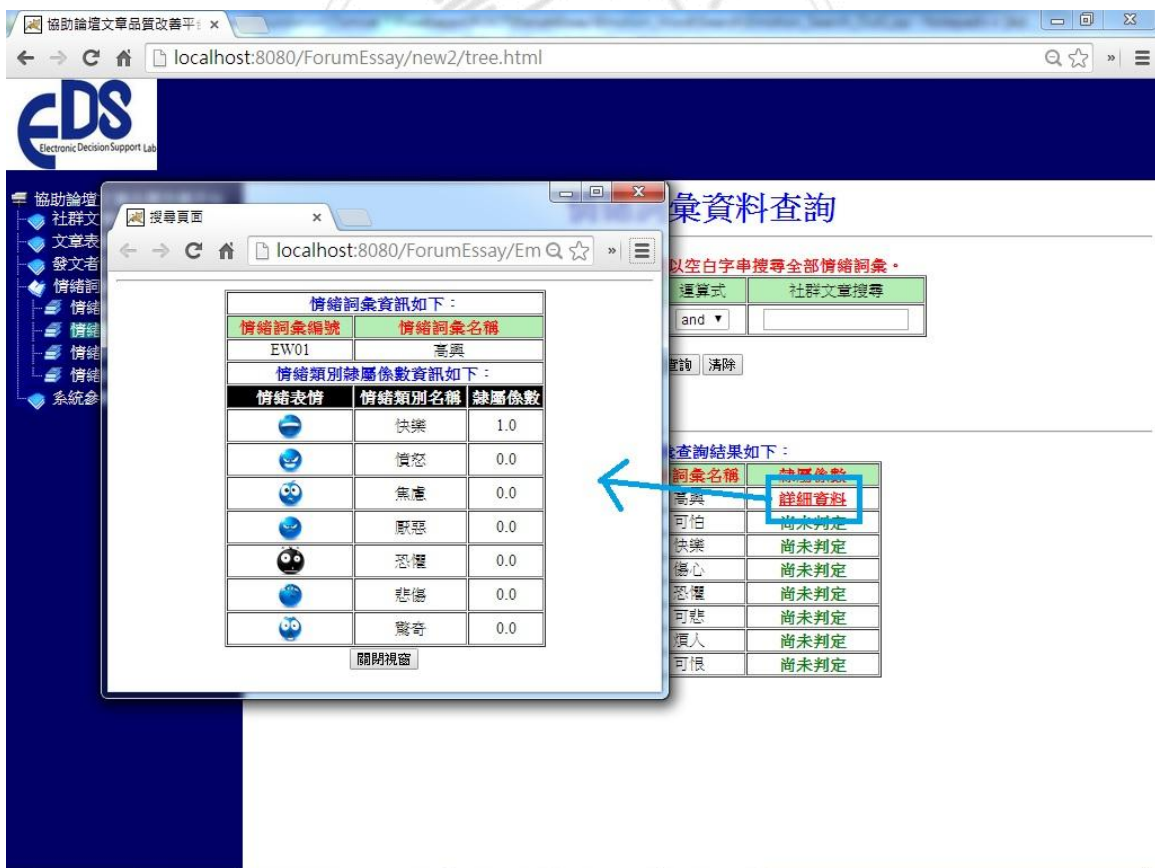


圖 B.5、情緒詞彙之類別隸屬係數結果

情緒詞彙修改功能

當權限內使用者執行修改功能時，系統即呈現條件式與關鍵字搜尋介面供使用者查詢欲修改之情緒詞彙，如圖 B.6 所示，待權限內使用者輸入完畢後，系統會於介面下方呈現查詢結果，並可透過文字格重新修改情緒詞彙之名稱，如圖 B.7 所示，使用者完成修改並按下確認修改鈕後，系統會根據權限內使用者修改之內容維護於資料庫中，同時，介面亦會呈現情緒詞彙修改前後之結果彈，如圖 B.8 所示。

舉例而言，當權限內使用者執行修改功能時，系統即呈現查詢介面予以使用者並依序輸入關鍵字「高」及「可」，條件式為「OR」後點擊查詢鈕，如圖 B.6 所示，系統即會呈現情緒詞彙「高興」、「可怕」、「可悲」及「可恨」情緒詞彙，同時，權限內使用將情緒詞「高興」修改為「高興的」、「可悲」修改為「可悲地」並按下確定修改鈕後，如圖 B.7 所示，系統會根據修改內容維護至資料庫中，且於介面中亦可查看情緒詞彙修改前與修改後之內容，如圖 B.8 所示。



圖 B.6、情緒詞彙查詢介面



圖 B.7、情緒詞彙修改介面



圖 B.8、情緒詞彙修改結果介面

(B-3)情緒詞彙刪除功能一程式說明

當權限內使用者執行刪除功能時，系統即會呈現條件式與關鍵字搜尋介面供使用者查詢欲刪除之情緒詞彙，如圖 B.9 所示，待權限內使用者輸入完畢後，系統會於介面下方呈現查詢結果，並可透過詳細資料連結查詢情緒詞彙詳細資料，接著，使用者可勾選

欲刪除之情緒詞彙並按下確定刪除鈕後，如圖 B.10 所示，系統會將權限內使用者指定之情緒詞彙資料於資料庫中予以刪除，同時，系統介面亦會呈現刪除成功之提示，如圖 B.11 所示。

舉例而言，當權限內使用者執行刪除功能時，系統即呈現查詢介面予以使用者並以空白字串條件後點擊查詢鈕，如圖 B.9 所示，系統即會呈現情緒詞彙「高興」、「可怕」、「可悲」、「可恨」...等 7 個情緒詞彙，之後權限內使用者勾選欲刪除之情緒詞彙「可怕」與「開心」並按下確定刪除鈕後，如圖 B.10 所示，系統會自動將目標情緒詞彙於資料庫中所有相關資料予以刪除，同時，系統亦會呈現刪除成功之提示於介面中，如圖 B.11 所示。



圖 B.9、情緒詞彙查詢介面



圖 B.10、情緒詞彙查詢結果及詳細資料



圖 B.11、情緒詞彙刪除結果

(C)文章情緒表達判定模組

本研究開發之「文章表達情緒判定模組」乃以論壇發文者文章為分析之基礎，並藉由文章各候選事件之前後文配對、標題加權與詞性相關性三項分數，計算候選事件之文章比重值以取得文章代表事件，再根據代表語句（即具有代表事件之語句）分析所有語

句之相似度，並計算相似語句之情緒機率值，即可取得代表語句之各情緒類別之穩定值，同時以代表語句之情緒類別作為文章之情緒類別。「文章表達情緒判定模組」乃分為「(C-1)文章代表事件解析功能」、「(C-2)情緒詞彙隸屬係數判定功能」、「(C-3)文章表達情緒判定功能」三大功能，其程式運作步驟如圖 C.1 所示。

「(C-1)文章代表事件解析功能」乃藉由斷詞將文章所有詞句進行分詞，並從文章中挑選具專有名詞或地方詞之代表詞彙作為「候選事件」，系統透過文章各候選事件之前後文分佈比例並取平均值以取得候選事件之前後文配對分數，同時根據候選事件與標題之關係計算標題加權分數，再依據文章之詞性種類數分析各候選事件所含詞性之分佈比例以獲取詞性相關性分數，系統即可將各候選事件之三項比重分數相加並以最大值者作為文章之代表事件。之後即可執行「(C-2)情緒詞彙隸屬係數判定」功能，此功能先利用基礎情緒詞建構對應之情緒語句，系統將統計情緒詞彙於各情緒語句之詞頻數，以初步取得情緒詞彙與各類別之關係係數後，系統即會將情緒詞彙予以正規化並取得情緒詞彙與各類別之類別隸屬係數。最後，「(C-3)文章表達情緒判定」功能首先利用向量空間相似法計算代表語句與訓練語句之關係，以獲取與代表語句相似之語句，系統同時可透過相似語句內情緒與重要詞彙所組合之極性項目，取得與文章較相關之相似語句，系統並可針對此相似語句計算情緒偏向機率值，再根據所有代表語句所含之相似語句計算情緒穩定值，即可取得代表語句之情緒類別並以此作為文章之情緒類別。



圖 C.1、文章表達情緒判定模組各功能步驟運作圖

(C-1) 文章代表事件解析功能—程式說明

➤ 程式運作邏輯：

權限內使用者執行文章代表事件解析功能時，首先可透過搜尋字串及搜尋條件查詢欲解析之論壇文章，系統即呈現符合條件之文章供權限內使用者選擇（如圖 C.2 所示），之後系統將文章進行斷詞處理，並以標點符號為劃分依據將語句分割，並同時抓取所有候選事件後，系統即會比對候選事件所屬之語句位置，如圖 C.3 所示，待系統完成文章斷詞及語句解析後，系統根據文章相異候選事件之個數、各候選事件於標題之頻率，計算所有候選事件於不同語句位置之標題加權分數（如圖 C.4 所示），之後系統將擷取候選事件前後各一個語句之詞彙，計算各詞彙於前後語句之分佈值，並將前後語句之詞彙分佈值相加即為前後文配對分數，如圖 C.5 所示，接著，系統將依據系統之詞性距離參數值（自然數），擷取候選事件前文參數數值之詞彙（如參數值為 6 即擷取前 6 個詞彙），並計算詞彙於所有候選事件前文之分佈比例並平均後即可取得詞性相關性分數，如圖 C.6 所示，待系統完成三項分數之解析後，系統即會將各候選事件於不同語句位置之三項分數加總並平均，並將各候選事件之三項總分予以正規化後，權限內使用者即可得知候選事件對於文章之代表程度，如圖 C.7 所示。

