

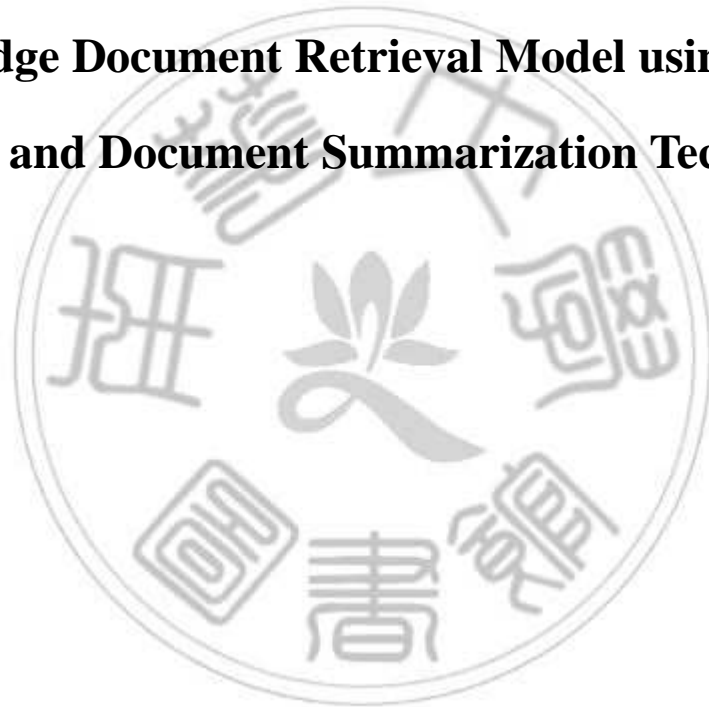
南 華 大 學

資訊管理學系

碩士論文

結合語意解析與文件摘要技術之知識文件檢索模式

**A Knowledge Document Retrieval Model using Semantic  
Analysis and Document Summarization Technologies**



研 究 生：龔鈺婷

指 導 教 授：楊士霆

中華民國 101 年 6 月 26 日

# 南 華 大 學

資訊管理學系

碩 士 學 位 論 文

結合語意解析與文件摘要技術之知識文件檢索模式

研究生： 龍金婷

經考試合格特此證明

口試委員： 楊七建  
李健良  
洪銘建

指導教授： 楊七建

系主任(所長)： 吳光岡

口試日期：中華民國 101 年 06 月 26 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： 龔鈺婷 之碩士畢業論文

中文題目：

結合語意解析與文件摘要技術之知識文件檢索模式

英文題目：

A Knowledge Document Retrieval Model using Semantic Analysis and Document Summarization Technologies

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：龔鈺婷 (請親自簽名)

指導老師：楊士霆 (請親自簽名)

中華民國 101 年 05 月 30 日

南華大學碩士班研究生

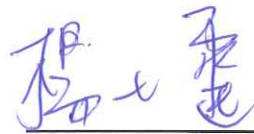
論文指導教授推薦函

資訊管理系碩士班龔鈺婷君所提之論文  
結合語意解析與文件摘要技術之知識文件檢索  
模式

---

係由本人指導撰述，同意提付審查。

指導教授



---

10 | 年 6 月 26 日

## 誌謝

在這兩年的求學過程中，無論是課業上、生活上或是精神上受到許多人的照顧與關心，也因為有你們的加油與支持讓我能順利完成論文，在此感謝你們。

首先，最要感謝我的指導教授楊士霆博士，感謝老師不厭其煩的提醒與指導，使我更謹慎、細心並具有規劃性的處理人事務，同時亦培養我獨立思考與創新的能力。感謝老師在我犯錯的時候指責、在我成長的時候鼓勵、在我遇到困境的時候一起想辦法面對。謝謝老師這兩年的指導，謝謝老師給我的自信與肯定，讓我有個特別又充實的碩士人生，在此致上由衷的感謝與敬意。

在學位考試期間，感謝清華大學的侯建良教授與南華大學的洪銘建教授，謝謝老師們用心審視學生的論文，並提供諸多寶貴建議與指正，使本論文內容更加完善。接著，我要感謝李翔詣老師在我求學期間給予的關心與協助，並提供寶貴的職場經驗與建議，使我提早了解職場環境，此外，我亦謝謝校內的老師們與伊汝姐的照顧與關心。

再來我要感謝 RFID 團隊的學姐與學長們，以及已畢業的士哲學長、瀨文學姐與世章學長，雖然學長姐們已就業或是正在服兵役，但仍在繁忙中關心我的狀況並給予鼓勵，讓我備感溫馨。謝謝我的好夥伴小胖以及在各地努力生活的大學朋友們，謝謝你們在畢業後沒忘了我，並且適時的關心我。謝謝研究室 208 與 204 的各位夥伴，無論是同學或是學弟妹，在這段日子謝謝你們的照顧與外送晚餐，大家庭的感覺讓我不孤單。感謝大學部的學弟妹們，謝謝有你們陪伴，使我歡樂不斷。最要感謝的是，EDS 團隊的胖子學弟們小紅、小偉與阿昌學弟，謝謝你們陪我熬夜、一起睡研究室、一起搶沙發、一起互相扶持、一起吃飯的日子，更讓我忘不了每次比賽大家一起緊張、互相幫忙的日子，再次謝謝你們的分享與一路的陪伴！

最後，我要感謝在高雄等我回去的朋友以及我的家人們，感謝朋友的支持、家人的包容與栽培，使我能專心並順利完成碩士班的學業。再次感謝所有人，並在此分享我的喜悅！

## 摘要

近年來，人們已習慣以網路進行資訊與知識之獲取，卻容易造成資訊過量等問題產生，因此延伸出關鍵字搜尋、文件分類等相關研究與技術並協助知識搜尋者能快速且有效取得資訊。雖有資訊搜尋窗口以縮小搜尋範圍，然而面對特定領域網站時，若無相關領域背景之知識搜尋者仍需不斷嘗試並取得回饋，關鍵字搜尋與文件分類乃缺乏有效地協助。因此，本研究乃以「勞工安全衛生知識網」為基礎，針對特定領域知識文件發展一套「結合語意解析與文件摘要技術之知識文件檢索」模式，以協助知識搜尋者有效地篩選資料與知識。

本研究所建立「結合語意解析與文件摘要技術之知識文件檢索」模式乃包含「知識文件表達項目解析模組」、「知識文件問答解析模組」與「知識文件結構化摘要推論模組」。首先，「知識文件表達項目解析模組」乃針對「勞工安全衛生知識網」之人因工程工作場所研究報告與技術叢書進行文件表達項目解析，以擷取此類型知識文件之表達方式、呈現內容以及相關性語彙，進而建構知識語彙庫並取得文件之觀念性語句，以作為知識文件問答解析之關鍵字語意比對與結構化摘要推論之依據。其次，「問答解析模組」乃進行文件關鍵字詞與搜尋字串語意相似性分析，透過語意相似性進而取得確實文件，最後，「結構化摘要模組」乃計算並擷取具有文件之代表性語句，並以制式摘要方式呈現於搜尋者，以避免個人閱讀偏好影響文件之篩選，進而加強勞工安全衛生知識網搜尋技術以提高知識網知識分享之成效。本研究除發展模式與方法論外，並依此方法論建構一套「結合語意解析與文件摘要技術之知識文件檢索」系統以進行案例驗證，從而確認方法論與技術之可行性。

整體而言，藉由知識語彙庫建立並透過觀念性語句擷取，並整合關鍵字索引與結構化摘要，以提供更方便、有效地方式解除知識搜尋者於專業領域中搜尋瓶頸，並能藉以整合技術更容易取得專業知識。

**關鍵詞：**勞工安全衛生知識網、知識管理、資料探勘、文件摘要技術、語意分析

## Abstract

It is a common practice to acquire information and knowledge from the Internet; thus, keyword searching, document classification and other technologies have been developed to facilitate searching. Although the search engine sites can narrow down the scope of search, knowledge demanders without background knowledge in the specific fields need to continuously search and receive feedbacks. Hence, this paper develops a Knowledge Document Retrieval model using semantic analysis and document summarization technologies for domain knowledge documents. First, this paper analyzes the ergonomic technology reports from the website of “Institute of Occupational Safety and Health” to capture the expressions and related vocabulary of domain knowledge documents to develop the knowledge vocabulary database. Second, through the Question and Answer Analysis (QAA) module, the correlations between proper names and query strings can be obtained. Third, based on Conceptual Vocabulary Determination (CVD) module, the most conceptual or representative sentences of domain documents can be derived and serve as candidate sentences for structured summarization. Finally, the Document Structured Summarization (DSS) module is used to calculate and retrieve representative sentences of the documents and integrate them into summary for knowledge demanders. It is expected that knowledge demanders can directly read the desired parts according to problems to ensure they can find document they want within a short time. In order to demonstrate applicability of the proposed methodology, a web-based knowledge document retrieval system is also established based on the proposed model. Furthermore, the knowledge documents (i.e., ergonomic technology reports) from the website of “Institute of Occupational Safety and Health” are applied as examples to evaluate the proposed model. The verification results show that the developed system is a high-performance knowledge document retrieval system. As a whole, this research provides an approach for knowledge demanders to efficiently and accurately acquire the domain knowledge documents.

**Keywords:** Institute of Occupational Safety and Health, Knowledge Management, Data Mining, Document Summarization Technology, Semantic Analysis

# 目錄

誌謝.....	I
摘要.....	II
Abstract.....	III
目錄.....	IV
圖目錄.....	VI
表目錄.....	X
第一章、研究背景.....	1
1.1 研究動機與目的.....	1
1.2 研究步驟.....	4
第二章、文獻回顧.....	7
2.1 研究定位.....	7
2.2 問答應用與技術分析.....	8
2.2.1 問答應用類型分析.....	8
2.2.2 問答技術分析.....	12
2.3 摘要應用與技術分析.....	16
2.3.1 摘要領域分析.....	16
2.3.2 摘要技術分析.....	20
2.4 小結.....	24
第三章、結合語意解析與文件摘要技術之知識文件檢索模式.....	25
3.1 知識文件表達項目解析模組.....	26
3.1.1 建立知識文件之表達項目.....	26
3.2 問答解析模組.....	33
3.2.1 詢問詞隱含目的判斷.....	34
3.2.2 詞彙類別相似度判斷.....	35
3.3 結構化摘要模組.....	37



3.3.1 簡略部分建立 .....	38
3.3.2 詳述部分建立 .....	44
3.4 結論 .....	50
第四章、系統架構規劃 .....	51
4.1 結合語意解析與文件摘要技術之知識文件檢索系統之流程架構 .....	51
4.2 系統功能架構 .....	52
4.3 資料模式定義 .....	56
4.4 系統流程 .....	58
4.4.1 系統功能流程 .....	58
4.4.2 系統資料流程 .....	65
4.5 系統開發工具 .....	65
第五章、系統實作與案例分析 .....	67
5.1 系統案例之應用流程 .....	67
5.2 系統驗證與評估 .....	83
5.3 小結 .....	105
第六章、結論與未來發展 .....	106
6.1 論文總結 .....	106
6.2 未來展望 .....	109
參考文獻 .....	110
附錄、系統功能操作說明 .....	116

## 圖目錄

圖 1.1、知識搜尋者取得資訊之既有模式 AS-IS MODEL.....	2
圖 1.2、知識搜尋者取得資訊之期望模式 TO-BE MODEL .....	3
圖 1.3、研究架構.....	6
圖 2.1、研究定位圖.....	8
圖 3.1、結合語意解析與文件摘要技術之知識文件檢索方法論之流程架構.....	25
圖 3.2、知識文件表達項目解析與觀念性語句擷取示意圖.....	26
圖 3.3、問答解析示意圖.....	33
圖 4.1、結合語意解析與文件摘要技術之知識文件檢索系統之流程架構.....	51
圖 4.2、結合語意解析與文件摘要技術之知識文件檢索系統之功能架構.....	53
圖 4.3、結合語意解析與文件摘要技術之知識文件檢索系統運作架構.....	56
圖 4.4、結合語意解析與文件摘要技術之知識文件檢索系統之資料關聯.....	58
圖 4.5、「知識文件表達項目解析模組」功能流程.....	59
圖 4.6、「知識文件問答解析模組」功能流程.....	60
圖 4.7、「知識文件結構化摘要推論模組」功能流程.....	61
圖 4.8、「知識文件問答維護模組」功能流程.....	62
圖 4.9、「知識文件管理模組」功能流程.....	63
圖 4.10、「系統參數設定模組」功能流程.....	64
圖 4.11、系統資料流程.....	65
圖 5.1、知識文件結構化摘要系統之應用流程.....	67
圖 5.2、勞工安全衛生知識網之研究報告.....	68
圖 5.3、知識文件管理模組-輸入文件資料.....	69
圖 5.4、知識文件管理模組-文件新增成功.....	69
圖 5.5、系統參數設定模組-系統參數檢視.....	70
圖 5.6、系統參數設定模組-系統參數設定.....	70
圖 5.7、詢問詞與回覆詞輸入畫面.....	71
圖 5.8、詢問詞與回覆詞配對結果畫面(1).....	71
圖 5.9、詢問詞與回覆詞配對結果畫面(2).....	72
圖 5.10、詢問詞與回覆詞新增成功畫面.....	72