➤ 程式運作流程：

舉例而言，當權限內使用者選擇標題「台灣與 FBI 犯罪心理分析官」、內容「這個頭銜或者...等」之文章後（如圖 C.2 所示），系統即自動抓取目標文章之候選事件「犯罪」、「家政」、「台灣」、「治安史」，並將文章以「，」、「、」、「。」等標點符號為依據分割語句，以得知候選事件「台灣」位於語句順序之位置為「7」、「3」，如圖 C.3 所示，之後權限內使用者點選解析標題加權分析鈕後，系統將計算目標文章相異候選事件個數為「4」個，並可得知候選事件「台灣」於標題中出現「1」次後，系統即會計算候選事件「台灣」於語句順序「3」、「7」、「9」、「17」、「26」、「29」中標題加權分數皆為「0.16」，如圖 C.4 所示，接著，權限內使用者點選解析前後文配對分數解析鈕後，系統將計算候選事件「台灣」、語句順序為「3」中前後語句詞彙之分佈值，分別取得「0.18」與「0.22」並相加兩分佈值形成此候選事件之前後文配對分數「0.4」，如圖 C.5 所示，之後權限內使用者點擊解析詞性相關性分數鈕後，系統將自動取得詞性參數值為 6，並擷取語句順序「3」、候選事件為「台灣」之前 6 個詞彙分別為「對」、「一樣」、「職業」、「個」、「這」、「婦」，同時系統將會計算此 6 個詞彙於所有候選事件前文之分佈平均值為「0.24」，如圖 C.6 所示，待完成三項分數之解析後，權限內使用者即可執行文章代表事件解析鈕，系統即會將候選事件「台灣」依據語句順序所取得之三項總分分別為「0.61」、「0.59」、「0.91」、「0.87」、「0.69」、「0.72」，同時將此分數加總平均取得「0.73」，並將此值予以正規化後即可得知候選事件「台灣」之代表分數為「0.196」，其中候選事件「犯罪」之代表分數「0.459」為最大值，故可得知「犯罪」為文章之代表事件，如圖 C.7 所示。



圖 C.2、文章查詢介面



圖 C.3、文章斷詞與語句分割介面



圖 C.6、候選事件詞性相關性分數解析介面



圖 C.7、文章代表事件解析介面

(C-2)情緒詞彙隸屬係數判定功能一程式說明

➤ 程式運作邏輯：

待完成「(C-1)文章代表事件解析」功能後，權限內使用者即可執行情緒詞彙隸屬係數判定功能，系統會擷取訓練文章中含有情緒詞彙之語句，如圖 C.8 所示，同時系統將

區分已判定及未判定類別隸屬係數之詞彙，並統計各情緒類別之詞彙數、詞彙於各情緒類別語句之頻率數，以初步計算「未判定情緒詞彙」之情緒類別關係係數，如圖 C.9 所示，最後，系統根據詞彙與各情緒類別之關係係數予以正規化，即可取得各情緒詞彙與情緒類別之類別隸屬係數，如圖 C.10 所示。

➤ 程式運作流程：

舉例而言，權限內使用者執行情緒詞彙隸屬係數判定功能，系統首先將取得情緒詞彙「欣賞」之所屬語句為「日前她出席電影節活動時表示很期待欣賞入圍影片」，並將此情緒語句儲存於資料庫中，如圖 C.8 所示，之後，系統將區分已判定情緒類別隸屬係數之詞彙個數為「7」，未判定詞彙數為「8」，並依據情緒類別之詞彙數、詞彙於情緒類別語句之頻率數，計算情緒詞彙「驚訝」於情緒類別「恐懼」、「驚奇」、「厭惡」與「焦慮」之關係係數，分別為「0.1」、「0.77」、「0.3」、「0.2」，而剩餘 3 個情緒類別之關係係數皆為「0」，如圖 C.9 所示，最後，系統乃將詞彙之關係係數正規化後即可得知情緒詞彙「驚訝」最大值類別隸屬係數為「0.56」，所對應之情緒類別為「驚奇」，權限內使用者亦可點選「查看」鈕得知詞彙「驚訝」於各情緒類別之隸屬分佈情況，如圖 C.10 所示。

The screenshot shows the EDS (Electronic Decision Support Lab) interface. The main title is "情緒詞彙類別隸屬係數判定" (Emotion Word Class Membership Coefficient Judgment) with the subtitle "取得各情緒類別之情緒語句" (Obtain emotion sentences for each emotion class). The interface is divided into three steps: Step 1 (selected), Step 2, and Step 3.

文章資訊 (Article Information):

文章篇數	文章平均語句數
7	28.5

情緒類別資訊 (Emotion Class Information):

情緒類別名稱	情緒詞彙個數
高興	1
悲傷	0
憤怒	3
恐懼	0
驚奇	0
厭惡	0
焦慮	3

情緒語句解析結果 (Emotion Sentence Analysis Results):

文章編號	情緒詞彙	語句內容 (藍色詞彙為情緒詞彙)
E01	可喜可賀	可喜可賀也就是說台灣變態殺人狂很罕見
E01	坐立難安	陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是壞壞不特定殺人
E03	憤怒	「我們應該憤怒」
E05	欣賞	日前她出席電影節活動時表示很期待欣賞入圍影片

圖 C.8、分析訓練文章所含之情緒語句



圖 C.9、各情緒詞彙之關係係數解析結果



圖 C.10、各情緒詞彙與類別隸屬係數解析結果

(C-3)文章表達情緒判定功能一程式說明

➤ 程式運作邏輯：

待完成(C-2)情緒詞彙隸屬係數判定後，權限內使用者即可執行文章表達情緒判定功能，系統將判斷文章是否已判定代表事件，如圖 C.11 所示，同時系統根據文章代表事

件，取得含有代表事件之語句後，如圖 C.12 所示，系統即利用向量空間模型之餘弦函數，計算代表語句與訓練文章所有語句之相似值，並根據系統門檻值取得代表語句之相似語句，如圖 C.13 所示，接著系統將判斷相似語句中是否存在轉折詞、否定詞與否，以篩選與代表語句語意不符之語句，並根據相似語句中所有之情緒詞與各類別隸屬係數值，分析相似語句與各情緒類別之情緒機率值，如圖 C.14 所示，之後系統乃依據代表語句對應之相似語句，計算各代表語句於各情緒類別之情緒穩定值，並以最趨近於 0 之情緒穩定值作為代表語句之情緒類別，且因文章具有多個代表語句，因此系統將依據文章所有代表語句對應之情緒類別，作為文章表達之情緒類別，如圖 C.15 所示。

➤ 程式運作流程：

舉例而言，權限內使用者可執行文章表達情緒判定功能，同時勾選已判定代表事件之文章內容為「這個頭銜或者說職業很特別吧！...等」、代表事件為「台灣」（如圖 C.11 所示），並按下取得相似語句解析鈕後，系統乃分析文章中具有代表事件「台灣」之語句有「對台灣來講都很陌生」、「可喜可賀也就是說台灣變態殺人狂很罕見」等六個語句，如圖 C.12 所示，接著，系統利用向量空間模型之餘弦函數，計算代表語句「對台灣來講都很陌生」與相似語句「而且被害人有特定連結但不認識的陌生人」之相似值為「0.31」，如圖 C.13 所示，之後權限內使用者即可點選解析相似語句情緒值鈕，系統首先判斷相似語句「可喜可賀也就是說台灣變態殺人狂很罕見」中是否存在轉折詞與否定詞，並依據否定值參數值「2」及「轉折詞存在與否」等篩選條件保留相似語句「可喜可賀..」，同時根據相似語句「可喜可賀也就是說」中情緒詞與類別隸屬係數關係，分析該語句於「高興」情緒類別之情緒機率值為「0.53」、於「憤怒」之情緒機率值為「0.28」、於「恐懼」之情緒機率值為「0.19」，而其餘之情緒類別之情緒機率值皆為「0」，如圖 C.14 所示，最後，權限內使用者點擊取得文章情緒類別鈕後，系統將依據代表語句「陳進興算是台灣「治安史」上唯一...等」對應之相似語句，計算「恐懼」情緒類別之情緒穩定值為「-0.726」、「驚奇」情緒類別為「-0.202」、「焦慮」情緒類別為「-0.152」，並以最趨近於 0 之穩定值「-0.152」得知代表語句「陳進興算是台灣「治安史」上唯一...等」之情緒類別為「焦慮」，系統亦透過目標文章所有代表語句所對應之情緒類別，判定文章內容「這個頭銜或者說職業很特別吧...等」所表達之情緒類別為「焦慮」及「高興」，如圖 C.15 所示。



圖 C.11、選擇已判定代表事件之文章



圖 C.12、取得文章語句中具有代表事件之語句

http://localhost:8080/ForumEssay/Avec_index1.htm

協助論壇文章品質改善平台

文章表達情緒判定
—取得文章代表語句之相似語句

步驟一 步驟二 步驟三

代表事件
【台灣】

系統相似參數值 參數說明
0.2 「相似門檻值」為代表語句與相似語句間相似值之門檻參數，用以精確取得與代表語句相似之語句。

文章代表語句
對台灣來講都很陌生
可喜可賀也就是說台灣變態殺人狂很罕見
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人
台灣的殺人犯都是有特定目標的行為
我們台灣社會整體還算健全
不然台灣有一天也會誕生像美國

分析相似語句

相似語句解析結果(僅列出高於門檻值之相似語句)

代表語句	相似語句	相似值
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	可喜可賀也就是說台灣變態殺人狂很罕見	0.25
可喜可賀也就是說台灣變態殺人狂很罕見	陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	0.25
對台灣來講都很陌生	而且被害人有特定連結但不認識的陌生人	0.31

圖 C.13、取得文章代表語句之相似語句

http://localhost:8080/ForumEssay/Avec_index1.htm

協助論壇文章品質改善平台

文章表達情緒判定
—解析相似語句之情緒機率值

步驟一 步驟二 步驟三

篩選相似語句

代表事件	相似語句數
【台灣】	3

否定詞參數值 參數說明
2 「否定詞參數值」為篩選與代表語句語意相反之相似語句，若否定詞出現頻率大於參數值則去除該相似語句。

相似語句	否定詞數	是否有轉折詞	是否保留
可喜可賀也就是說台灣變態殺人狂很罕見	0	無	✓
陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人	1	無	✓
而且被害人有特定連結但不認識的陌生人	1	無	✓

分析情緒機率值

相似語句情緒機率值解析結果(紅色字體為最大值)

相似語句	高興	悲傷	憤怒	恐懼	驚奇	厭惡	焦慮
可喜可賀也就是說... ...繼續閱讀	0.53	0	0.28	0.19	0	0	0
陳進興算是台灣「... ...繼續閱讀	0	0	0.19	0.1	0	0	0.71
而且被害人有特定連結但不認識的陌生人	0	0	0	0	0	0	0

圖 C.14、取得相似語句之情緒機率值



圖 C.15、取得文章表達之情緒類別

(D)發文者文章語句重組模組

本研究開發之「發文者文章語句重組模組」乃以論壇文章為分析之基礎，先行分析論壇文章所使用之詞語種類數以給予初始分數，並以投票演算法計算文章間之相似程度，同時以累加式方式分析論壇文章評閱分數，以判斷論壇文章是否須進行後續改寫內容之程序，接著依據語句內詞彙之延伸定義，去除語意相近之句子，以取得文章之重要語句，同時分析文章之重要語句與訓練語句之相似性，將與重要語句相似之語句彙集成多重組合句，最後，透過多重組合句之組合形成數個論壇文章之語句架構，依據各語句架構之評閱分數最高者作為目標論壇文章之新語句架構。「發文者文章語句重組模組」乃分為「(D-1)文章評閱分數判定功能」、「(D-2)文章多重組合句建立功能」、「(D-3)文章語句架構重組功能」三大功能，其程以取得文章之重要語句式運作步驟如圖 D.1 所示。

「(D-1)文章評閱分數判定功能」首先乃去除文章重複出現之詞語，系統將相同之詞語視為1種詞語種類以取得初始分數，接著，系統根據目標論壇文章與訓練文章之共用詞語分析相似標準差，再依據共用詞語之語意差異性，解析目標論壇文章之評閱分數，並依據五種門檻值判斷目標論壇文章語句重組之合適性。之後即可執行「(D-2)文章多重組合句建立功能」功能，系統將利用中英翻譯器取得語句中各詞性詞彙之英文詞彙，並依據 WordNet 建立英文詞彙之延伸定義，系統即會計算各語句中中英文詞彙之歧義相似性，以將相似之語句去除，接著，系統乃標定論壇文章之語句位置，並計算文章語句與

訓練語句之相似性，將相似程度最高者作為候選填充句，同時系統將去除不與論壇文章原先表達情緒一致之候選填充句，此外，系統亦將候選填充句之關鍵詞與訓練文章進行比對，以替換與候選填充句之關鍵詞，進而形成論壇文章之多重組合句。之後執行「(D-3)文章表達情緒判定」功能後，系統將依據文章語句之順序組合多重組合句，以形成數個文章語句架構，最後，為能從中取得語句架構較完整之組合，系統將重新評閱各語句架構之分數，並以最高評閱分數之語句架構，作為論壇文章語句重組後之內容。

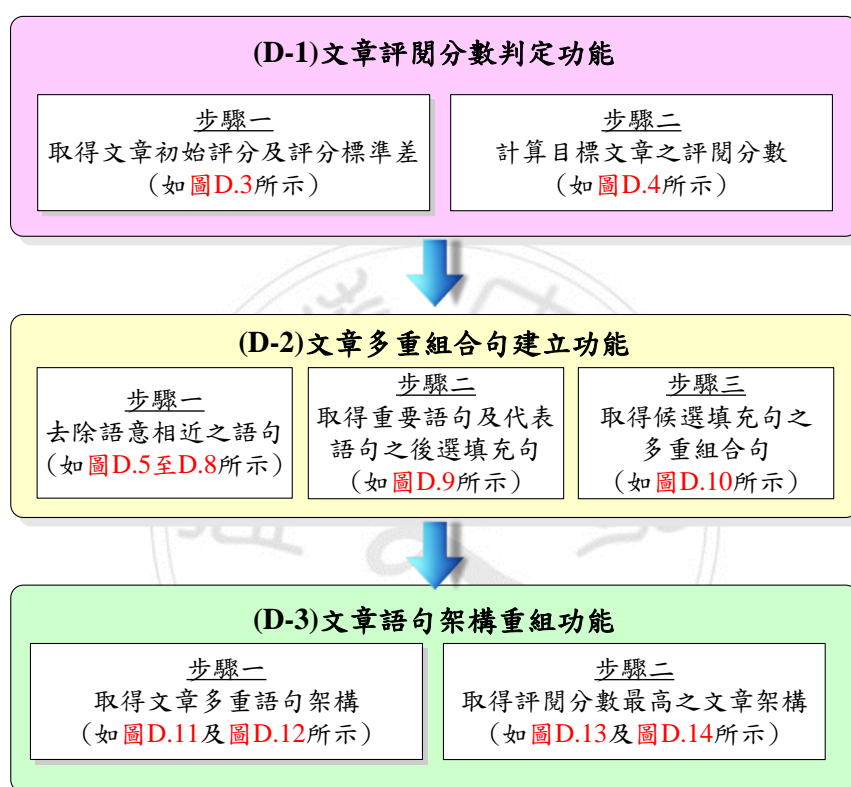


圖 D.1、發文者文章語句重組模組各功能步驟運作圖

(D-1)文章評閱分數判定功能

➤ 程式運作邏輯：

權限內使用者執行文章評閱分數判定功能時，首先可透過搜尋字串及搜尋條件查詢欲解析之論壇文章，系統即呈現符合條件之文章供權限內使用者選擇(如圖 D.2 所示)，之後系統將根據目標論壇文章之相異詞語數計算初始分數，同時系統利用訓練文章之相異詞語數分析評閱分數之標準差，如圖 D.3 所示，之後系統將根據評閱標準差計算目標論壇文章與訓練文章之語意差異性，同時與訓練文章之詞彙進行相似性比較，以取得兩文章之相似性，最後，系統再依據各訓練文章之相似值與語意差異程度，分析目標訓練

文章之評閱分數，如圖 D.4 所示。

➤ 程式運作流程：

舉例而言，當權限內使用者選擇標題「台灣這些球員還嫌看球的觀眾太多嗎...等」、內容「2012 觀護盃籃球賽是第 10 季 SBL 熱身賽，雖然戰績...等」之文章後（如圖 D.2 所示），系統即分析文章中不同之詞語數有「172」個，同時根據訓練文章之相異詞語數計算評分標準差為「10」，如圖 D.3 所示，之後權限內使用者即可點選分析文章評閱分數鈕後，系統將依據評分標準差計算目標論壇文章與訓練文章內容「最近親人都在問說...等」之語意差異程度為「-12.7」、共同詞語數為「29」、與訓練文章內容「他就說不高興那就不要!分手阿...等」之語意差異程度為「12.7」、共同詞語數為「49」，最後，系統乃依據語意差異程度與共同詞語數計算目標論壇文章評閱分數為「264」，如圖 D.4 所示。



圖 D.2、文章查詢介面



圖 D.3、文章初始分數計算介面

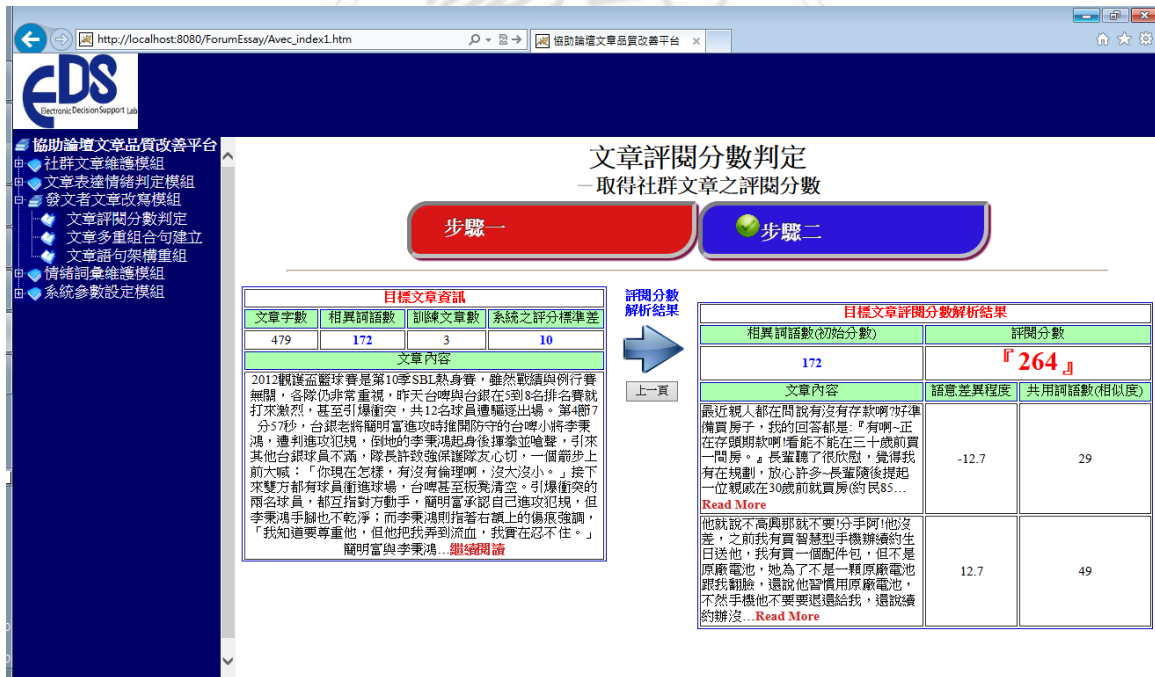


圖 D.4、文章評閱分數計算介面

(B-2)文章多重組合句建立功能

➤ 程式運作邏輯：

待完成文章評閱分數判定功能後，權限內使用者即可執行文章多重組合句建立功能，首先系統利用中英翻譯器 (Denisowsl's CEDICT) 將目標文章內重要語句之詞彙翻