圖 5.11、知識文件表達項目解析查詢介面 .....	73
圖 5.12、知識文件表達項目解析查詢介面 .....	73
圖 5.13、知識文件表達項目解析結果(1) .....	74
圖 5.14、知識文件表達項目解析結果(2) .....	74
圖 5.15、知識文件表達項目解析結果(3) .....	75
圖 5.16、知識文件問答解析-詢問詞解析目的 .....	75
圖 5.17、知識文件問答解析-文件篩選設定 .....	76
圖 5.18、知識文件問答解析-文件篩選結果 .....	76
圖 5.19、知識文件問答解析-文件下載 .....	77
圖 5.20、知識文件結構化摘要-查詢文件 .....	78
圖 5.21、知識文件結構化摘要-文件查詢結果 .....	78
圖 5.22、知識文件結構化摘要-簡略結構化摘要解析 .....	79
圖 5.23、知識文件結構化摘要-簡略結構化摘要呈現 .....	79
圖 5.24、知識文件結構化摘要-詳述結構化摘要呈現 .....	80
圖 5.25、知識文件問答維護-詢問詞與回覆詞查詢畫面 .....	80
圖 5.26、知識文件問答維護-詢問詞與回覆詞查詢結果畫面 .....	81
圖 5.27、知識文件管理模組-文件查詢 .....	81
圖 5.28、知識文件管理模組-文件查詢結果(1) .....	82
圖 5.29、知識文件管理模組-文件查詢結果(2) .....	82
圖 5.30、知識文件管理模組-文件查詢結果(3) .....	82
圖 5.31、驗證資料蒐集與建置 .....	86
圖 5.32、知識文件問答解析第一階段之回覆詞召回率之分布趨勢 .....	89
圖 5.33、知識文件問答解析第一階段之回覆詞正確率之分布趨勢 .....	90
圖 5.34、知識文件問答解析各驗證週期之績效分布趨勢 .....	90
圖 5.35、第一階段知識文件之摘要滿意度之分佈趨勢 .....	100
圖 5.36、第一階段知識文件之簡略摘要呈現效果之分佈趨勢 .....	101
圖 5.37、第一階段知識文件之詳述摘要呈現效果之分佈趨勢 .....	101
圖 5.38、各驗證週期之結構化摘要滿意度之分佈趨勢 .....	103
圖 5.39、各驗證週期之結構化摘要文字呈現效果之分佈趨勢 .....	103
圖 A.1、知識文件表達項目解析 .....	117

圖 B.1、詢問詞解析與文件篩選(1).....	119
圖 B.2、詢問詞解析與文件篩選(2).....	119
圖 B.3、詢問詞解析與文件篩選(3).....	119
圖 B.4、詢問詞解析與文件篩選(4).....	120
圖 B.5、詢問詞解析與文件篩選(5).....	120
圖 B.6、詢問詞解析與文件篩選(6).....	120
圖 B.7、詢問詞解析與文件篩選(7).....	121
圖 B.8、詢問詞解析與文件篩選(8).....	121
圖 B.9、詢問詞解析與文件篩選(9).....	121
圖 B.10、詢問詞解析與文件篩選(10).....	122
圖 B.11、詢問詞解析與文件篩選(11).....	122
圖 B.12、詢問詞解析與文件篩選(12).....	123
圖 B.13、詢問詞解析與文件篩選(13).....	123
圖 C.1、簡略結構化摘要推論畫面(1).....	125
圖 C.2、簡略結構化摘要推論畫面(2).....	126
圖 C.3、簡略結構化摘要推論畫面(3).....	126
圖 C.4、簡略結構化摘要推論畫面(4).....	126
圖 C.5、簡略結構化摘要推論畫面(5).....	127
圖 C.6、詳述摘要畫面(1).....	128
圖 C.7、詳述摘要畫面(2).....	129
圖 C.8、詳述摘要畫面(3).....	129
圖 C.9、詳述摘要畫面(4).....	129
圖 C.10、詳述摘要畫面(5).....	130
圖 D.1、知識文件新增(1).....	131
圖 D.2、知識文件新增(2).....	131
圖 D.3、知識文件新增(3).....	132
圖 D.4、知識文件查詢(1).....	133
圖 D.5、知識文件查詢(2).....	133
圖 D.6、知識文件查詢(3).....	133
圖 D.7、知識文件查詢(4).....	134

圖 D.8、知識文件查詢(5)	134
圖 D.9、知識文件查詢(6)	134
圖 D.10、知識文件修改(1)	135
圖 D.11、知識文件修改(2)	136
圖 D.12、知識文件修改(3)	136
圖 D.13、知識文件刪除(1)	137
圖 D.14、知識文件刪除(2)	138
圖 D.15、知識文件刪除(3)	138
圖 E.1、系統門檻值檢視畫面	140
圖 E.2、系統門檻值設定畫面	140
圖 E.3、系統門檻值修改成功畫面	141
圖 E.4、系統權重值檢視畫面	142
圖 E.5、系統權重值設定畫面	142
圖 E.6、系統權重值設定成功畫面	143
圖 F.1、詢問詞與回覆詞輸入畫面	144
圖 F.2、詢問詞與回覆詞配對結果畫面(1)	144
圖 F.3、詢問詞與回覆詞配對結果畫面(2)	145
圖 F.4、詢問詞與回覆詞新增成功畫面	145
圖 F.5、詢問詞與回覆詞查詢畫面	146
圖 F.6、詢問詞與回覆詞查詢結果畫面	146
圖 F.7、詢問詞與回覆詞查詢畫面	147
圖 F.8、詢問詞與回覆詞查詢結果畫面	147
圖 F.9、詢問詞與回覆詞修改畫面	148
圖 F.10、詢問詞與回覆詞修改成功畫面	148
圖 F.11、詢問詞與回覆詞查詢畫面	149
圖 F.12、詢問詞與回覆詞查詢結果畫面	149
圖 F.13、詢問詞與回覆詞資料確認畫面	150
圖 F.14、詢問詞與回覆詞刪除成功畫面	150

## 表目錄

表 2.1、問答應用領域文獻彙整表 .....	11
表 2.2、問答技術文獻彙整表 .....	15
表 2.3、摘要應用領域文獻彙整表 .....	19
表 2.4、摘要建立文獻彙整表 .....	23
表 3.1、表達細項與對應內容及表達方式(1) .....	26
表 3.1、表達細項與對應內容及表達方式(2) .....	27
表 3.2、輔助表達細項與對應內容及表達方式 .....	27
表 5.1、使用者所蒐集之研究報告 .....	68
表 5.2、知識搜尋者詢問句資料表 .....	83
表 5.3、勞工安全衛生研究所定義之雙語詞彙資料表 .....	84
表 5.4、知識文件表達項目資料表 .....	85
表 5.5、知識文件問答解析之詢問句與回覆句組合 .....	88
表 5.6、知識文件問答解析第一階段之受測詢問句結果分析 .....	89
表 5.7、知識文件問答解析績效彙整 .....	91
表 5.8、知識文件問答解析綜合兩階段之驗證績效彙整 .....	92
表 5.9、知識文件結構化摘要驗證之測試資料（其中 8 份） .....	93
表 5.10、知識文件結構化摘要驗證之測試資料（其中 7 份） .....	94
表 5.11、評級項目與評級考量標準資料表 .....	95
表 5.12、知識文件結構化摘要驗證之測試資料（其中 5 份） .....	98
表 5.13、知識文件結構化摘要第一階段實際結果呈現（以 5 份為例）(1) .....	98
表 5.13、知識文件結構化摘要第一階段實際結果呈現（以 5 份為例）(2) .....	99
表 5.14、第一階段知識文件之簡略摘要滿意度指標績效彙整 .....	99
表 5.15、第一階段知識文件之摘要呈現效果指標績效彙整 .....	100
表 5.16、第二階段第六週期實際結果呈現(1) .....	102
表 5.16、第二階段第六週期實際結果呈現(2) .....	103
表 5.17、知識文件結構化摘要綜合兩階段之驗證績效彙整 .....	104

# 第一章、研究背景

本章乃說明本論文之研究背景，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

由於網際網路發達，人們已習慣以網路進行資訊與知識之獲取，卻容易造成資訊過量等問題產生，因此延伸出相關研究與技術以協助知識搜尋者能快速且有效取得資訊。為協助知識搜尋者方便取得資訊，多數解決方案乃建立網站將相關領域文件彙整並提供給搜尋者分享，此類網站即具有該領域之代表特性，亦表示知識搜尋者欲查詢相關資訊時，期望可以直覺方式於該網站查詢所需資訊，進而節省資訊搜尋時間。

針對上述，雖有特定領域網站作為資訊搜尋窗口以縮小搜尋範圍，但由於網路資源龐大，從中仍需進行文件分類、關鍵字查詢等方式以準確獲取所需文件。本論文以「勞工安全衛生知識網」為例，該網站彙集各行業職業傷害防制等勞工安全相關知識與文件，雖已針對主題進行文件分類，但主題下所含眾多議題之文件，導致無法立即判斷是否為所需資料，該知識網雖有關鍵字搜尋方式以進行查詢，但知識文件多為專業人員進行訪視、研究後撰寫，其文件內容大多以專業名詞、專業知識描述。因此，若無相關領域背景之知識搜尋者以關鍵字搜尋方式亦無法確切取得回饋，或是知識搜尋者無法定義關鍵字詞反而以揣測方式進行搜尋。此外，搜尋者必須透過文件標題先行篩選再以閱讀摘要方式判定文件取決，且該網頁之文件摘要呈現方式乃以作者自行撰寫，其自由形式呈現方式易影響搜尋者判斷文件取得。

根據上述內容，雖然目前網頁針對文件主題或是定義關鍵字以協助知識搜尋者取得資訊。然而，針對特定領域知識網若以此方式可能因專業知識不足，而導致知識搜尋者無法正確取得資訊；另外，文件摘要之呈現沒有統一格式或是字數之控管，導致搜尋者容易因個人喜好而影響文件篩選，進而降低知識網站分享知識功能。綜合上述，其既有之運作模式如圖 1.1 之 AS-IS Model 所示。



圖 1.1、知識搜尋者取得資訊之既有模式 As-Is Model

如圖 1.1 所示，目前知識搜尋者取得資訊之方式可透過關鍵字搜尋、主題分類與文件摘要閱讀等，但以上述方式若針對特定知識領域仍有部分瑕疵，本研究乃將問題彙整並列點如下：

1. 知識搜尋者無從定義專有名詞以進行關鍵字搜尋：於勞工安全領域中有許多該領域之專有名詞，並且多數文件之題目、關鍵字皆以專有名詞撰寫與定義，於此無相關背景或是間接人員之搜尋者較無明確關鍵字定義並搜尋，以至於無法立即、確切取得所需資訊。
2. 主題分類眾多、自由形式摘要呈現方式不一：雖然勞工安全衛生知識網已針對主題進行文件分類，但主題下所包含議題眾多且複雜，此外文件摘要呈現方式乃由作者自行撰寫，並以自由形式方式呈現，以此方式容易因文字長短、文章編排等因素影響搜尋者閱讀習慣，進而影響知識文件篩選條件。

以工廠之老闆舉例說明，假設工廠老闆欲改善員工工作所造成身體痠痛等職業傷害，透過關鍵字「痠痛」於勞工安全衛生知識網搜尋其結果卻包含許多無正相關之文件，其原因乃該領域所定義之專有名詞為「肌肉骨骼傷害」而非一般搜尋者所解讀之「痠痛」，於此造成關鍵字無法確實發揮其功能。假設工廠老闆已縮小搜尋範圍，但文件摘要尚未結構化，容易因個人習慣選擇摘要字數較短之文件作為取決考量，而錯過重要知識文件。