譯為英文詞彙，以得知重要語句（即含有候選事件之語句）代表詞彙之英文定義，如圖 D.5 所示，之後系統透過 WordNet 詞彙庫取得代表詞彙之延伸英文定義，以更廣泛取得詞彙之多重定義，如圖 D.6 所示，系統即會根據名詞詞彙延伸定義、各語句詞彙交集程度、各語句英文代表詞彙總數等數據，計算兩語句間之語意關聯程度，如圖 D.7 所示，之後權限內使用者可根據門檻值條件篩選欲去除之語句，並按下去除語意相似之語句鈕後，系統即會根據篩選門檻對應之語意關聯程度分佈情況，去除符合篩選條件門檻之語句，如圖 D.8 所示，接著，權限內使用者點擊取得候選填充句鈕後，系統即會將代表語句及重要語句彙集成擴充語句集合，並將擴充語句與訓練資料庫之語句進行相似性比對，同時將最高相似度之語句作為候選填充句，如圖 D.9 所示，之後權限內使用者點擊候選填充句之多重組合句鈕後，系統將透過候選填充句之關鍵詞彙（名詞、情緒詞、地方詞及形容詞）與訓練文章庫內所有語句進行比對，取得與關鍵詞彙完全相符之語句，並彙整成候選填充句之多重組合句，即完成文章多重組合句之建立，如圖 D.10 所示，其文章多重組合句建立結果將用於後續文章語句架構重組模組中。

➤ 程式運作流程：

舉例而言，系統由先前模組所取得之重要語句將目標文章「下雪天開車，可怕的經驗」、內容「小弟住的紐約天氣...等」中，五個重要語句「小弟住的紐約天氣..等」、「但這個禮拜下了一場雪」等利用中英翻譯器得知重要語句「但這個禮拜下了一場雪」之英文語句為「but;this;week;lower;one;stage;snowfall」，如圖 D.5 所示，接著，權限內使用者即可點選取得詞彙延伸定義集合鈕後，系統乃分析目標語句「但這個禮拜下了一場雪」之代表詞彙為「禮拜；Week」及「雪；Snowfall」，爾後，系統透過 WordNet 英文詞彙庫提供之詞義延伸定義，如圖 D.6 所示，取得目標語句英文代表詞彙之延伸定義集合為「hebdomad,workweek,snow」，之後權限內使用者即可點擊取得語句關聯程度鈕後，系統乃分析目標語句之英文詞彙延伸定義集合與其他語句「小弟住的紐約天氣還算穩定」之交集詞彙為「Snowfall」，詞彙於文章出現頻率為 3 次，且文章英文代表名詞詞彙總數為 21 後，系統即根據上述數據分析得知目標語句「但這個禮拜下了一場雪」與重要語句「小弟住的紐約天氣還算穩定」之關聯程度為「0.65」，如圖 D.7 所示，接著，權限內使用者選擇篩選門檻條件「最大值」並按下去除語意相近之語句鈕後，系統即會根據門檻條件「最大值」，將語意關聯程度「0.85」、語句內容為「但還是有餘雪結冰的狀況」予以去除，如圖 D.8 所示，之後權限內使用者按下取得候選填充句鈕後，系統將重要語

句「但這這個禮拜下了一場雪」與訓練文章庫所有語句進行相似性比對得知語句「但上個禮拜下了一場大雪」之相似度為「0.707」，該語句之相似度為所有語句中最高值，因此系統將判定該語句為重要語句「但這這個禮拜下了一場雪」之候選填充句，如圖 D.9 所示，待取得候選填充句後，權限內使用者可點選取得候選填充句之多重組合句鈕，系統乃依據候選填充句「小弟住的紐約天氣還算穩定」之關鍵詞彙「天氣、紐約、穩定」，與訓練文章庫比對相符之語句，將具有關鍵詞彙之語句「最主要的原因就是紐約當天的天氣不穩定」、「紐約天氣狀況不穩定」等五個組合句，彙集成候選填充句之多重組合句，如圖 D.10 所示，即完成目標文章標題為「下雪天開車，可怕的經驗」多重組合句之建立。

The screenshot shows the EDS (Electronic Decision Support Lab) interface. The main content area is titled "文章多重組合句建立 - 去除語意相近之語句" (Building Multiple Sentence Combinations for an Article - Removing Similar Sentences). It features a progress bar with three steps: "步驟一" (Step 1), "步驟二" (Step 2), and "步驟三" (Step 3). Below the progress bar, there are two tables and a "Chinese Reader" window.

目標文章資訊		
文章字數	文章作者	文章事件
406	henman	下雪
斷詞文章內容		
小弟住的紐約這禮拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住的附近要找車位停雖然下雪後路面有積雪, 但還是有餘雪結冰的狀況就在我找停車位與準備前面左轉時..我的車竟然打滑了...衝突忘了當時的時速..感覺可能有20 mph左右吧右腳用力的踩煞車, ABS傳回陣陣的回饋, VSA的燈狂亮, 但是車速並未減慢 無言		
重要語句資訊		
小弟住的紐約天氣還算穩定		
但這個禮拜下了一場雪		
雖然下雪後路面有積雪		
但還是有餘雪結冰的狀況		
如果要到下雪的地方開車		

重要語句資訊	
中文語句(中文)	英文語句
小弟住的紐約天氣還算穩定	I live; of; New York; Weather; entire; stable
但這個禮拜下了一場雪	but; this; week; lower; one; stage; snowfall
雖然下雪後路面有積雪	even though; to snow; escape route; side; have; shovel; cross
但還是有餘雪結冰的狀況	but; still; have; snowfall; freeze; of situation
如果要到下雪的地方開車	snow; of; region; drive car

The "Chinese Reader" window shows the word "鏟" (shovel) with its English translation: "to shovel; to remove; spade; shovel".

圖 D.5、取得重要語句之英文詞彙介面

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

EDS Electronic Decision Support Lab

- 協助論壇文章品質改善平台
- 社群文章維護模組
- 文章表達情緒判定模組
- 發作者文章改善模組
- 文章評閱分數判定
- 文章多重組合句建立
- 文章語句架構重組
- 情緒詞彙維護模組
- 系統參數設定模組

文章多重組合句建立 — 去除語意相近之語句

步驟一 步驟二 步驟三

目標文章資訊		取得英文詞彙 延伸定義	
文章作者	henman	文章事件	下雪
斷詞文章內容		英文詞彙延伸定義集合	
小弟住的紐約這各種拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住... Read More		中文語句	英文名詞詞彙
重要語句資訊		中文語句	延伸詞彙集合
中文語句	英文語句	小弟住的紐約天氣都算穩定	I, weather
但這個禮拜下了一場雪	but, this	但這個禮拜下了一場雪	week, snowfall
雖然下雪後路面有積雪	even the route, si	雖然下雪後路面有積雪	無詞彙
但還是有積雪結冰的狀況	but, still	但還是有積雪結冰的狀況	snowfall, situation
如果要到下雪的地方開車	snow, of	如果要到下雪的地方開車	region

WordNet 2.1 Browser

Search Word: weather

Searches for Weather: Noun Verb

The noun weather has 1 sense (first 1 from tagged texts)

1. (20) weather, weather condition, atmospheric condition -- (the

取得語句關聯程度 上一頁

圖 D.6、取得重要語句之英文延伸定義詞彙介面

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

EDS Electronic Decision Support Lab

- 協助論壇文章品質改善平台
- 社群文章維護模組
- 文章表達情緒判定模組
- 發作者文章改善模組
- 文章評閱分數判定
- 文章多重組合句建立
- 文章語句架構重組
- 情緒詞彙維護模組
- 系統參數設定模組

文章多重組合句建立 — 去除語意相近之語句

步驟一 步驟二 步驟三

目標文章各語句關聯程度			
文章資訊			
文章英文代表名詞總數(ASL)		文章事件	
21		下雪	
文章中文內容		文章英文內容	
小弟住的紐約這各種拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住的附近要找車位停雖然下雪後路面有積雪, 但還是有積雪結冰的狀況就在找停車位與準備前面左轉時...我的... Read More		I live in New York this week have some snow The road was wet and slippery, if you walk or drive will... Read More	
選擇計算之語句:		但這個禮拜下了一場雪	
選擇篩選門檻條件:		最大值	
語句內容	中文代表名詞	英文代表名詞	關聯程度
小弟住的紐約天氣都算穩定	小弟, 天氣	I, Weather	0.65
雖然下雪後路面有積雪	無詞彙	無詞彙	0.00
但還是有積雪結冰的狀況	雪, 狀況	snowfall, situation	0.845
如果要到下雪的地方開車	地方	region	0.00

去除語意相近句子 上一頁

圖 D.7、取得文章各語句之關聯程度介面

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

EDS Electronic Decision Support Lab

- 協助論壇文章品質改善平台
- 社群文章維護模組
- 文章表達情緒判定模組
- 發文者文章改善模組
- 文章評閱分數判定
- 文章多重組合句建立
- 文章語句架構重組
- 情緒詞彙維護模組
- 系統參數設定模組

文章多重組合句建立 — 去除語意相近之語句

步驟一 步驟二 步驟三

去除語意相近之語句				
文章中文內容		文章資訊		
文章英文內容				
小弟住的紐約這各種拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住的附近要找車位停雖然下雪後路面有縫過... 繼續閱讀		I live in New York this week have some snow The road was wet and slippery, if you walk or drive will... Read More		
目標語句:		但這個禮拜下了一場雪		
篩選門檻條件:		最大值		
語句內容	中文代表名詞	英文代表名詞	關聯程度 (底線組字為最高關聯)	是否予以去除
小弟住的紐約天氣都算穩定	小弟,天氣	I,Weather	0.65	<input type="radio"/> NO
雖然下雪後路面有縫過	無詞彙	無詞彙	0.00	<input type="radio"/> NO
但還是有餘雪結冰的狀況	雪,狀況	snowfall, situation	0.845	<input checked="" type="radio"/> YES
如果要到下雪的地方開車	地方	region	0.00	<input type="radio"/> NO