有鑑於此，為掌握搜尋者需求並針對問題提供確實回饋，本研究認為可在關鍵字搜尋技術與文件摘要呈現方式上加強搜尋者篩選之思維模式，以協助知識搜尋者能快速且確實求得需求文件。因此，如期望之運作模式如圖 1.2 所示，本論文之研究動機與目的可歸納為以下兩點：

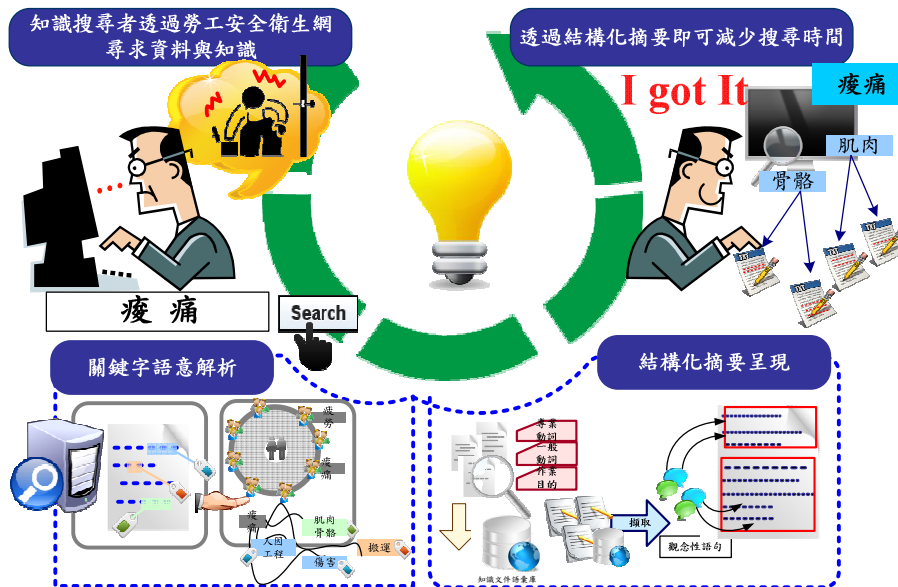


圖 1.2、知識搜尋者取得資訊之期望模式 To-Be Model

### 1. 建立知識語彙庫並加強關鍵字語意搜尋

由於勞工安全衛生知識網多為專業知識背景人士所撰寫，因而使用許多專有名詞做為文件之標題與摘要，但此方式造成無相關知識背景之一般搜尋者無法定義關鍵字以進行搜尋。因此，為加強關鍵字語意搜尋技術，本研究乃將文件進行解析並取得文件關鍵字詞，並與使用者搜尋字串進行語意關聯，以提高關鍵字搜尋之效能。

### 2. 結構化摘要以輔佐知識搜尋者檢視文件

當搜尋者縮小搜尋範圍欲以摘要閱讀方式篩選文件，但其摘要呈現乃由作者自行撰寫，以此方式易根據作者風格、喜好而有不同摘要呈現方式，以至於知識搜尋者容易因個人閱讀偏好，進而影響篩選文件。為改善搜尋者篩選文件之判斷，本研究乃將知識文件進行解析以取得文件結構、表達語彙等，並以文件重要性語彙為基，擷取文件觀念性語彙同時結合文件起承轉合呈現方式以形成結構化摘要。

整體而言，為協助知識搜尋者快速且有效取得資訊與知識，本研究乃解析知識文件之表達方式、呈現內容以及相關性語彙，並以此為基建立知識語彙庫，進而與搜尋字串進行語意關聯，以加強關鍵字搜尋技術。另外，亦可透過知識語彙庫取得文件之觀念性語句，並依照文件結構建立結構化摘要，以避免個人閱讀偏好影響文件之篩選，本研究之研究目的即縮短知識搜尋者搜尋文件之時間，進而加強勞工安全衛生知識網搜尋技術以提高勞工安全衛生知識網知識分享之成效。

## **1.2 研究步驟**

本研究之目的乃協助知識搜尋者快速且確實取得需求文件，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，並依照方法論建置一套結合語意解析與文件摘要技術之知識文件檢索系統，最後則以勞工安全衛生知識網進行系統案例驗證，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，本研究乃針對知識搜尋者搜尋索引行為以及透過文件摘要建立方式以進行探討，並彙整所涉及主題包含「問答應用與技術分析」與「摘要應用與技術分析」等議題，透過文獻蒐集與研讀，以了解知識文件解析與表達項目、關鍵字語意檢索與結構化摘要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，知識搜尋者習慣以自我解讀方式進行關鍵字串搜尋，但既有知識網多以關鍵字比對等方式進行搜尋文件，透過「問答應用與技術分析」之相關研究以得知問答解析中語意關聯方式與技術之建立；「摘要應用與技術分析」之相關研究以了解文件摘要於知識網適用性，並了解相關研究主要乃根據文件結構解析、語意問題等推論並建立摘要。由於既有知識網多僅以關鍵字進行搜尋，故解析後之文件較不符合需求者所需，且尚無明確、結構性之摘要呈現於知識搜尋者。因此，本研究乃針對知識文件先以解析，並取得文件之表達項目與觀念性語句，進而發展一套「結合語意解析與文件摘要技術之知識文件檢索模式」，此模式將解析後之專業語彙進行語意關聯，以協助關鍵字搜尋，再透過文件之表達項目建立語彙庫，同時亦擷取最具代表性語句並根據文件結構組合成結構化摘要，以協助知識搜尋者加速判定、篩選文件內容。

### **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「知識文件資料解析與探討」、「知識文件語意問答與摘要技術之探討」、「結合語意解析與文件摘要技術之知識文件檢索方法論發展」與「系統功能開發」，以下為各項主題之細述說明：

#### 主題一、知識文件資料解析與探討

- 蒐集並解析相關知識文件
- 定義具有知識文件表達項目
- 解析知識文件表達項目，並定義各表達細項
- 擷取具有觀念性與代表性之觀念語句，並建置知識語彙庫

#### 主題二、知識文件語意問答與摘要技術之探討

- 解析知識文件專有名詞語意關聯
- 計算觀念性語句具有文件代表性權重分配
- 建立結構化摘要判定法則

#### 主題三、語意解析與文件摘要技術方法論發展

- 建立知識文件表達項目解析方法論
- 建立知識文件觀念性語句擷取方法論
- 建立知識文件語意問答解析方法論
- 建立知識文件結構化摘要方法論

#### 主題四、系統功能開發

- 開發知識文件具有觀念性語句自動擷取功能
- 開發知識文件專有名詞語意問答功能
- 開發知識文件結構化摘要推論功能

### **步驟四、案例驗證**

於此步驟乃將本研究所建構之「結合語意解析與文件摘要技術之知識文件檢索系統」為基，並以勞工安全衛生知識網之技術叢書為驗證資料，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，並評估本研究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，本研究首先乃根據研究動機與目的蒐集知識文件表達項目分析、語意問答技術、摘要建立等相關文獻，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（即等知識文件資料解析與擷取、知識文件專有名詞語意關聯與知識文件結構化摘要推論等），並以此為基礎開發一套結合語意解析與文件摘要技術之知識文件檢索系統。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。本研究之研究架構如圖 1.3 所示。

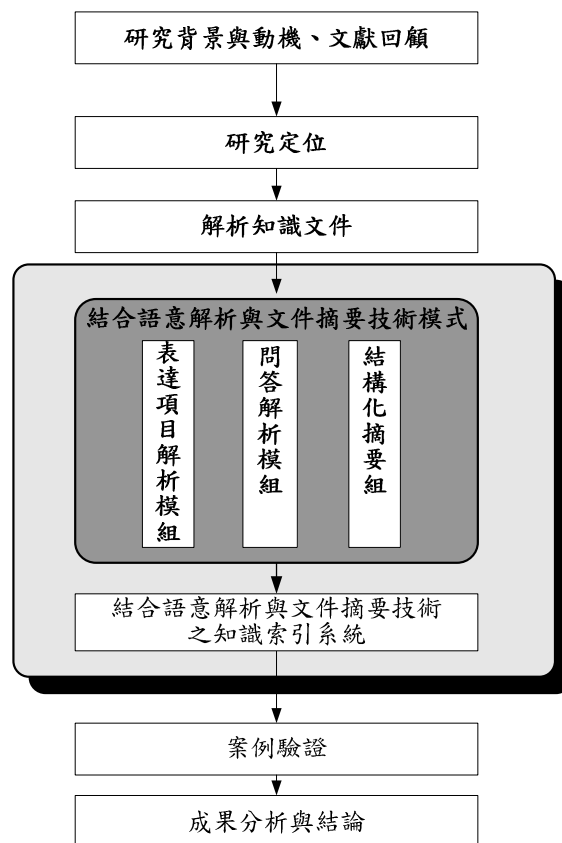


圖 1.3、研究架構

## 第二章、文獻回顧

本研究之目的乃協助資訊搜尋者於龐大之網路文件中，迅速且有效地搜尋其所需要之文件資料，以節省搜尋者過濾與篩選文件之時間。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「問答應用與技術分析」與「摘要應用與技術分析」等兩大研究方向，以下即針對此兩大主題之相關研究進行文獻回顧及探討。

於問答應用與技術之議題中，根據過去研究可將問答應用與技術分為「問答應用領域分析」以及「問答技術分析」兩方面進行探討，於問答應用領域中，相關研究之成果可以問答或是檢索方式進行探討，而問答技術分析課題而言，根據過去相關研究可根據問答生成或篩選條件等方式分成以「以文件主題分類」、「以文件特徵分類」及「以文件語意解析」等三方面進行探討。

於摘要應用與技術之議題中，過去相關研究可分為「摘要應用領域分析」與「摘要技術分析」等二方面進行探討。於摘要應用領域中，大多乃針對具有專業知識之特定知識領域，或是資料量過大或是具有相同主題文件之網際網路領域進行摘要應用，而摘要技術分析課題而言，過去相關研究之成果可將摘要建立技術分為機器學習之監督式學習以及非監督式學習等兩種進行探討。

綜合以上所述，本論文所涉及之各項主題領域可以圖 2.1 呈現其架構關係；圖中灰色部分乃代表本研究所強調研究之主題。如圖 2.1 所示，以本研究所發展之方法論為主軸，藉由過去許多相關之研究成果，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「問答應用與技術分析」與「摘要應用與技術分析」兩大議題進行說明。

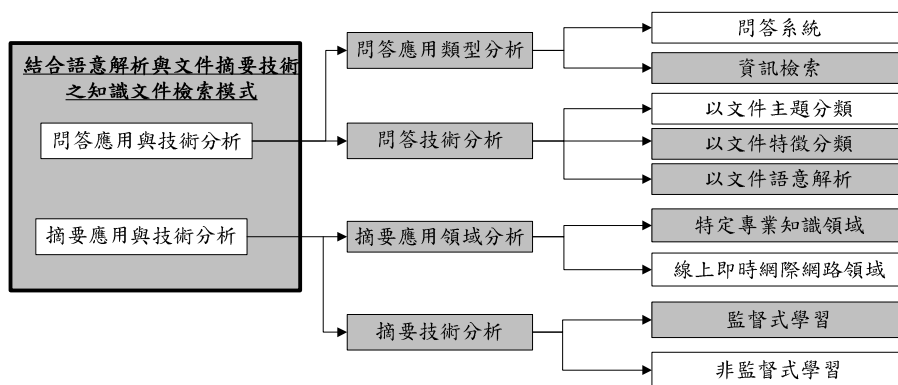


圖 2.1、研究定位圖

## 2.2 問答應用與技術分析

對於問答應用與技術議題而言，本研究乃問答應用類型以及問答技術進行相關文獻探討，期望於其中觀察應用於不同類型之不同角度與層面解析，以更深層瞭解問答應用與技術特性。

### 2.2.1 問答應用類型分析

於問答應用領域探討中，本研究乃針對「問答系統」以及「資訊檢索」等兩主題進行相關文獻探討，期望從中探討問答應用所涉及之範圍與領域。以下即針對「問答系統」以及「資訊檢索」進行文獻探討。

#### (A) 問答系統

針對問答應用類型大多以搜尋者提出詢問句方式提出問題，並取得回覆句作為答案以回應，常以一問一答之方式進而取得答案，以此方式即可應用於醫學領域上協助搜尋者擷取複雜、無格式化之臨床報告重點，如 **Cao 等人 (2011)** 乃針對醫學上臨床報告建立一套線上問答系統 (AskHERMES)，以並將擷取重點總結後形成答案，而 **Terol 等人 (2007)** 則提出一套醫療互聯網問答系統，該研究透過自然語言處理以及廣泛解析器 (Broad Coverage Parser, MINIPAR) 解析問句中各詞性與依賴關係 (Dependency Relationship)，進而結合統一醫學語言系統 (Unified Medical Language System, UMLS) 與醫療術語網 (WordNet) 選擇相關資訊並擷取作為回覆句，以確保問答系統之資訊正確性，且能同時處理專有名詞與自然語言之問答，並以互聯網方式更能廣泛使用。**Guo 及 Zhang (2009)** 藉由語意關係探索提出資料庫於領域知識上表示本體語意之方法，以醫療領域為主題，並透過自然語言分析 (Natural Language; NL)，將醫療領域之語言及

完整口語問答方式進行關鍵連結。另外，為了取得最確實答案，**Dalmas 和 Webber (2007)** 建立一套問題解答模式 (QA Answer Model, QAAM)，即避免以單一答案作為回覆句易導致其他資訊價值流失，進而問題分類並透過模式圖型程序 (Model-View-Controller, MVC) 進行準則分析，以確立回覆句與回覆句之間之關聯性，除了篩選回覆句外，亦分析回覆句之背景並作為評估準則，以確定回覆句能於不同背景下使用，並表示以組合回覆句方式更具有資訊參考能力。**Scheffer (2004)** 先制定回答選項類別，並將客戶提問之電子郵件以支援向量機 (Support Vector Machine; SVM) 擷取郵件內容特徵，再以天真貝氏分類法 (Naive Bayes Classifier) 分析郵件特徵，以區分客戶於郵件中所提問題之所屬類別，再根據郵件所屬回答類別回應客戶對應回答內容，以達到客戶提問自動化回覆，**Lorch 等人 (2001)** 認為傳統檢索系統乃著重於快速回應使用者需求，而非針對問題給予確實回應。因此，提出一套醫學問答系統 (Medical Definitional Question Answering System, MedQA)，該研究以監督學習方式將問題自動分類 (如：分成 What、How 等，What 表示問定義而 How 則是問作法)，使用者即可確實從問題中取得答覆，改善傳統檢索系統以關鍵字方式快速回應。

## (B) 資訊檢索

資訊檢索多以定義關鍵字詞進行搜尋，待搜尋完後即可取得相關文件或資訊之回饋。對此，大多研究乃針對生物醫學領域進行資訊檢索，**Huang 等人 (2006)** 乃針對生物醫學文獻之關聯性分析，並提出一套混合式關係模式以擷取生物醫學文獻，以加強文獻之間關聯性同時提高資訊檢索之正確性。**Hahu 等人 (2002)** 基於醫療調查結果提出一套醫療自然語言系統 (MedSynDiKATe)，使檢索功能更加口語化。透過詞彙連結關係先將文件分析，並根據概念、語意關係將知識概念轉換成邏輯結構以整合成醫療知識庫。**Uzuner 等人 (2010)** 則針對臨床記錄提出以問題為導向之醫療記錄模式，以建立病歷之關聯性並應用於醫療記錄索引，進而可提供醫護人員較為正確之醫療訊息。該研究藉由醫學語言統一系統 (Unified Medical Language System, UMLS) 定義疾病類型與症狀，再以藉由支援向量機 (Support Vector Machines, SVM) 將病歷之重要語句、特點進行排序與分類，並透過特徵向量取得詞彙關聯性。**Rindfleisch 和 Fisman (2003)** 以 Hypernymic 概念為基提出一套語言結構以解釋並應用生物醫學領域檢索系統，透過統一醫學語言系統 (Unified Medical Language System, UMLS) 與自然語言處理以建立語意網路，再加入 Hypernymic 生物分類學概念使語意更具體，以此方式能提高語意精確度

亦能應用於資訊檢索技術以減少系統錯誤。此外，針對資訊量過於龐大，故無法針對使用者確切篩選文件之問題，Mingxin (2011) 以語意網為基建立一套資訊檢索系統 (Information retrieval, IR)，該研究以本體論先針對網頁文件之 HTML 語法以資訊描述架構 (RDF) 提取重要訊息，並將具有相同概念之資訊相互關聯，從中亦針對過去搜尋紀錄以取得網頁與關鍵字之關聯性並建立資料庫，其代理人即可根據資料庫與搜尋字串進行關聯並提供正確回應於使用者。Teng 等人 (2010) 則透過群集文件從中尋找主題方式，使文件更具關聯性。該研究乃蒐集被多人點選閱讀之文件，同時研究文件之關聯性與共同特徵 (即基本要素 Basic Element, BE: 包含語句之開頭、修飾詞與連結關係)，利用共同特徵作為文件篩選條件並進行群集分析。當中，Yin (2011) 即以引用圖特性建立生物醫學領域文獻檢索模式，以取得文獻檢索之最佳路徑。該研究先以連鎖分析演算法 (Degree Distribution) 分析文獻上下文之內容被引用機率，並透過資訊關聯方式程度 (Degree) 作為程度範圍，最後以關鍵字權重 (Okapi) 作為檢索基準線，以判斷該圖之節點是否為範圍內，進而增進文獻檢索之效果。

綜上所述，問答應用領域可針對特定知識領域進行詢問方式取得回應，或以檢索方式取得資訊，皆有許多文獻進行研究各領域之問答機制效果，以下乃針對各研究之問答系統應用層面分別彙整成表，如表 2.1 所示。



表 2.1、問答應用領域文獻彙整表

應用領域	相關研究	技術名稱	目的與結果
問答系統	Cao 等人 (2010)	線上問答系統 (AskHERMES)	以擷取複雜、無格式化之臨床報告重點，並將擷取重點並自動分類問題主題以確立關鍵字，總結後形成答案。
	Terol 等人 (2007)	醫療互聯網問答系統	透過自然語言處理以及廣泛解析器，結合統一醫學語言系統與醫療術語網以確保問答系統之資訊正確性，且能同時處理專有名詞與自然語言之問答。
	Dalmas 和 Webber (2007)	問題解答模式	透過模式圖型程序進行準則分析，以確立回覆句與回覆句之間之關聯性，並考量回覆句之背景並作為評估準則，以確定回覆句能於不同背景下使用。
	Guo 及 Zhang (2009)	醫療領域問答系統	藉由語意關係探索提出資料庫於領域知識上表示本體語意之方法，以醫療領域為主題，並透過自然語言分析進行關鍵連結。
	Scheffer (2004)	電子郵件自動回覆系統	以支援向量機擷取郵件內容特徵，再以天真貝氏分類法分析郵件特徵，以區分客戶於郵件中所提問題之所屬類別，再根據郵件所屬回答類別回應客戶對應回答內容。
	Lorch 等人 (2001)	醫學問答系統	該研究以監督學習方式將問題自動分類，使用者即可確實從問題中取得答覆，改善傳統檢索系統以關鍵字方式快速回應。
資訊檢索	Huang 等人 (2006)	混合式關係模式	擷取生物醫學文獻，透過淺層句法解析 (Shallow Parsing) 取得語法、語意之結構，再以貪婪法進行匹配以訓練模式擷取生物醫學文獻之主題。
	Hahu 等人 (2002)	醫療自然語言系統	透過詞彙連結關係先將文件分析，並根據概念、語意關係將知識概念轉換成邏輯結構以整合成醫療知識庫。
	Uzuner 等人 (2010)	以問題為導向之醫療記錄模式	透過語意關係提出以問題為導向之醫療記錄模式，藉由醫學語言統一系統定義疾病類型與症狀，再以藉由支援向量機將病歷之重要語句、特點進行排序與分類。
	Mingxin (2011)	以語意網為基建立一套資訊檢索系統	以本體論先針對網頁文件之 HTML 語法以資訊描述架構提取重要訊息，並將具有相同概念之資訊相互關聯，從中亦針對過去搜尋紀錄以取得網頁與關鍵字之關聯性並建立資料庫。
	Rindflesch 和 Fiszman (2003)	統一醫學語言系統	以自然語言處理以建立語意網路，再加入 Hypernymic 生物分類學概念使語意更具體，以此方式能提高語意精確度亦能應用於資訊檢索技術以減少系統錯誤。
	Teng 等人 (2010)	以使用者為導向之文件摘要檢索	根據文件之關聯性與共同特徵 (即基本要素 Basic Element, BE: 包含語句之開頭、修飾詞與連結關係)，利用共同特徵作為文件篩選條件並進行群集分析。
	Yin (2011)	引用圖特性建立生物醫學領域文獻檢索	以連鎖分析演算法分析文獻上下文之內容被引用機率，並透過資訊關聯方式程度作為程度範圍，最後以關鍵字權重作為檢索基準線，以判斷該圖之節點是否為範圍內，進而增進文獻檢索之效果。

## 2.2.2 問答技術分析

於問答技術中，過去研究多數都針對文件先行解析並取得文件之特徵與主題等，再進行回覆之任務；因此本研究乃針對「以文件主題分類」、「以文件特徵分類」及「以文件語意分類」進行相關文獻探討。

### (A)以文件主題分類

針對文件主題分析部分，Oh 等人 (2012) 乃提出一套問答系統學習機制，透過現有問答文件以分析結構（如詞性與專有名詞等），再以詞義消歧法進行語意分析，從中取得問句與回覆句之組合（回答格式、回答主題、目標與預期回答內容），經由學習機制即可取得問句之重要字詞並連結相關回覆文件以取得回覆句，以此方式能更精確解決問題，同時亦提高問題解決之效率。Yangarber 等人 (2000) 提出一套非監督式自動化關聯評估模式（The Discovery Procedure, ExDISCO），首先將文件進行文法結構拆解（即主詞、動詞與受詞），再根據詞彙鏈、詞頻等方式找出一組候選模式（即種子模式 Seed Patterns），並透過種子模式與文件內結構單字比對以判斷文件所屬類別，最後即可根據類別進行文件關聯分析，以此方式無需透過讀者自行解讀或是標記即可自動關聯相關資料以加強篩選。Han 等人 (2007) 提出以先確定問題類型並建立各類型相關詞彙，於此即可先從問題中判決問題目標並分析問題檢索之類別，以進行擴大查詢並提取過去共同出現之資訊作為外部資源。該研究乃以功能附屬分析器（Conexor FDG Parser）提取動詞或名詞，並以擴展技術獲取相關資訊之語句，最後以各種準則（冗餘或具有術語）進行評分與排序與擷取，以此方式能消除過於冗餘以精簡語句，並遞補更具意義之回覆句作為參考。

### (B)以文件特徵分類

藉由文件特徵分析，即可根據分析結果與條件進行關聯以及相似性連結。因此，Jones 和 Love (2007) 提出於文件關聯中具有主題角色扮演模式，若關係越具相似性則表示兩者之間存在一個共同角色，此外亦有其他扮演關聯中之某角色（即掠食者則會有狩獵關係作用出現），透過背景環境、關係為匹配準則以取得文件之共同關係。透過模式以角色作用推理出具有相似之文件，其方式能更深入與擴展關聯分析機制。當中針對詢問句可包含多種問題條件，如：世界上最長的河流是什麼？當中條件有區域性（世界上）、比較性（最長的）以及詢問物（河流），因而無法以單一角度進行回覆。於此，Oh 等人

(2011) 乃提出一套複合式問答系統 (Compositional QA)，針對問題格式 (單一問項或是多重問項)、主題、問題限制 (時間或地點等) 作為判斷準則並以此為基以得知回覆句之類型與格式，進而以語意匹配模式 (Lexico-Semantic Patterns, LSPs) 針對問題條件分別找尋答案，最後再將答案組成並回覆。Ko 等人 (2004) 提出以重要性語句作為文件分類之根據，以加強文件分類技術。該技術之概念乃由文件摘要延伸，透過語句內容與標題符合度、語句辭頻等權重計算取得重要性得分，根據重要性得分數取得文件分類之依據。於過去研究常以關鍵字詞進行分類，而忽視作者資訊亦可成為分類準則。因此，Legara 等人 (2011) 提出一套根據作者署名自動化分類專欄文章，根據作者寫作風格生成四個準則：(1)句法、(2)結構、(3)詞彙、(4)內容特點，並依照上述準則進行頻率排序 (Frequency Ranking Method, FRM) 以及影響大小排序 (Effect Size Ranking Method, ESRM)，最後乃以排序結果作為分類標準，該研究將此作法應用於意見專欄中並發現以此方式結果乃優於頻率分類法。