取得候選填充句 上一頁

圖 D.8、去除語意相似之語句介面

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

EDS Electronic Decision Support Lab

- 協助論壇文章品質改善平台
- 社群文章維護模組
- 文章表達情緒判定模組
- 發文者文章改善模組
- 文章評閱分數判定
- 文章多重組合句建立
- 文章語句架構重組
- 情緒詞彙維護模組
- 系統參數設定模組

文章多重組合句建立 — 取得重要語句與代表語句之候選填充句

步驟一 步驟二 步驟三

文章資訊		擴充語句之候選填充句		
文章內容		選擇目標擴充語句	小弟住的紐約天氣都算穩定	
文章擴充語句		候選填充句	相似值	是否採用
小弟住的紐約這各種拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住的附近要找車位停雖然下雪後路面有縫過, 但還是有餘雪結冰的狀況就在我找停車位與準備前面左轉時..我的... Read More		大雪不緊不慢地下了個禮拜	0.043	<input type="radio"/> NO
小弟弟住的紐約天氣都算穩定		但上禮拜下了一場大雪	0.707	<input checked="" type="radio"/> YES
雖然下雪後路面有縫過		新年的第一個禮拜就下大雪	0.042	<input type="radio"/> NO
但這個禮拜下了一場雪		生平第一次堆大雪人	0.00	<input type="radio"/> NO
如果要到下雪的地方開車		今年第一場雪好興奮啊!	0.032	<input type="radio"/> NO

取得候選填充句

取得候選填充句之多重組合句 上一頁

圖 D.9、取得代表及重要語句之候選填充句介面

The screenshot shows a web browser window with the URL localhost:8080/ForumEssay/new2/tree.html. The page title is '文章多重組合句建立 - 取得候選填充句之多重組合句'. The interface is divided into three steps: Step 1 (Article Information), Step 2 (Candidate Sentence Combinations), and Step 3 (Final Selection). The 'Article Information' table contains the following data:

文章資訊	
文章內容	
小弟住的紐約這禮拜下了些雪 路面又濕又滑, 不管走路或開車真的都要很小心話說下過雪後的一天, 在住的附近要找車位停雖然下雪後路面有積雪, 但還是有餘雪結冰的狀況就在我找停車位與準備前面左轉時..我的...Read More	
候選填充句	關鍵詞 (名詞、情緒、地方及形容詞)
小弟住的紐約天氣還算穩定	紐約、天氣、穩定
踏下雪後路面有積雪	路面
但這個禮拜下了一場大雪	禮拜
如果要到下雪的地方開車	地方

The 'Candidate Sentence Combinations' table contains the following data:

候選填充句之多重組合句	
選擇目標候選填充句	小弟住的紐約天氣都算穩定
關鍵詞	紐約、天氣、穩定
編號	組合句
第1組	最主要原因就是紐約當天的天氣不穩定
第2組	紐約天氣狀況不穩定
第3組	在紐約停留的期間如果天氣不穩定
第4組	紐約股市受到天氣影響呈現不穩定的浮動
第5組	紐約天氣不穩定 洋基1比6不敵雙城

圖 D.10、取得候選填充句之多重組合句

(B-3) 文章語句架構重組功能

➤ 程式運作邏輯：

待完成文章多重組合句建立後，權限內使用者即可執行文章語句架構重組功能，系統首先會依據語句之順序，將候選填充句之多重組合句與文章代表語句、重要語句進行替換，同時權限內使用者可選擇文章語句之組合，系統將根據語句替換之組合（即組合句替換後之文章為不重複之文章）呈現不同之文章語句架構於介面中，如圖 D.11 及圖 D.12 所示，之後，權限內使用者即可執行取得最高評閱分數之文章鈕後，系統乃將計算各文章架構之相似程度標準差，同時根據訓練文章取得語意差異性，最後，系統將兩數據結合後，即可得知各文章架構之評閱分數，如圖 D.13 所示，最後，權限內使用者即可點選顯示最高評閱分數之文章架構鈕後，系統將會顯示評閱分數最高之語句架構於畫面中以最為最中輸出結果，而本系統所替換之語句乃與原文章表達之情緒相符，故可得知評閱分數最高之文章所表達情緒將與原文章相符，如圖 D.14 所示。

➤ 程式運作流程：

舉例而言，當權限內使用者點選文章語句架構重組功能後，系統首先取得目標文章內容為「小弟住的紐約天氣都算穩定，但這個..等」之欲替換語句共四句，同時根據系統

於前一功能所取得之多重組合句，當中包含「但這個禮拜下了一場雪」有其他 5 組組合句、「如果要到下雪的地方開車」有其他 1 組組合句，得知文章語句共有「12」種不同之語句架構，當權限內使用者選擇第 1 組組合時，系統即會將「但這個禮拜下了一場雪」替換為「但上禮拜下了場大雪」、「如果要到下雪的地方開車」替換為「如果要再下雪天開車」並將新的文章架構呈現於介面中，如圖 D.11 所示，而第 2 組新的文章架構內容如圖 D.12 所示，之後，權限內使用者即可執行最後步驟，並點選取得最高評閱分數之文章鈕後，系統乃根據相異詞語數計算得知第 1 組文章架構相似程度標準差為「1138.01」、第 2 組為「1143.81」，並根據語意差異性、共用詞語數，計算得知所有語句組合中最高評閱分數為第 4 組「1183.62」，如圖 D.13 所示，最後，權限內使用者按下顯示最高評閱分數之文章架構鈕後，系統乃顯示第 4 組之語句架構，其語句架構乃為原文語句重組後之最終輸出結果，其中，語句架構乃替換「下雪不緊不慢下了一個禮拜」及「如果要在下雪天開車」兩語句，而替換之語句乃與情緒不相關，因此系統判定後得知新文章架構之情緒類別為「恐懼」，並與原文章之情緒類相符，如圖 D.14 所示。

文章語句架構重組
- 取得文章多重語句架構

步驟一 **步驟二**

目標文章資訊		
文章字數	文章作者	
406	henman	
原始文章內容		
小弟住的紐約天氣都算穩定，但這個禮拜下了一場雪，路面又濕又滑，不管走路或開車真的都要很小心，話說下過雪後的一天，在住的附近要找車位停，雖然下雪後路面有縫通，但還是有餘雪結冰的狀況，就在我找停車位與準備前面左轉時，我的車竟然打滑了... 衝刺，忘了當時的時速... 感覺可能有 20 mph 左右吧... 繼續閱讀		
語句順序	欲替換語句	可替換語句總數
1	小弟住的紐約天氣都算穩定	1
2	但這個禮拜下了一場雪	6
7	雖然下雪後路面有縫通	1
38	如果要到下雪的地方開車	2
取得評閱分數最高之文章 上一頁		

取得文章語句架構

新的文章語句架構	
文章語句架構組合數	12組
選擇組合	第1組
新文章語句架構(藍色語句為替換之語句)	
小弟住的紐約天氣都算穩定， 新年的第一個禮拜就下大雪 ，路面又濕又滑，不管走路或開車真的都要很小心，話說下過雪後的一天，在住的附近要找車位停，雖然下雪後路面有縫通，但還是有餘雪結冰的狀況，就在我找停車位與準備前面左轉時，我的車竟然打滑了... 衝刺，忘了當時的時速... 感覺可能有 20 mph 左右吧，右腳用力的踩煞車，ABS 陣陣的回轉，VSA 的燈亮，但是車速並未減慢無言，眼看前面就是 stop sign.. 兩旁有一兩位在縫雪的人選用怪異的眼光看我，就在我滑過 stop sign 之後，順勢左轉，車子才重新有煞車力道，就在我完成左轉後，本來直行的車也在我後面跟了上來，想當然就被按了喇叭，因為他那各方向是沒有 stop sign 的，當時並沒有特別驚嚇，不過事後想想，真的滿可怕的，還好當時在巷子裡，而且是單行道，運氣好，前面路口剛好沒車經過，不然小弟我開車十幾年從沒發生過的紀錄就要終止了(被撞的不算)，說不定也會因此而受傷，大家出門開車千萬要小心， 如果要在下雪天開車 ，更是千萬要注意，我還想回台中吃排骨麵...	

圖 D.11、取得文章多重語句架構-第 1 組

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

文章語句架構重組 - 取得文章多重語句架構

步驟一

目標文章資訊		
文章字數	文章作者	
406	henman	
原始文章內容		
<p>小弟住的紐約天氣都算穩定,但這個禮拜下了一場雪,路面又濕又滑,不管走路或開車真的都要很小心,話說下過雪後的一天,在住的附近要找車位停,雖然下雪後路面有積雪,但還是有餘雪結冰的狀況,就在我找停車位與準備前面左轉時,我的車竟然打滑了...嚇死,忘了當時的時速,感覺可能有20 mph左右吧...繼續閱讀</p>		
語句順序	欲替換語句	可替換語句總數
1	小弟住的紐約天氣都算穩定	1
2	但這個禮拜下了一場雪	6
7	雖然下雪後路面有積雪	1
38	如果要到下雪的地方開車	2
取得評閱分數最高之文章 上一頁		

取得文章語句架構

新的文章語句架構	
文章語句架構組合數	12組
選擇組合	第2組
新文章語句架構(藍色語句為替換之語句)	
<p>小弟住的紐約天氣都算穩定,但上禮拜下了一場大雪,路面又濕又滑,不管走路或開車真的都要很小心,話說下過雪後的一天,在住的附近要找車位停,雖然下雪後路面有積雪,但還是有餘雪結冰的狀況,就在我找停車位與準備前面左轉時,我的車竟然打滑了...嚇死,忘了當時的時速,感覺可能有20 mph左右吧,右腳用力的踩煞車,ABS傳回陣陣的回饋,VSA的燈亮,但是車速並未減慢無言,眼看前面就是stop sign,兩旁有一兩位在縫簷的人選用怪異的眼光看我,就在我滑過stop sign之後,順勢左轉,車子才重新有煞車的力道,就在我完成左轉後,本來直行的車也在我後面跟了上來,想當然爾被按了喇叭,因為他那各方向是沒有stop sign的,當時並沒有特別驚嚇,不過事後想想,真的滿可怕的,還好當時在巷子裡,而且是單行道,運氣好,前面路口剛好沒車經過,不然小弟我開車十幾年從沒肇事過的紀錄就要終止了(被撞的不算),說不定也會因此而受傷,大家出門開車千萬要小心,如果要在下雪天開車,更是千萬要注意,我還想回台中吃排骨麵啊...</p>	

圖 D.12、取得文章多重語句架構-第 2 組

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

文章語句架構重組 - 取得最高評閱分數之文章架構

步驟一

目標文章資訊		
文章標題	下雪天開車, 可怕的經驗	
文章字數	相異詞語數	情緒類別
406	259	恐懼 😨
文章內容		
<p>小弟住的紐約天氣都算穩定,但這個禮拜下了一場雪,路面又濕又滑,不管走路或開車真的都要很小心,話說下過雪後的一天,在住的附近要找車位停,雖然下雪後路面有積雪,但還是有餘雪結冰的狀況,就在我找停車位與準備前面左轉時,我的車竟然打滑了...嚇死,忘了當時的時速,感覺可能有20 mph左右吧,右腳用力的踩煞車,ABS傳回陣陣的回饋,VSA的燈亮,但是車速並未減慢無言,眼看前面就是stop sign,兩旁有一兩位在縫簷的人選用怪異的眼光看我,就在我滑過stop sign之後,順勢左轉,車子才重新有煞車的力道,就在我完成左轉後,本來直行的車也在我後面跟了上來,想當然爾被...繼續閱讀</p>		
顯示最高評閱分數之文章架構 上一頁		

各語句組合**目前系統不含目標文章之訓練文章數為3份

評分結果

所有語句架構評閱分數				
組合數	12組			
組別	相似標準差	評閱分數	是否為最高評分	
1	13.49	1138.01	NO	NO
2	15.87	1143.81	NO	NO
3	14.42	1062.72	NO	NO
4	13.72	1183.62	YES	YES
5	15.22	1081.3	NO	NO
6	14.82	998.65	NO	NO
7	14.6	1043.75	NO	NO
8	14.11	1045.2	NO	NO
9	15.65	989.81	NO	NO
10	13.72	1105.21	NO	NO
11	15.65	1008.72	NO	NO
12	16.05	1033.56	NO	NO

圖 D.13、取得評閱分數最高之文章組合

協助論壇文章品質改善平台

localhost:8080/ForumEssay/new2/tree.html

Electronic Decision Support Lab

- 協助論壇文章品質改善平台
- 社群文章維護模組
- 文章表達情緒判定模組
- 發文者文章改善模組
- 文章評閱分數判定
- 文章多重組合句建立
- 文章語句架構重組**
- 情緒詞彙維護模組
- 系統參數設定模組

文章語句架構重組

—取得最高評閱分數之文章架構

步驟一 步驟二

目標文章資訊			
文章字數	相異詞語數	情緒類別	文章內容
406	259	恐懼	小弟住的紐約天氣都算穩定,但這個禮拜下了一場雪,路面又濕又滑,不管走路或開車真的都要很小心,話說... 繼續閱讀

最高評分之文章內容

目標文章之新文章架構資訊及內容			
組別	評閱分數	情緒類別	是否與原文情緒一致
4	1183.62	恐懼	YES

小弟住的紐約天氣都算穩定,但大雪不緊不慢的下了一個禮拜,路面又濕又滑,不管走路或開車真的都要很小心,話說下過雪後的一天,在住的附近要找車位停,雖然下雪後路面有積雪,但還是有絲毫結冰的狀況,就在我找停車位與準備前面左轉時..我的車竟然打滑了...衝突,忘了當時的時速..感覺可能有20 mph左右吧,右腳用力的踩煞車,ABS傳回陣陣的回線,VSA的燈狂亮,但是車速並未減慢無言,眼看前面就是stop sign,兩旁有一兩位在積雪的人還用怪異的眼光看我,就在我滑過stop sign之後,順勢左轉,車子才重新有煞車的力道,就在我完成左轉後,本來直行的車也在我後面跟了上來,想當然爾被按了喇叭,因為他那各方向是沒有stop sign的,當時並沒有特別驚嚇,不過事後想想,真的滿可怕的,還好當時在巷子裡,而且是單行道,運氣好,前面路口剛好沒車經過,不然小弟我開車十幾年從沒肇事過的紀錄就要終止了(被撞的不算),說不定也會因此而受傷,大家出門開車千萬要小心,如果要在下雪天開車,更是千萬要注意,我還想回台中吃排骨麵啊..

圖 D.14、評閱分數最高之文章資訊與內容

附錄(2)、模組步驟運算與應用說明

本研究所發展之以論壇文章蘊含情緒為基之文章重組方法論乃分為「文章表達情緒判定模組」與「發文者文章語句重組模組」，其流程圖與對應步驟如圖 E.1 與圖 E.5 所示，當中第一模組乃以標題「台灣與 FBI 犯罪心理官」、「唉~無奈」與「女友吵架」等實際文章作為計算說明（如圖 E.2 至圖 E.4 所示）；於第二模組中以標題「台灣這些球員...」（如圖 E.5 所示）、「下雪天開車...」（如圖 E.6 所示）、「唉~無奈」與「女友吵架」等實際文章內容做為計算說明，以下乃針對各模組之步驟以實際文章內容作為運算說明：

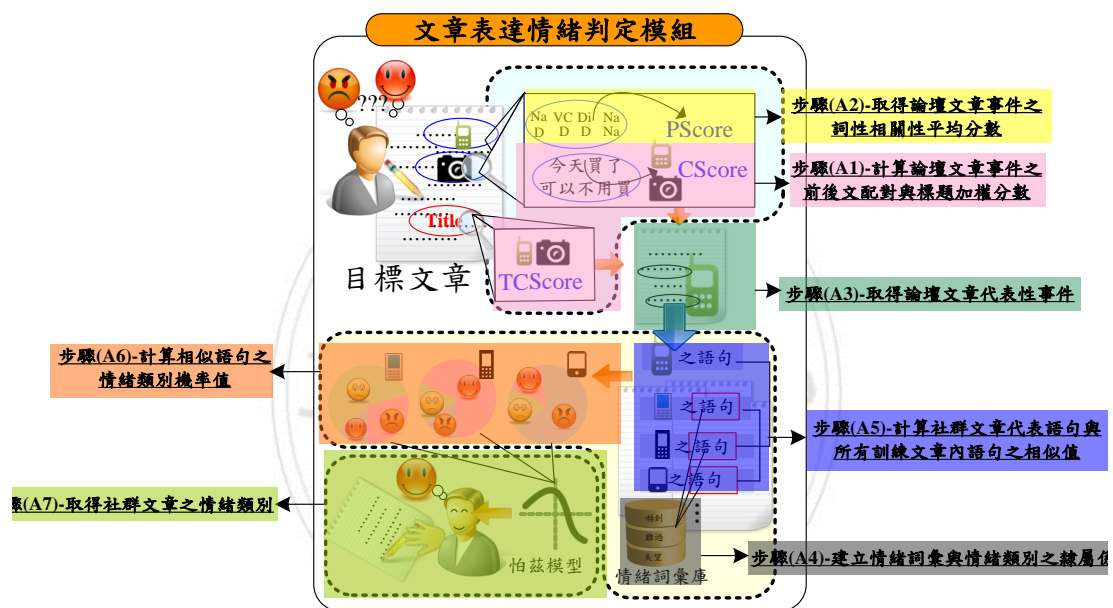


圖 E.1、文章表達情緒判定模組之流程步驟

步驟(A1)—計算論壇文章事件之前後文配對及標題加權分數

由圖 E.2 之論壇實際文章作為計算說明，詳細內容如表 E.1 所示，該論壇文章之候選事件分別為「治安史」及「犯罪」，其中候選事件「治安史」於文章中出現 1 次，且該事件前一個前文語句之詞彙為「如果說」及「有」，後一完整語句之詞彙分別為「他」、「只是」、「一」、「件」、「綁架案」、「的」、「延伸」，其中詞彙「有」於所有候選事件前一個前文語句分佈比例為 0.22，且該詞彙於論壇文章內之分佈比例為 4，故將各詞彙之兩個分佈比例值相加並取得平均值，即可取得候選事件「治安史」前一個前文語句之平均分佈比例值，此外，候選事件後一個後文語句之計算方式亦計算後文語句之詞彙平均分佈比例值，之後將前後文語句之平均分佈比例值相加，即為候選事件「治安史」之前後文配對平均分數。此外，文章之候選事件「犯罪」於標題中出現 1 次，故可得知候選

事件「犯罪」加權分數為 0.5。



圖 E.2、標題「台灣與 FBI 犯罪心理官」之實際文章

表 E.1、「治安史」事件之論壇文章資訊

文章標題	台灣 FBI 犯罪心理分析官
文章內容	<p>這個頭銜或者說職業很特別吧！跟家政婦這個職業一樣，對台灣來講都很陌生，也不存在的行業，但這 2 個職業在日本都有，職業跟需求成正比，可喜可賀也就是說台灣變態殺人狂很罕見，如果說有，陳進興算是台灣「治安史」上唯一令一般人坐立難安的殺人犯其實他真的不算是隨機不特定殺人，他只是一件綁架案的延伸，也就是失敗逃亡，為了自保才有後續的犯案.....「犯罪」心理分析～是針對連續殺人，而且被害人有特定連結但不認識的陌生人，足以影響一般民眾的恐慌，所以這類人一定要趕快抓到，不然會不斷有受害人增加，台灣的殺人犯都是有特定目標的行為，基於衝動或預謀，一般安份守己的老百姓不會成為受害者或在生活中形成一種恐懼，如：殺父母兄弟姐妹，殺自己小孩，殺情人配偶朋友，加害人和受害人有特定關係，有時不小心只想殺這些特定人卻波及無辜，如：縱火等。我們台灣社會整體還算健全，目前培養不出連續殺人魔，但健全家庭學校社會功能仍然刻不容緩，不然台灣有一天也會誕生像美國，日本這樣的連續殺人魔，不可不慎啊！</p>
文章字數	427