### (C)以文件語意解析

大多問答系統 (Question Answering, QA) 乃先定義特定名詞 (Named Entities, NEs) 以建立問答機制，但若以生活語言 (Common Noun, CN) 為基之詢問句則無法確實取得回覆。是故，Moreda 等人 (2011) 以語意規則與語彙網 (WordNet) 為基建立一套語意問答系統。該研究首先進行詢問句演算分析以了解詢問句之語意角色，進而加入語彙網 (WordNet) 語意模式以取得過去類似問題之回覆句，再以主要關鍵標記 (Prop Bank, PB) 方式進行專有名詞與生活語言關聯。Dunlavy (2007) 等人提出一套整合資訊查詢系統 (The Query, Cluster, Summarize, QCS)，該系統先根據文句位置、文件內容等特性標記文件之重要語句，透過潛在語意索引技術 (Latent semantic indexing, LSI) 進行相關性查詢分析，再以修剪法 (Trimming) 去除具有相關語句群集之冗長語句以形成摘要，最後以相關性將遞減列出提高查詢效益。Erdogan 等人 (2005) 將語法結構結合語意並提出三種方法以建立語言模式加強語音辨識能力。透過潛在語意分析、雙連語意詞彙模型 (Two-Level Semantic-Lexical Modeling) 解析資料之相關性並以機率模型計算詞彙序列，最後以最大熵法 (Maximum Entropy, ME) 結合成語言模式，經由此模式能增加語意之辨識能力亦能提高詞彙整合之緊密度。Ruiz-Casado 等人 (2007) 提出一套以詞彙模式自動識別語意關係，首先以最小距離法計算兩者間之相似性並以矩陣方式呈現，並從中去除無相關性，最後整合成一套特定關係，以擷取語意間特殊關係 (例如：

木星之於太陽系關係：木星是太陽系的一部分）亦能基於文件、特定領域形成整體-部分之關係（Holonymy）。透過統計分析可自動預測兩事件之關聯性，但由於語意關係易造成語句具有多種解釋，因此 Dorr 和 Gaasterland (2007) 提出一套考量時態與語意關係之結合模式，基於時序關係、事件觀點等將相關事件結合。首先乃根據動詞時態（Basic Tense Structure, BTS, 包含現在、過去、未來）轉成矩陣並以相關性形成時態結構（Complex Tense Structure, CTS），從中加入語法約束時態結構形成延伸時態結構（Constraint on Derived Tense Structures, CDTS），最後以組合排序法組織主題之關聯性即可擷取相關性語句，由於該模式乃將因果關係以及時序連結作為考量準則，透過重新分析與解釋可提高多筆事件關聯之整合性。

綜合上述，針對問答技術層面皆有許多文獻研究，無論根據文件之主題、文件之特徵或根據文件語意方式，皆使資訊解析方式更加多元、其結果更加精準，而依其不同應用與技術分別彙整成表，如表 2.2 所示。

表 2.2、問答技術文獻彙整表

技術部分	相關研究	技術名稱	目的與結果
以文件主題分類	Oh 等人 (2012)	問答系統學習機制	現有問答文件以分析結構如 (詞性與專有名詞等), 再以詞義消歧法進行語意分析, 從中取得問句與回覆句之組合。
	Yangarber 等人 (2000)	自動化關聯評估模式	將文件進行文法結構拆解 (即主詞、動詞與受詞), 再根據詞彙鏈、詞頻等方式找出一組候選模式, 並透過種子模式與文件內結構單字比對以判斷文件所屬類別。
	Han 等人 (2007)	分析問題檢索之類別	確定問題類型並建立各類型相關詞彙, 於此即可先從問題中判決問題目標並分析問題檢索之類別, 以進行擴大查詢並提取過去共同出現之資訊作為外部資源。
以文件特徵分類	Jones 和 Love (2007)	主題角色扮演模式	透過潛在語意分析並透過背景環境、關係為匹配準則以取得文件之共同關係。透過模式以角色作用推理出具有相似之文件。
	Ko 等人 (2004)	以重要性語句作為文件分類	透過語句內容與標題符合度、語句辭頻等權重計算取得重要性得分, 根據重要性得分數取得文件分類之依據。
	Legara 等人 (2011)	根據作者署名自動化分類	根據作者寫作風格生成四個準則: (1)句法、(2)結構、(3)詞彙、(4)內容特點, 並依照上述準則進行頻率排序以及影響大小排序, 最後乃以排序結果作為分類標準。
	Oh 等人 (2011)	複合式問答系統 (Compositional QA)	以準則判斷方式進行問題分析, 針對問題格式 (單一問項或是多重問項)、主題、問題限制 (時間或地點等) 作為判斷準則並以此為基以得知回覆句之類型與格式。
以文件語意解析	Moreda (2011)	語意問答系統	進行詢問句演算分析以了解詢問句之語意角色, 進而加入語彙網 (WordNet) 語意模式以取得過去類似問題之回覆句, 再以主要關鍵標記 (Prop Bank, PB) 方式進行專有名詞與生活語言關聯。
	Dunlavy (2007)	整合資訊查詢系統	根據文句位置、文件內容等特性標記文件之重要語句, 透過潛在語意索引技術進行相關性查詢分析, 再以修剪法去除具有相關語句群集之冗長語句以形成摘要, 最後以相關性將遞減列出提高查詢效益。
	Erdogan 等人 (2005)	語音辨識能力	透過潛在語意分析、雙連語意詞彙模型解析資料之相關性並以機率模型計算詞彙序列。
	Ruiz-Casado 等人 (2007)	以詞彙模式自動識別語意關係	以最小距離法計算兩者間之相似性並以矩陣方式呈現, 並從中去除無相關性, 最後整合成一套特定關係, 以擷取語意間特殊關係
	Dorr 和 Gaasterland (2007)	考量時態與語意關係之結合模式	根據動詞時態轉成矩陣並以相關性形成時態結構, 從中加入語法約束時態結構形成延伸時態結構, 最後以組合排序法組織主題之關聯性。

## 2.3 摘要應用與技術分析

對於摘要應用與技術議題而言，本研究針對摘要應用方式與摘要建立技術進行相關文獻探討，當中摘要應用可分為領域分析與文件類型分析與摘要技術分析三部分，期望於其中觀察摘要文件之解析方式，以更深一層瞭解文件潛在之資料特性，從中建立摘要。

### 2.3.1 摘要領域分析

於摘要應用領域探討中，本研究乃針對「特定專業知識領域」以及「線上即時網際網路領域」等兩主題進行相關文獻探討，期望從中探討問答應用所涉及之範圍與領域。以下即針對「特定專業知識領域」以及「線上即時網際網路領域」進行文獻探討。

#### (A) 特定專業知識領域

對於文件摘要大多乃應用於資料過量大或是具有過多相似主題文件之領域，如醫療領域、線上新聞與法律領域等。針對醫學領域文刊數量過多問題，Elhadad 等人 (2005) 乃建立一套統一摘要模式，基於搜尋檢索技術將其結果進行摘要總結以協助使用者更有效瀏覽。該系統乃以主題樹概念將主題作為連結節點，先將文件進行主題分群再根據檢索主題找出相關文件，再將相關文件進行多文件摘要以形成搜尋總結。於此，該模式能過濾大量醫學資訊，並提供使用者特定查詢之具體總結。Ling 等人 (2007) 提出一套以基因學角度為基之半結構化摘要技術，醫學學者可藉由此技術於醫學資料庫中將文獻進行半結構化摘要，更為方便與查詢於該領域最新資訊。該技術經由字典檢索、過濾同義詞方式訓練模型取出文件關鍵字並作為目標基因，透過目標基因作為語句擷取條件，再以空間向量模型 (Vector Space Model, VSM)、機率語言模型 (Probabilistic Language Model) 訓練模型於語意部分之評分，最後乃擷取語句並形成摘要。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，經由圖形介面工具建立視覺化摘要以產生概念與主題，透過主題樹即可判定該術語之語意類型並擷取。該模式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。此外，針對法律領域文刊數量過多問題，為改善手寫法案之效率，Moens (2007) 將法院判決紀錄等形式文書予以電子化，並提出一套自動化個案摘要與檢索功能，使用者可根據相關案例找尋論點與諮詢。由於法院判決紀錄屬高度結構化文件，並具有固定出現概念序列，該研究將法院判決文件劃分為：(1)基本要素 (包

含被告、原告、案由、投訴類型等)、(2)背景、(3)相關法律問題、(4)辯護論點、(5)裁決與處置,根據上述劃分作為結構摘要關鍵字擷取依據。藉由分配概念將文件之內容分配至所屬概念,並基於圖形概念方法將鏈結相關性形成摘要,此作法不僅節省人力與時間,更能替律師將相關案件進行總結以方便尋找論據。**Moens 等人 (2005)** 提出一套透過通用於單一文件與多文件摘要模式,並以層級化方式呈現主題、摘要及細節部分。該模型先以詞性標記 (Part-Of-Speech, POS) 標籤具有文法結構和語意關係之語句,再以演算法根據主題、詞彙於文件之頻率、結束符號等,將文件層級化分割建立主題樹,透過主題樹取得語句以及對應語句以形成摘要。此外針對多文件而言,該研究將群聚技術結合詞性標記技術,即可藉由主題樹方式可依照層級形成摘要。該模式可應用於不同類型文件(如英文雜誌、荷蘭百科全書等);此外,該模式亦可將文件壓縮以取得重要資訊並應用於新聞頭條上。**Xie 和 Liu (2010)** 則針對特定會議記錄提出以迴歸監督式學習方式統計分類擷取具代表改善會議記錄總結,透過支援向量機 (Support Vector Machines, SVM) 監督學習方式標記並分類具有代表性文句,並以增加取樣 (Up-sampling)、減少取樣 (Down-sampling) 和重複取樣 (Re-sampling) 三種不同取樣方式,以提高於少數類別 (Minority Class) 中擷取文句之預測準確性。

## (B)線上即時網際網路領域

針對大量網頁文件搜尋不易之問題,**Bouras 等人 (2008)** 提出一套個人化新聞索引系統 (PeRSSonal)。該研究將網路新聞、文章自動化摘要並透過分類以產生文字標籤 (Text Labeling),結合網路新聞 RSS 訂閱服務即可進行個人化服務。藉由字頻計算取得新聞文件之關鍵字,並以餘弦相似性權重計算與分類新聞,最後選擇具有代表性語句形成摘要。針對網路新聞於手持行動裝置呈現與負載等問題,**Yang 和 Wang (2007)** 針對手持行動裝置之文件進行自動化摘要。該研究利用碎形理論 (Fractal Theory) 形成摘要模型,並以樹狀結構或是分層方式加強手持行動裝置視覺化操作。該研究乃以碎形理論可將文件結構進行層級化(章節、段落、文句、字詞等),藉由分層即可取得章節區塊並計算文句於父節點之價值,找尋該章節之代表,再根據壓縮比決定文件摘要擷取文句之數量。此外,**Zajic 等人 (2007)** 為克服新聞標題文字長度之限制,運用文件壓縮技術將擷取單一文件之摘要語句,以形成候選語句並彙整成多文件摘要。其壓縮方式可透過兩種方法進行:(1)解析與整理 (Parse-And-Trim)、(2)基於隱藏式馬可夫鏈模型 (Hidden Markov Model, HMM) 之改良模型:HMM Hedge (Hidden Markov Model

HEaDline GEnerator)。透過解析與整理可去除冗長語句（即動詞、連接詞、備註等）同時解析語句長度是否符合限制規定，直到符合規定則擷取。其考量因素包含：字詞出現位置、字詞出現位置與開頭之距離、語意偏差，透過模型訓練進行語句壓縮與過濾最後擷取最符合之標題。**Lin 和 Liang (2008)** 表示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，待時間過去讀者難以追朔該新聞事件。因而提出將事件主軸納入摘要建立條件機制（Story-line based Topic Retrospection, SToRe），透過事件主軸能使讀者更了解事件發展與概念。該研究乃先界定事件再由事件建立議題，透過議題發展出脈絡並去除較無相關性事件，從中再擷取代表性語句並以議題主軸為依據形成摘要。該研究所提出之機制能提供更了解議題發展，同時亦可作為索引提高讀者閱讀效率。