步驟(A2)–取得論壇文章事件之詞性相關性平均分數

本步驟乃以表 E.1 論壇文章之候選事件「治安史」作為計算說明，首先將詞性距離參數值 $S_{dis}(POS.)$ 設定為 6，並將候選事件「治安史」及「犯罪」之前 6 個詞彙集合於 Set(VEF) 中，其中治安史前 6 個詞彙分別為「罕見」、「如果說」、「有」、「陳進興」、「算

是」、「台灣」，而6個詞彙於詞彙集合 Set(VEF)中分別出現1、1、2、1、1及1次，於論壇文章中分別出現1、1、2、9、2及6，之後利用計算「治安史」事件之詞性相關性平均分數即為0.314。

步驟(A3)—取得論壇文章代表性事件

藉由步驟(A1)與步驟(A2)可得知，候選事件「治安史」之前後文配對分數為「0.23」，標題加權分數為「0.5」，詞性相關性配對分數為「0.314」，各權重值皆設為1時，全部相加後即可得知候選事件「治安史」之代表分數為「1.044」。

步驟(A4)—建立情緒詞彙與情緒類別之隸屬係數

完成前三步驟後，本步驟以「無奈」之情緒詞彙為計算範例，並以圖E.2與圖E.3作為實際訓練文章，當中詳細內容整理如表E.2所示，首先擷取表E.2中論壇文章含有情緒詞彙之語句形成情緒語句集合（如表E.2粗色字體之語句），例如「害怕」情緒類別之情緒語句「這種關心對我來說已經算是一種壓力了」，「難過」情緒類別之情緒語句「我無奈到心灰意冷！」等，接著統計詞彙「無奈」於「害怕」情緒類別之情緒語句集合中出現1次，而於「難過」情緒類別之情緒語句集合中出現2次，且「害怕」情緒類別之情緒詞彙於文章出現總數為3次，「難過」情緒類別總數為3次，因此可得知「無奈」情緒詞彙與「害怕」及「難過」情緒類別之關係係數分別為0.333及1.056，並予以正規化後即可得知「無奈」情緒詞彙與「害怕」及「難過」之隸屬係數分別為「0.24」及「0.76」。



圖 E.3、標題「唉~無奈」之實際文章



圖 E.4、標題「女友吵架...」之實際文章



表 E.2、訓練文章資訊

文章標題	文章內容	文章字數	文章作者
唉~無奈	<p>最近親人都在問說有沒有存款啊?好準備買房子,我的回答都是:『有啊~正在存頭期款啊!看能不能在三十歲前買一間房。』長輩聽了很欣慰,覺得我有在規劃,放心許多~長輩隨後提起一位親戚在 30 歲前就買房(約民 85 年)大大稱讚~回答完這問題的我,心想其實對於購屋,根本都沒頭緒、毫無規劃,目前連助學貸款都還沒還完,買房,對我來說真的還好遙遠~我也想買房子啊~現在我都好怕長輩問但無奈就是財力不足,這種關心對我來說已經算是一種壓力了...年紀稍長的網友們,如有關心你妳晚輩此類問題,麻煩多少也體諒一下年輕人,尤其一比較,壓力還不小...</p>	414	22K 新世代
女友吵架常說不高興那就不要!分手阿!她沒差!這樣的女友你能接受嗎?	<p>每次不順我女友的意或吵架,他就說不高興那就不要!分手阿!他沒差,之前我有買智慧型手機辦續約生日送他,我有買一個配件包,但不是原廠電池,她為了不是一顆原廠電池跟我翻臉,還說他習慣用原廠電池,不然手機他不要要退還給我,還說續約辦沒誠意送他,省東省西小氣!後來我打電話跟他道歉我會改進才罷休.又有一次我買吃的東西給他,冰在我家冰箱裡後來被我姊小朋友吃掉,我告訴他後又跟我翻臉,說我不尊重他,沒誠意,我說都吃啦怎麼辦,不要跟小朋友計較,不然要去罵小孩子嗎?我再買給你就好啦!還說我說他跟小朋友計較,又說我們觀念不同不適合,不高興那不要阿 分手阿 又跟他道歉哄他 因為我很愛她!後來我有跟他溝通不要輕言分手 這樣你有在乎這段感情嗎?後來次數比較少一點.有一次我們聊天我問她 我做生意沒有 100%之百成功 萬一我生意失敗怎麼辦 他說他會離開我 他很怕吃苦 我無言!而且每次吵架都是我先低頭認錯 請她原諒 以下他要求沒給絕對翻臉 過年要給紅包 情人節 聖誕節 生日 要送禮物 他說我送他的東西他都不需要 沒品味,要折現 我也同意避免不必要的爭執,後來又說沒照顧他生活,最近要我給她生活費!(我想跟他說我們還沒結婚耶)出去外面吃喝玩的都是我付錢耶!(裝假牙也是我幫他付錢喔) 每星期都要求去餐廳吃飯他還說以後他不想上班,還說有能力的男人不會給老婆上班,我說我有一天能力許可這沒問題,花別人的錢臉不紅氣不喘!有一次吃飯我錢帶不夠 叫他先墊一下一上車馬上跟我要 怕我不給他?這件事讓我無奈到心灰意冷! 他從來不送我東西,還嫌我送的 COACH 手錶爛,說他自己買的手表都比這個貴,有好幾次去逛街他看衣服問我意見,我說還好!她說我怕付錢才說還好,我說你要跟我說我會付錢,又不理我!說我都不照顧他生活(之前帶她出國玩過好幾次).最近他說結婚後,房子要買他名下,我說 OK! 又說錢全部他要管,還說如果我有外遇全部財產都要給他,我問他說那你有外遇呢?馬上翻臉,說我沒誠意不夠爽快!每次吵架 我在開車的時候他都會一直把車門開又關要衝下車,很危險!下車時車門甩很大聲,好幾次都把我的車都弄得有異音還要花錢去修理,這次他下車時候叫他也不理我.等</p>	1137	sirichai

步驟(A5)—計算論壇文章代表語句與所有訓練文章內語句之相似值

本步驟以表 E.2 文章作為計算說明，當中語句「現在我都好怕長輩問但無奈就是財力不足」為目標語句（以下代稱為 ADS1），該語句斷詞後詞彙為「現在」、「我」、「都」、「好」、「怕」、「長輩」、「但」、「無奈」、「就是」、「財力」、「不足」，並計算相似語句「我無奈到心灰意冷!」，斷詞後詞彙為「這件」、「事」、「讓」、「我」、「無奈」、「到」、「心灰意冷」，接著統計目標語句與相似語句中所有詞彙出現頻率，如表 E.3 所示，並計算目標語句與相似語句相似值之分子約為 8.775、分母為 2，故相除後即可取得相似度約為 0.228，該相似值愈趨近於 1 即表示與目標語句愈相似。

表 E.3、目標語句與相似語句之詞彙頻率表

斷詞詞彙 語句	現在	我	都	好	怕	長輩	讓	但	無奈	這件	事	就是	財力	不足	到	心灰 意冷
目標語句	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0
相似語句	0	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1

步驟(A6)—計算相似語句之情緒類別機率值

接著以表 E.2 中目標語句「現在我都好怕長輩問但無奈就是財力不足」作為計算範例，其中極性項目乃由相似語句「這件事讓我無奈到心灰意冷!」之重要詞彙「這件事」及情緒詞彙「無奈」所組成，且藉由「無奈」情緒詞彙得知該極性項目與「害怕」及「難過」之情緒類別隸屬係數分別為「0.24」及「0.76」，但相似語句中無轉折詞存在，因此指標函數 $I[F_i(ADS_q, S_u), TW]$ 為 1，最後計算相似語句對於「害怕」及「難過」情緒類別機率值之分子分別約為 2.869 與 3.219 而分母約為 6.088，並取得兩情緒類別之機率值分別約為「0.471」及「0.529」。

步驟(A7)—取得論壇文章之情緒類別

最後，以目標語句「現在我都好怕長輩問但無奈就是財力不足」為例，其中相似語句「讓我無奈到心灰意冷!」對於「害怕」及「難過」情緒類別機率值分別為「0.471」及「0.529」，透過公式(3.12)即可得知目標語句與「害怕」及「難過」情緒類別之穩定值約為「-0.424」與「-0.503」，由於「害怕」情緒類別之穩定值「-0.424」較趨近於 0，故目標語句極可能偏向「害怕」之情緒類別。