另外，針對資料量過大之議題或是具有相似主題文件，**Ye 等人 (2007)** 提出一套以文件概念格（Document Concept Lattice, DCL）方式建立多文件摘要。該研究乃將相關主題、議題之對應語句於概念格中，從中挑取最具概念之語句形成摘要。該研究運用詞彙鏈以及反詞頻方式刪除冗長、不必要之詞彙，並將具有意義之詞彙形成一個有方向但非循環之概念圖。此外，**Kuperberg 等人 (2006)** 表示許多文章乃增加描述語句（即無相關語句）或是銜接語句（即中度相關語句）以強調閱讀重點（即高度相關語句），但過多描述反延長閱讀時間。因此，為了加強語句順暢性，利用功能性磁共振造影（Functional Magnetic Resonance Imaging, fMRI）研究於事件之因果關係提出以簡短訊息（即兩句背景以及一句關鍵語句）確立與了解主要訊息。**Sweeney 等人 (2008)** 表示由於自動化文件摘要多為壓縮文件內容而擷取重要片段，但此方式易影響原始文件之正確解釋。因此，提出以新語句方式建立摘要，並透過延伸閱讀方式（Show Me More）給予更多資訊以完整表達文件。該研究先以搜尋者查詢為需求條件建立摘要（Query-biased Summarisation Methods）並基於文件結構（標題、開頭）、詞頻和查詢條件等條件進行得分計算，再依據得分排序形成摘要順序；此外基於類似文件即去除重複性語句並形成摘要。**Ko 和 Seo (2008)** 考量非所有文件皆以特定格式（如標題與位置）撰寫於此影響文件摘要之形成，因而針對無標題樣式文件進行摘要，該研究提出上下文資訊以及混合統計方式，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，再以詞頻計算方式（TF-based Query Method）篩選出關鍵字詞，將關鍵字詞作為語句相似度衡量依據，以取得出最重要、最具代表性之語句以形成摘要。



綜合上述，針對摘要應用面無論於專業知識領域或是即時網際網路等領域，皆有許多相關文獻研究並依其不同應用與技術分別彙整成表，如表 2.3 所示。

表 2.3、摘要應用領域文獻彙整表

應用領域	相關研究	技術名稱	目的與結果
特定專業知識領域	Elhadad 等人 (2005)	統一摘要模式	以主題樹概念將主題作為連結節點，先將文件進行主題分群再根據檢索主題找出相關文件，再將相關文件進行多文件摘要以形成搜尋總結。
	Ling 等人 (2007)	基因學角度為基之半結構化摘要	經由字典檢索、過濾同義詞方式訓練模型取出文件關鍵字並作為目標基因，透過目標基因作為語句擷取條件，再以空間向量模型、機率語言模型訓練模型於語意部分之評分，最後乃擷取語句並形成摘要。
	Zhou 等人 (2006)	多文件摘要	透過解析與整理並基於隱藏式馬可夫鏈模型之改良模型，以解析與整理可去除冗長語句，同時解析語句長度是否符合限制規定，直到符合規定則擷取。
	Moens (2007)	自動化個案摘要與檢索	將法院判決文件劃分為：(1)基本要素（包含被告、原告、案由、投訴類型等）、(2)背景、(3)相關法律問題、(4)辯護論點、(5)裁決與處置，根據上述畫分作為結構摘要關鍵字擷取依據。
	Moens 等人 (2005)	以層級化方式	將文件層級化分割建立主題樹，透過主題樹取得語句以及對應語句以形成摘要。
	Xie 和 Liu (2010)	特定會議記錄摘要	以迴歸監督式學習方式統計分類擷取，透過支援向量機監督學習方式標記並分類具有代表性文句。
線上即時網際網路領域	Bouras 等人 (2008)	個人化新聞索引系統 (PeRSSonal)	將網路新聞、文章自動化摘要並透過分類以產生文字標籤，結合網路新聞 RSS 訂閱服務即可進行個人化服務。
	Yang 和 Wang (2007)	手持行動裝置之文件進行自動化摘要	利用碎形理論形成摘要模型，並以樹狀結構或是分層方式加強手持行動裝置視覺化操作。
	Zajic 等人 (2007)	--	為克服新聞標題文字長度之限制，運用文件壓縮技術將擷取單一文件之摘要語句，以形成候選語句並彙整成多文件摘要。
	Lin 和 Liang (2008)	以使用者為導向之文件摘要檢索	根據文件之關聯性與共同特徵，利用共同特徵作為文件篩選條件並進行群集分析。
	Ye 等人 (2007)	文件概念格方式建立多文件摘要。	該研究乃將相關主題、議題之對應語句於概念格中，從中挑取最具概念之語句形成摘要。
	Kuperberg 等人 (2006)	基於因果關係建立摘要	透過 fMRI 技術取得語句相關程度，當中再根據合理性將中度相關語句連結高度相關語句以構成因果關係。
	Sweeney 等人 (2008)	延伸閱讀摘要	以搜尋者查詢為需求條件建立摘要並基於文件結構、詞頻和查詢條件等條件形成摘要。
	Ko 和 Seo (2008)	上下文資訊以及混合統計方式建立摘要	透過雙連搜尋技術將文件之兩個連續語句結合成一個虛擬語句，以詞頻計算方式篩選最重要、最具代表性之語句以形成摘要。

### 2.3.2 摘要技術分析

於摘要建立技術部分，過去研究將摘要建立技術分為「監督式機器學習」方式與「非監督式機器學習」方式等兩種技術進行探討。

#### (A) 監督式學習

針對監督式機器學習可細分為各種技術，如詞彙鏈、支援向量機或主題樹等技術。透過詞彙進行標記以形成詞彙鏈，但語意排序正確性乃影響摘要之可讀性，若語意排序錯誤可能造成閱讀混淆或是形成錯誤觀念。因此，**Bollegala 等人 (2010)** 提出一套由下而上方式排序語句以結合兩份文件摘要技術，該研究以年代、主題適配度、繼承、優先等級作為擷取標準，並以監督式訓練模型計算語句間之方向性與強度排序，同時結合支援向量機 (Support Vector Machine, SVM)，以進行分類即可群集兩份文件之語句與排序並擷取形成摘要。**Jung 等人 (2005)** 透過假設語句結合統計計算方式以提出一套自動化摘要模式，藉由雙向連詞 (Bi-gram) 結合兩兩相鄰語句以建立假設語句，並以假設語句為基計算該語句於文件之重要性 (即標題、位置等準則) 以刪除不必要語句，最後再將假設語句分離以進行群聚分析取得相似性，取得更具代表性語句已形成摘要。**Benedí 和 Sánchez (2005)** 將 SCFGs 技術結合 N 連詞提出一套結合語言模型，透過 N 連詞擷取詞彙關聯性，再以隨機語法技術進行關聯性分類，以此方式即可降低詞彙關聯複雜性進而提高 SCFGs 之運算。**Li 和 Chen (2010)** 利用統計語言模型先進行判斷文件相關性，再以機率排序和潛在馬可夫鏈模式擷取出具代表性之片段。藉由此方式能精確取得文字片段開頭與結尾以提高語意之協調性，亦可針對使用者需求應用於文件自動化摘要。**Oliva 等人 (2011)** 針對簡短文件提出一套基於結構 (即名詞、形容詞等) 之語意相似度系統 (Syntax-based Measure For Semantic Similarity, SyMSS)，透過 WordNet 概念模擬人類共同詞彙以強化語意合理性與相似性，同時利用註解測量 (Gloss-based Measures) 方式比較字詞語意、相似性與詞性，並以加總方式計算語句之間相似性。**Ouyang 等人 (2011)** 建立一套以特徵為基擷取模型，其方式乃預先定義七個篩選條件 (根據語意、與主題之相關性、字詞出現頻率、是否具有結尾語意以及語句存在位置) 並作為評分準則，透過支援向量迴歸模型 (SVR) 學習方式將具有相同得分條件之語句關聯群集以完成分類，從中進行二元決策方法篩選語句，根據篩選即可留下最具價值、重要性語句之最佳組合。針對多文件摘要技術大多針對需求者之要求、主題建立，進而要求自動化摘要能於有限字數內表達涵義。是故，**Vanderwende 等人 (2007)** 以多文件

摘要系統 SumBasic 為基提出一套多文件摘要總結系統，透過擷取、語句簡化、和詞彙擴充方式並以任務為目的進行多文件主題摘要。基於 SumBasic 系統主要透過詞頻並考量人類手動摘要方式提出多文件摘要，於此該研究加入語句壓縮以及加強同義詞等詞彙關係以擷取重要內容，並以簡化方式滿足需求者，經由此系統能將主題詞彙擴充以提高相關性文件之擷取。

## (B)非監督式學習

針對非監督式機器學習技術，可分為向量空間模型與潛在語意分析兩種。**Chan (2006)** 基於潛在語意分析提出一套擷取最具代表性語句之量化模式，透過潛在語意加強人類理解模式並將具有代表性或連結性之語彙組織成詞彙網路圖以表示文件語句間之關聯性，並以此為基建立模式，透過此模式能加強文件摘要生成之連續性，並以關聯方式加強語句相關性。**He 等人 (2009)** 提出一套自動化摘要評估系統，為減少篩選時間藉由潛在語意分析 (Latent Semantic Analysis, LSA) 以矩陣方式取得文字出現頻率並以餘弦定理確定相似性，同時結合 N 連詞共生 (N-gram Co-occurrence) 以評估、篩選作品，再以機器翻譯排名系統 (BLEU Algorithm) 機器學習方式標記作品並藉由匹配訓練以提高精確度，透過 LSA 以及 N 連詞共生方式乃優於自動彙整評估系統 (Recall-Oriented Understudy for Gisting Evaluation, ROUGE) 評估更能加強輔佐以提高篩選分級效率。**Steinberger 等人 (2007)** 以潛在語意分析 (Latent Semantic Analysis, LSA) 為基建立自動化摘要系統，並文件壓縮百分比方式提出摘要字數之限制。該研究以 TERM-BASED 方法找出具重要性詞彙並延伸文件之詞彙關係，同時結合潛在語意分析 (Latent Semantic Analysis, LSA) 以自動擷取出相關語句生成摘要，以此方式能提高多文件摘要之彙整性亦提高摘要之一致性。針對未涉獵之領域文件資料，文件摘要多數需以人工手動標記方式形成有系統性之規則，且非常費力與耗時。因此，**Nomoto 和 Matsumoto (2001)** 提出一套非監督式多樣性多文件摘要技術，首先利用群聚演算 (K-means) 並以最小距離 (Minimum Description Length Principle, MDLP) 於各領域中尋找相關主題文件，再以 Zechner 所提出之加權模型 (Z-model) 去除多餘語句同時亦擷取最具代表性語句以自動形成多樣性多文件摘要，以此方式不僅能取代人力亦能應用於檢索技術，透過檢索結果形成多文件摘要以加強篩選。**Yeh 等人 (2005)** 分別為以「改良語彙庫為基技術」(Modified Corpus-based Approach, MCBA) 與「潛在語意分析 LSA 為基結合 TRM 技術」，以建立文件摘要。當中，MCBA 方式乃以特徵評分 (即位置、

正向關鍵字、排除負向關鍵字、向心性以及與標題相似性等），計算結合基因演算法（Genetic Algorithm, GA），以擷取得最佳組合形成摘要；第二種技術則透過潛在語意分析（LSA）則取得文件或語彙庫之語意矩陣，並建立語意關聯圖（Text Relationship Map, TRM），以取得顯著語意並形成摘要。Ko 等人（2003）利用詞彙群聚提出一套以主題為基之文件摘要技術，首先將文件進行上下關聯分析並以空間向量呈現，從中取得詞彙群集並確定群集核心以產生主題和關鍵字，最後則以壓縮比率作為之擷取語句之固定數量形成摘要。此外，Yangarber（2003）以特定名詞（人、位置、組織及日期等）進行分類並解析文件結構（主詞、動詞與受詞）提出一套非監督式學習語意模式。首先以權重配置方式取代二元決策進行相關性匹配，即可以權重準確取決具有相關性候選文件，透過此方式即可自然延伸出停止學習之終點，以提高非監督式學習之精確性。

綜述而論，針對摘要建立之機器學習技術可分為監督式以及非監督式等技術，其各技術皆有許多相關文獻研究與探討，依其不同技術分別彙整成表，如表 2.4 所示。

表 2.4、摘要建立文獻彙整表

技術部分	相關研究	技術名稱	目的與結果
監督式學習	Bollegala 等人 (2010)	支援向量機	以年代、主題適配度、繼承、優先等級作為擷取標準,並以監督式訓練模型計算語句間之方向性與強度排序,並結合支援向量機,以進行分類即可群集兩份文件之語句與排序並擷取形成摘要。
	Jung 等人 (2005)	雙向連詞、統計計算	藉由兩兩相鄰語句以建立假設語句,並以假設語句為基計算該語句於文件之重要性)以刪除不必要語句,並群聚分析取得相似性,取得更具代表性語句已形成摘要。
	Benedi 和 Sánchez (2005)	結合語言模型	透過 N 連詞擷取詞彙關聯性,再以隨機語法技術進行關聯性分類即可降低詞彙關聯複雜性進而提高 SCFGs 之運算。
	Li 和 Chen (2010)	統計語言模型、馬可夫鏈	擷取出具代表性之片段。藉由此方式能精確取得文字片段開頭與結尾以提高語意之協調性,亦可針對使用者需求應用於文件自動化摘要。
	Oliva 等人 (2011)	語意註解測量	比較字詞語意、相似性與詞性,並以加總方式計算語句之間相似性。
	Ouyang 等人 (2011)	支援向量迴歸模型	將文章之語句進行分段,再提取該文章之第一列句子,並以文句相似度進行群組分析同時刪除重複出現之語句,透過相似性得分計算擷取出該群組中最具代表性文句,再與其他群組文句組合以提高可讀性。
	Vanderwende 等人 (2007)	詞彙擴充	透過詞頻並考量人類手動摘要方式提出多文件摘要,於此該研究加入語句壓縮以及加強同義詞等詞彙關係以擷取重要內容,並以簡化方式滿足需求者。
	Ko 和 Seo (2008)	雙連搜尋技術	透過雙連搜尋技術將文件之兩個連續語句結合成一個虛擬語句,再以詞頻計算方式篩選出關鍵字詞,將關鍵字詞作為語句相似度衡量依據,以取得出最重要、最具代表性之語句。
非監督式學習	Chan (2006)	潛在語意分析	將具有代表性或連結性之語彙組織成詞彙網路圖以表示文件語句間之關聯性,並作為基建立模式,透過此模式能加強文件摘要生成之連續性。
	He 等人 (2009)	潛在語意分析、N 連詞	評估、篩選作品,再以機器翻譯排名系統 (BLEU Algorithm) 機器學習方式標記作品並藉由匹配訓練以提高精確度,
	Steinberger 等人 (2007)	潛在語意分析	找出具重要性詞彙並延伸文件之詞彙關係,同時結合潛在語意分析以自動擷取出相關語句生成摘要。
	Nomoto 和 Matsumoto (2001)	非監督式多樣性多文件摘要技術	以最小距離尋找相關主題文件,再以加權模型 (Z-model) 去除多餘語句同時亦擷取最具代表性語句以自動形成多樣性多文件摘要。
	Yeh 等人 (2005)	潛在語意分析、基因演算法	以特徵評分 (即位置、正向關鍵字、排除負向關鍵字、向心性以及與標題相似性等),計算結合基因演算法以擷取得最佳組合形成摘要。
	Ko 等人 (2003)	上下關聯分析、空間向量	將文件進行上下關聯分析並以空間向量呈現,從中取得詞彙群集並確定群集核心以產生主題和關鍵字,最後則以壓縮比率作為之擷取語句之固定數量形成摘要。
	Yangarber (2003)	非監督式學習語意模式	以特定名詞進行分類並解析文件結構,再以權重配置方式取代二元決策進行相關性匹配,

## 2.4 小結

綜上所述，本研究之主要目的乃以知識文件為基礎，以進行專業語彙語一般詞彙之語意關聯性以及知識文件結構化摘要推論，本研究藉由此二大方向進行文獻探討，因此，由2.2節之文獻回顧可知，過去「問答應用與技術」之相關研究可發現針對問答系統之可以文件主題、文件特徵或是文件語意作為主要分析因素並等部分進行探勘，以透過問項類型，以進行問題之解答。由2.3節之文獻回顧可知，過去「摘要應用與技術」之相關研究，可得知過去針對摘要之技術大多乃應用於資料量過大或是具有相似主題之文件，此外根據摘要建立之技術可以機器學習方式或是根據文章內容與結構評估等方式。

整體而言，有別於先前之研究，本研究藉由2.2節之文獻回顧所探討「問答應用與技術」議題，可得知文件主題、特徵以及語意作為分析因素，並以此為依據建構「知識文件表達項目解析模組」進行表達項目分析與觀念性語句擷取，以及建構「知識文件問答解析模組」進行主要詢問詞分析，並透過詢問詞與回覆詞配對解析找出具有關聯性之語意詞，進而取出對應知識文件，接著藉由2.3節之文獻回顧所探討「摘要應用與技術」議題，所得知摘要可應用於知識領域以輔佐知識文件之呈現，此外針對摘要呈現方式不同進而影響搜尋者閱讀觀感。因此本研究乃建構「知識文件結構化摘要推論模組」進行知識文件摘要建立，對於知識文件呈現方式，本研究乃以簡略結構化摘要描述文件之主要動機與目的，另以詳述結構化摘要描述文件之細部內容。最後知識搜尋者則有效縮短知識搜尋者搜尋文件之時間，以及提高知識分享之成效。

### 第三章、結合語意解析與文件摘要技術之知識文件檢索模式

本研究所提出之「結合語意解析與文件摘要技術之知識文件檢索模式」乃以勞工安全衛生知識網所提供之技術叢書與研究報告為分析基礎，先行解析知識文件之結構特性以提出八個表達項目、二十八個表達細項特性同時建立「知識語彙庫」，並以知識語彙庫為基建立語彙法則，再將其文件與語彙法則進行比對進而取得觀念性語句與文件之關鍵字詞，並配置於各自歸屬語彙集合中，以為後續問答解析與結構化摘要等模組之用。透過「知識語彙庫」即可取得專有名詞並針對搜尋字串進行語意關聯，透過搜尋字串隱含目的解析並找出具有關聯或是相同目的之專有名詞，進而取出對應知識文件以提高檢索準確性；此外亦可根據結構化摘要之法則從語彙集合中擷取出具有代表性語句以建立摘要，即形成一份制式知識文件以釐清與完整表達報告內容。因此本研究之主要流程可分為三大部份，分別為如圖 3.1 之 Part1「知識文件表達項目解析模組」、Part2「知識文件問答解析模組」、Part3「知識文件結構化摘要模組」所示。

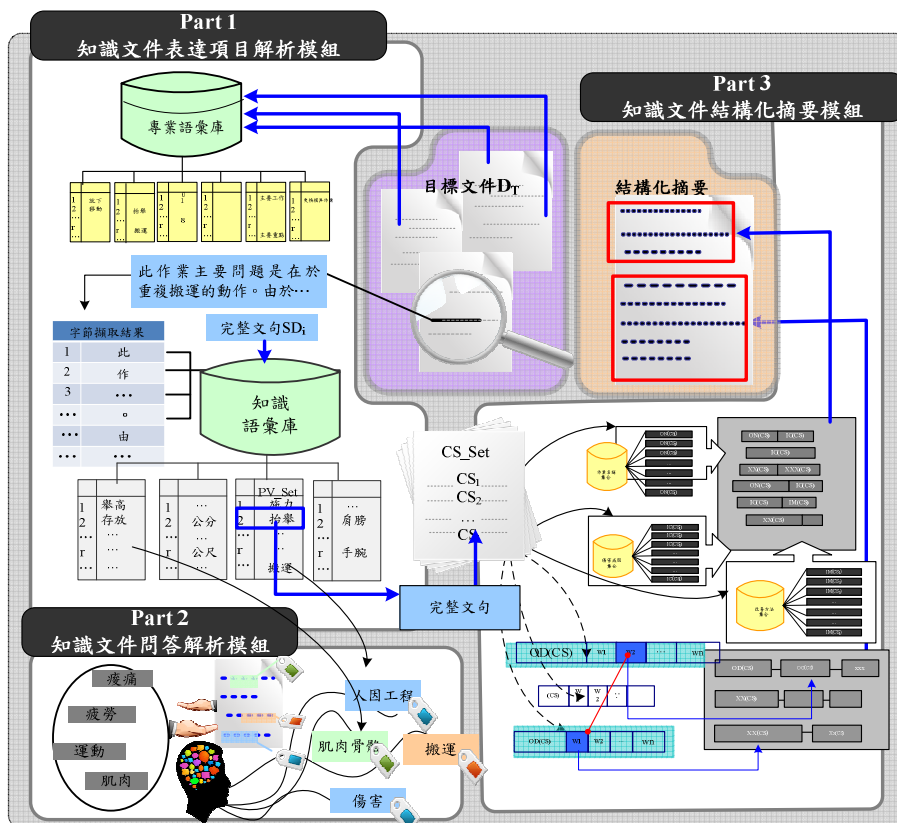


圖 3.1、結合語意解析與文件摘要技術之知識文件檢索方法論之流程架構

### 3.1 知識文件表達項目解析模組

由於人因工程工作場所研究報告皆由專業人員進行訪視後撰寫，多以專業名詞、專業知識描述報告，為協助無背景知識之使用者進行搜尋。因此，針對表達項目建立乃參照人因工程改善報告與技術叢書內文，如於技術叢書「人因工程危害預防技術」提到造成肌肉骨骼傷害之五個主要成因為過度施力、高重複動作與震動等因素，本研究與人因工程相關人員進行討論，並於討論後將改善報告與技術叢書內文中重複出現或重要之描述詞彙彙整，再進行分項以建立知識文件之表達項目。待本研究解析知識文件並建立表達項目與觀念性語句擷取，以為後續問答解析與結構化摘要推論之用（如圖 3.2 所示）。針對各不同表達項目與表達細項之細部說明如下：

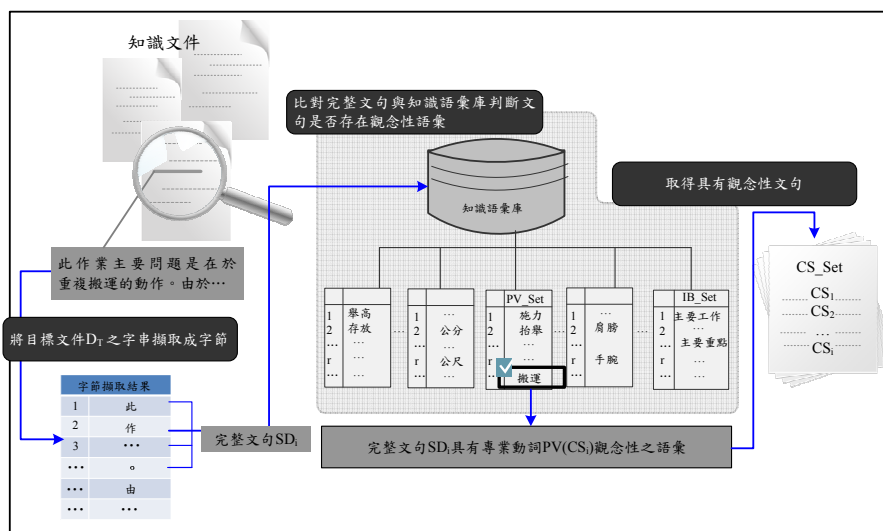


圖 3.2、知識文件表達項目解析與觀念性語句擷取示意圖

#### 3.1.1 建立知識文件之表達項目

針對人因工程工作場所研究報告文件之內容解析，可分為八個表達項目、十九個表達細項，為加強語句順暢性，本研究另外增加九個表達細項進行語句輔助。故「知識語彙庫」共計包含二十八個語彙集合，其二十八個語彙集合如表 3.1 與表 3.2 所示，並定義符號如下：

表 3.1、表達細項與對應內容及表達方式(1)

表達項目	表達細項	表達方式
作業領域	作業領域	表達行業類別之集合；其內容包含「農林漁牧業」、「礦業及土石採取業」、「食品製造業」、「紡織成衣業」等。
作業名稱	作業名稱	表達作業在此行為之名稱集合；其內容包含「更換模具做業」、「納箱作業」、「搬運作業」。



表 3.1、表達細項與對應內容及表達方式(2)