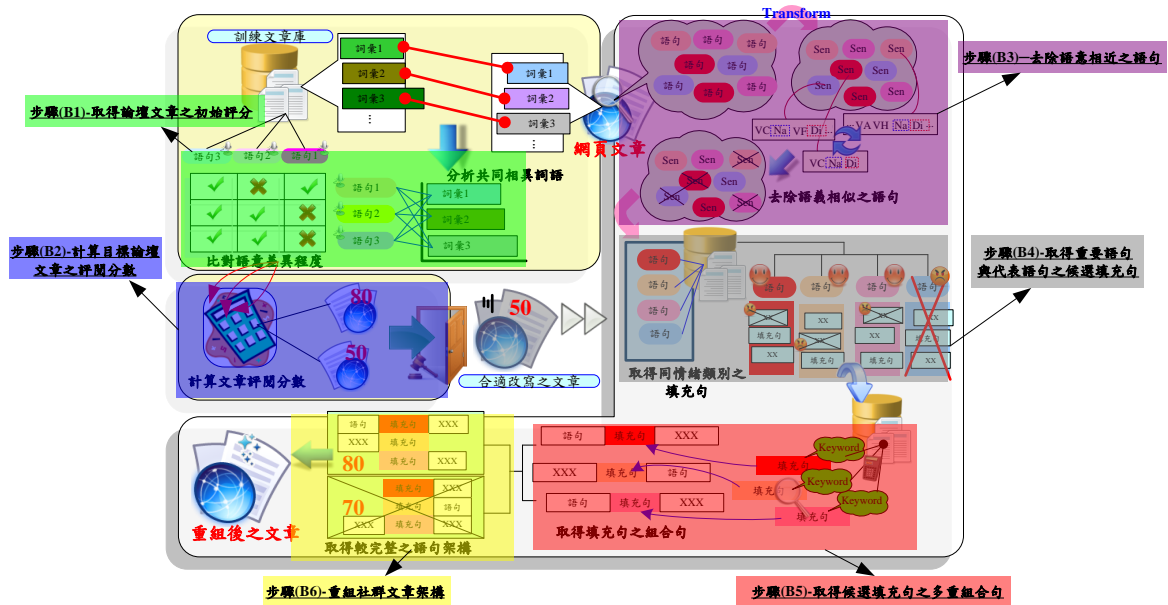


圖 E.5、發文者文章語句重組模組之流程步驟

步驟(B1)-取得論壇文章之初始評分

於此步驟中以圖E.6之實際論壇文章「台灣這些球員…」為例，如詞彙「台啤」與「台銀」分別於文章中重複出現6次與5次，但本步驟將11個詞彙視為2種詞語數，因此該文章篩選後之相異詞語數為172個。



圖 E.6、標題「台灣這些球員…」之實際文章

表 E.4、論壇之文章資訊

文章標題	台灣這些球員還嫌看球的觀眾太多嗎?又打架了
文章內容	2012 觀護盃籃球賽是第 10 季 SBL 熱身賽，雖然戰績與例行賽無關，各隊仍非常重視，昨天台啤與台銀在 5 到 8 名排名賽就打來激烈，甚至引爆衝突，共 12 名球員遭驅逐出場。第 4 節 7 分 57 秒，台銀老將簡明富進攻時推開防守的台啤小將李秉鴻，遭判進攻犯規，倒地的李秉鴻起身後揮拳並嗆聲，引來其他台銀球員不滿，隊長許致強保護隊友心切，一個箭步上前大喊：「你現在怎樣，有沒有倫理啊，沒大沒小。」接下來雙方都有球員衝進球場，台啤甚至板凳清空。引爆衝突的兩名球員，都互指對方動手，簡明富承認自己進攻犯規，但李秉鴻手腳也不乾淨；而李秉鴻則指著右額上的傷痕強調，「我知道要尊重他，但他把我弄到流血，我實在忍不住。」簡明富與李秉鴻起衝突後，台啤周資華也衝向台銀板凳席，企圖找未登錄的王男桂理論，被桑一慶與林瑞瑩擋下，才沒有引爆大規模肢體衝突。最終裁判將引發衝突的兩人及所有應在板凳席上、卻衝進場的球員全部吹判奪權出場，台銀有 4 人，台啤則因板凳清空，共 8 人遭判出場。台啤最終只剩 4 人出賽，執行教練劉嘉發不滿裁判有意沒收比賽，也遭判奪權出場。又看到周資華，連教練也被趕出場，真是夠了，出國打有這麼會打就好了。
文章作者	DSYMSN
文章字數	479
相異詞語數	172

步驟(B2)—計算目標論壇文章之評閱分數

接著，以表 E.4 論壇文章作為目標文章之計算範例，該目標論壇文章與訓練文章(如表 E.5 所示，實際文章如圖 E.3 與圖 E.4 所示)之相似度分別為 29 與 49，首先計算論壇文章與訓練文章相似程度標準差為 10，並計算目標論壇文章與訓練文章之語意差異程度分別為-12.7 與 12.7，最後即可取得目標論壇文章評閱分數為 264。

表 E.5、訓練文章資訊

文章標題	文章內容	文章字數	文章作者	相異詞語數
女友吵架常說不高興那就不要!分手阿!她沒差!這樣的女友你能接受嗎?	，他就說不高興那就不要!分手阿!他沒差，之前我有買智慧型手機辦續約生日送他，我有買一個配件包，但不是原廠電池，她為了不是一顆原廠電池跟我翻臉，還說他習慣用原廠電池，不然手機他不要要退還給我，還說續約辦沒誠意送他，省東省西小氣!後來我打電話跟他道歉我會改進才罷休.又有一次我買吃的東西給他，冰在我家冰箱裡 後來被我姊小朋友吃掉，我告訴他後又跟我翻臉，說我不尊重他，沒誠意，我說都吃啦怎麼辦..等	1137	sirichai	395

表 E.5、訓練文章資訊(續)

文章標題	文章內容	文章字數	文章作者	相異詞語數
唉~無奈	最近親人都在問說有沒有存款啊?好準備買房子,我的回答都是:『有啊~正在存頭期款啊!看能不能在三十歲前買一間房。』長輩聽了很欣慰,覺得我有在規劃,放心許多~長輩隨後提起一位親戚在30歲前就買房(約民85年)大大稱讚~回答完這問題的我,心想其實對於購屋,根本都沒頭緒、毫無規劃,目前連助學貸款都還沒還完,買房,對我來說真的還..等	414	22K 新世代	141

步驟(B3)—去除語意相近之語句

本步驟乃以圖E.7之論壇實際文章(彙整如表E.6所示)資訊為例,兩語句「小弟住的紐約天氣都算穩定」與「但這個禮拜下了一場雪」中「天氣」與「雪」為一名詞,其英文名稱為「weather」與「snow」,並利用WordNet取得詞彙「weather」之延伸詞彙包含「snowfall、weather condition、atmospheric condition、physical phenomenon..等」,並放入延伸定義詞彙集合ASLE_Set_m中,而詞彙「snow」則可取得「snowfall、weather condition、physical process」等延伸詞彙,亦放入延伸定義詞彙集合ASLE_Set_m中,將兩集合交集後可得知兩詞彙之共同延伸定義詞彙集合CL_Set內包含「snowfall」與「weather condition」詞彙,其中,論壇文章各語句之英文代表名詞詞彙總數ASL_i為21個(如表E.6標註底線之詞彙),「snowfall」下雪一詞於論壇文章內出現頻率N(CL_Set[E L_i])為3次(如表E.6黑底詞彙),但「weather condition」天氣狀況一詞於論壇文章中並無出現,故將忽略計算該詞彙,因此計算兩英文名詞之關聯程度為「0.845」,由於兩語句其他詞性之詞彙關聯程度皆為0,故計算後兩語句之關聯程度為0.845,此外,因語句「小弟住的紐約天氣都算穩定」詞性數為7個,「但這個禮拜下了一場雪」語句詞性數為8個,故於本步驟乃去除詞性種類數較少之語句「小弟住的紐約天氣都算穩定」。



圖 E.7、標題「下雪天開車，可怕的經驗」之實際文章

表 E.6、論壇文章資訊

文章標題	下雪天開車，可怕的經驗
文章內容	小弟住的紐約天氣都算穩定，但這個禮拜下了一場雪，路面又濕又滑，不管走路或開車真的都要很小心。話說下過雪後的一天，在住的附近要找車位停，雖然下雪後路面有鏟過，但還是有餘雪結冰的狀況，就在我找停車位與準備前面左轉時..我的車竟然打滑了...忘了當時的時速..感覺可能有 20 mph 左右吧，右腳用力的踩煞車，ABS 傳回陣陣的回饋，VSA 的燈狂亮，但是車速並未減慢，眼看前面就是 stop sign..兩旁有一兩位在鏟雪的人還用怪異的眼光看我，就在我滑過 stop sign 之後，順勢左轉，車子才重新有煞車的力道，就在我完成左轉後，本來直行的車也在我後面跟了上來，想當然爾被按了喇叭，因為他那各方向是沒有 stop sign 的，當時並沒有特別驚嚇，不過事後想想，真的滿可怕的，還好當時在巷子裡，而且是單行道，運氣好，前面路口剛好沒車經過，不然小弟我開車十幾年從沒肇事過的紀錄就要終止了(被撞的不算)。說不定也會因此而受傷，大家出門開車千萬要小心，如果要到下雪的地方開車，更是千萬要注意，我還想回台中吃排骨麵啊..
文章字數	406
文章作者	henman

步驟(B4)—取得重要語句與代表語句之候選填充句

之後，表 E.6 論壇文章之情緒類別為「害怕」，首先將該論壇文章之重要語句「但這個禮拜下了一場雪」、「我的車竟然打滑了」等，以及代表語句「下雪天開車」、「雖然下雪後路面有鏟過」等彙集成擴充語句集合 Ex_Set，並尋找與語句「但這個禮拜下了一場雪」相似程度最高之語句為「但上禮拜下了一場大雪」(相似度約 0.707)，而前後語句分別為「原本行程是要出去玩，」及「，行程只能延期」，兩語句中皆無情緒詞彙，因此將語句「原本行程是要出去玩」及「行程只能延期」放置候選填充句集合 Con_Set 中。

步驟(B5)–取得候選填充句之多重組合句

接著，表 E.6 論壇文章之候選填充句「小弟住的紐約天氣都算穩定」之關鍵詞有「紐約」、「天氣」及「穩定」，於訓練文章庫中句有此些關鍵詞之語句為「最主要的原因就是紐約當天的天氣不穩定」，藉由彙集填充詞後即可得知候選填充句之組合句為「紐約當天的天氣不穩定」。

步驟(B6)–重組論壇文章架構

根據前一步驟所取得候選填充句之多重組合句，本步驟乃依據語句之順序，將目標論壇文章候選填充句之多重組合句 $CS_MUS_{q,z}$ 、代表語句及重要語句進行組合，因候選填充句具有多個組合句，故重組後之論壇文章將有數個不同之語句架構，為使從中取得架構較完整之論壇文章，本步驟乃將重組後之論壇文章藉由步驟(B1)與步驟(B2)取得評閱分數 Rec_Grade ，並由評閱分數最高之語句架構作為目標社群文章改寫後之內容。