表達項目	表達細項	表達方式
作業身分	作業員性別	形容該作業之作業員性別語彙之集合；其內容包含「男性」「女性」或是「男女不拘」。
	作業員年齡	形容該作業之作業員年齡範圍之語彙，如「20~30 歲」；其內容包含「中年」、「青年」、「無年齡限制」等。
	作業員職稱	為形容該作業之作業員職位之語彙，其內容包含「護士」、「技術師」、「作業員」等。
作業環境	設備語彙	為描述該作業所使用之設備語彙之集合，如「採血床」、「保溫櫃」、「熱鍛機台」等。
	設備佈置	即設備擺放與佈置語彙之集合，如「輸送帶的高度為 75 公分」。
	工具介紹	為描述作業或是改善流程中所使用之工具語彙集合；其內容包含「蝴蝶籠」、「手臂支撐靠墊」、「升降台車」、「手推車」等。
作業行為	作業目的	為作業目標對應作業名稱語彙之集合；其內容包含「主要工作」、「主要功能」、「主要重點」等。
	作業描述	為對應作業名稱語彙、作業目的語彙、作業工具語彙之集合。
	專業動詞	為描述作業員對應作業描述語彙之姿勢集合；其內容包含「站姿」、「彎腰」、「低頭」、「施力」、「抬舉」。
作業時間	作業次數/天	為描述作業員在此作業所重複動作之次數；其內容包含「一天需要」、「一天必須」等。
	作業時間/次	為描述作業員在此作業所花費時間之集合；其內容包含「一次所花時間」、「一次需要」、「一次必須」等。
	作業距離/次	為描述作業員在此作業所需距離之集合；其內容包含「距離」、「最近距離」、等。
傷害成因	傷害因素	為描述傷害成因之集合；其內容包含「過度施力」、「高重複性動作」、「震動」、「低溫」、「不良工作姿勢」。
	傷痛部位	表示作業員對應作業描述語彙之姿勢部位及傷痛部位之集合；其內容包含「頭頸」、「軀幹」、「手部」、「手腕」、「腿部」。
改善方法	改善目的	表明對應傷害成因語彙之改善目的之集合；其內容包含「主要改善」、「有效改善」、「大程度地」、「明顯降低」等。
	改善流程	描述流程改善語彙之集合；其內容包含「考慮」、「利用」、「建議」、「只要」等。
	改善評估	解釋其作業改善後評估語彙之集合；其內容包含「行動水準」、「評級點數」、「檢核總分」、「風險等級」、「負荷明顯降低」等。

表 3.2、輔助表達細項與對應內容及表達方式

表達細項	表達方式
連結語彙	其內容包含「與」、「和」、「但是」、「以」。
一般性動詞	一般性動詞之語彙集合；其內容包含「是」、「主要是」、「為」、「舉高」、「放下」、「移動」、「存放」。
數字語彙	記錄「0」至「9」及其所組合之語彙。
金錢單位語彙	包含「元」、「台幣」等。
年齡單位語彙	包含「歲」。
長度單位語彙	包含「距離」、「公分」、「公尺」、「長」等。
時間單位語彙	包含「時間」、「分鐘」、「小時」等。
重量單位語彙	包含「公斤」、「公克」、「噸」等。
頻率單位語彙	包含「次」、「下」等。

## 符號定義

$D_T$	目標文件
$N(DT)$	目標文件DT所包含文句之個數
$SD_i$	文件之第i個語句
CS	具有人因工程工作場所改善報告觀念語文句
OF(CS)	具作業領域觀念之語彙
ON(CS)	具作業名稱觀念之語彙
ORS(CS)	具作業員性別觀念之語彙
ORA(CS)	具作業員年齡觀念之語彙
ORT(CS)	具作業員職稱觀念之語彙
F(CS)	具設備觀念之語彙
FL(CS)	具佈置觀念之語彙
OT(CS)	具作業工具觀念之語彙
OG(CS)	具作業目的觀念之語彙
OD(CS)	具作業描述觀念之語彙
OFQ(CS)	具作業頻率觀念之語彙
OH(CS)	具作業時間觀念之語彙
ODT(CS)	具作業距離觀念之語彙
IC(CS)	具傷害成因觀念之語彙
B(CS)	具作業部位觀念之語彙
IG(CS)	具改善目的觀念之語彙
IR(CS)	具改善流程觀念之語彙
R(CS)	具評估觀念語彙之語彙
C(CS)	具連結語彙
GV(CS)	具一般動詞語彙
PV(CS)	具專業動詞觀念之語彙
N(CS)	具數字語彙
MU(CS)	具金錢單位語彙
AU(CS)	具年齡單位語彙
LU(CS)	具長度單位語彙

TU(CS)	具時間單位語彙
WU(CS)	具重量單位語彙
FU(CS)	具頻率單位語彙
OF_Set	具作業領域觀念文句之集合
ON_Set	具作業名稱觀念文句之集合
OR_Set	具作業身分觀念文句之集合
OE_Set	具作業環境觀念文句之集合
OV_Set	具作業行為觀念文句之集合
OH_Set	具作業時間觀念文句之集合
IC_Set	具傷害成因觀念文句之集合
IM_Set	具改善方法觀念文句之集合

### 3.1.2 觀念性語句擷取

先將目標文件 $D_T$ 進行字節斷句方式取得完整語句 $SD_i$ ，並以領域專家所建立之「知識語彙庫」為基增加語彙法則，再將其完整語句 $SD_i$ 與語彙法則進行觀念性語彙比對，進而取得觀念性語句並配置於各自歸屬語彙集合中。如公式(3.1)所示，首先以字節擷取方式取得完整語句 $SD_i$ 。

$$SD_i = \{SD_{i,1}, SD_{i,2}, SD_{i,3} \dots, SD_{i,j}, \dots\} \quad (3.1)$$

為更具完整性與精確性，本研究乃增加語彙法則以明確定義觀念性語句之結構。若該完整語句內含語彙法則中之詞彙，則歸納至所屬之表達細項。因此本研究針對各語彙法則建立八種篩選規則，以取得各語彙所代表之語句，其詳細說明與定義語彙法則之表達方式如下：

1. 作業領域語彙法則 (R\_OF)：作業領域語彙法則乃表達行業之類別，故法則如公式(3.2)所示。若完整語句存在於作業領域觀念語彙中，即表示該完整語句  $SD_i$  為作業領域之觀念語句 OF\_Set。

$$\text{IF } SD_{i,j} \text{ exist in OF(CS)} \forall j \text{ Then } SD_i \in \text{OF\_Set} \quad (3.2)$$

2. 作業名稱語彙法則 (R\_ON)：作業名稱語彙法則表達此行為之名稱，故法則如公式(3.3)所示。若完整語句存在於作業名稱觀念語彙中，即表示該完整語句 SD<sub>i</sub> 為作業名稱之觀念語句 ON\_Set。

$$\begin{aligned} & \text{IF } SD_{i,j} \text{ exist in ON(CS)} \forall j \\ & \text{Then } SD_i \in \text{ON\_Set} \end{aligned} \quad (3.3)$$

3. 作業身分語彙法則 (R\_OR)：在此階段將設置寬鬆法則以及嚴謹法則乃確保擷取無誤。其寬鬆法則與嚴謹法則之規則定義如下：

✓ 寬鬆法則：能以一個至兩個語彙即能表示作業身分之觀念，如公式(3.4)至(3.5)所示。

$$\begin{aligned} & \text{IF } SD_{i,j} \text{ exist in ORT(CS)} \forall j \\ & \text{Then } SD_i \in \text{OR\_Set} \end{aligned} \quad (3.4)$$

$$\begin{aligned} & \text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{ORT(CS)} \\ \text{and ORA(CS)} \end{array} \right) \forall j \\ & \text{Then } SD_i \in \text{OR\_Set} \end{aligned} \quad (3.5)$$

$$\begin{aligned} & \text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{ORT(CS)} \\ \text{and ORS(CS)} \end{array} \right) \forall j \\ & \text{Then } SD_i \in \text{OR\_Set} \end{aligned} \quad (3.6)$$

✓ 嚴謹法則：以多個語彙形成嚴謹結構以表達其作業身分之觀念，如公式(3.7)所示。

$$\begin{aligned} & \text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{ORT(CS) and N(CS)} \\ \text{and AU(CS)} \end{array} \right) \forall j \\ & \text{Then } SD_i \in \text{OR\_Set} \end{aligned} \quad (3.7)$$

4. 作業環境語彙法則 (R\_OE)：根據公式(3.8)至公式(3.9)所示，對於作業環境之描述須以設備語彙 F(CS) 結合設備佈置語彙 FL(CS)，同時結合數字語彙 N(CS) 以及長度單位語彙 LU(CS)、乃明確表達作業設備與工具之規格。

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{F(CS) and FL(CS)} \\ \text{and N(CS) and LU(CS)} \end{array} \right) \forall j \quad (3.8)$$

Then  $SD_i \in OE\_Set$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OT(CS) and N(CS)} \\ \text{and LU(CS)} \end{array} \right) \forall j \quad (3.9)$$

Then  $SD_i \in OE\_Set$

5. 作業行為語彙法則 (R\_OV)：根據公式(3.10)至公式(3.12)所示，對於作業目的之表達描述須以作業目的語彙 OG(CS)結合一般動詞語彙 GV(CS)、作業工具語彙 OT(CS)以及專業動詞語彙 PV(CS)。

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OG(CS)} \\ \text{and GV(CS)} \end{array} \right) \forall j \quad (3.10)$$

Then  $SD_i \in OV\_Set$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OG(CS) and GV(CS)} \\ \text{and PV(CS)} \end{array} \right) \forall j \quad (3.11)$$

Then  $SD_i \in OV\_Set$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{ON(CS) and GV(CS)} \\ \text{and PV(CS) and OT(CS)} \end{array} \right) \forall j \quad (3.12)$$

Then  $SD_i \in OV\_Set$

6. 作業時間語彙法則 (R\_OH)：作業時間語彙法則即表示作業之頻率與所花時間。其描述方式包含作業頻率語彙 OFQ(CS)、作業時間語彙 OH(CS)、作業距離語彙 ODT(CS)等。透過公式(3.13)、公式(3.14)、公式(3.15)篩選並將其符合之語句列為作業時間語彙之觀念語句 OH\_Set。

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OFQ(CS) and PV(CS)} \\ \text{and N(CS)} \end{array} \right) \forall j \quad (3.13)$$

Then  $SD_i \in OH\_Set$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OH(CS) and N(CS)} \\ \text{and FU(CS)} \end{array} \right) \forall j \quad (3.14)$$

Then  $SD_i \in \text{OH\_Set}$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{OT(CS) and ODT(CS)} \\ \text{and N(CS) and LU(CS)} \end{array} \right) \forall j \quad (3.15)$$

Then  $SD_i \in \text{OH\_Set}$

7. 傷害成因語彙法則 (R\_IC)：傷害成因語彙法則乃表達作業所造成之傷害，其表達方式包含傷害成因語彙以及部位語彙，根據公式(3.16)找尋具有傷害成因語彙之觀念語句 IC\_Set。

$$\text{IF } SD_{i,j} \text{ exist in } (\text{IC(CS) and B(CS)}) \forall j \quad (3.16)$$

Then  $SD_i \in \text{IC\_Set}$

8. 改善方法語彙法則 (R\_IM)：根據公式(3.17)至公式(3.20)所示，改善方法語彙法則表達方式包含改善目的 IG(CS)、改善流程 GV(CS)、改善評估等 IR(CS)，並結合一般動詞 GV(CS)、專業動詞語彙 PV(CS)以及作業工具語彙 OT(CS)表達之。

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{IG(CS) and GV(CS)} \\ \text{and PV(CS)} \end{array} \right) \forall j \quad (3.17)$$

Then  $SD_i \in \text{IM\_Set}$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{IR(CS) and GV(CS)} \\ \text{and OT(CS)} \end{array} \right) \forall j \quad (3.18)$$

Then  $SD_i \in \text{IM\_Set}$

$$\text{IF } SD_{i,j} \text{ exist in } R(\text{CS}) \forall j \quad (3.19)$$

Then  $SD_i \in \text{IM\_Set}$

$$\text{IF } SD_{i,j} \text{ exist in } \left( \begin{array}{l} \text{RV(CS) and ORT(CS)} \\ \text{and PV(CS)} \end{array} \right) \forall j \quad (3.20)$$

Then  $SD_i \in \text{IM\_Set}$

在此模組可得八種觀念語句之集合，分別為作業領域、作業名稱、作業身分、作業環境、作業行為、作業時間、傷害成因以及改善方法等。透過知識文件表達項目解析階段乃擷取知識文件表達項目之觀念性語句，即為後續問答解析與結構化摘要之應用。

### 3.2 問答解析模組

由於知識搜尋者所輸入之搜尋字串，大多屬知識搜尋者以個人直覺且不屬於知識文件專業語彙之詢問詞。針對傳統關鍵字檢索技術，若以專業語彙搜尋即可找到所屬相關知識文件；反之，自行定義之詢問詞可能因不具有明確定義，則有找出非所屬相關文件。因此，為加強自然語言搜尋彈性，本研究提出知識文件問答解析模組，針對知識搜尋者之搜尋字串進行主要詢問詞分析，並透過詢問詞與回覆詞配對解析找出具有關聯性之語意詞（如圖 3.3 所示），進而取出對應知識文件以提高檢索準確性。於此階段所使用符號定義如下：

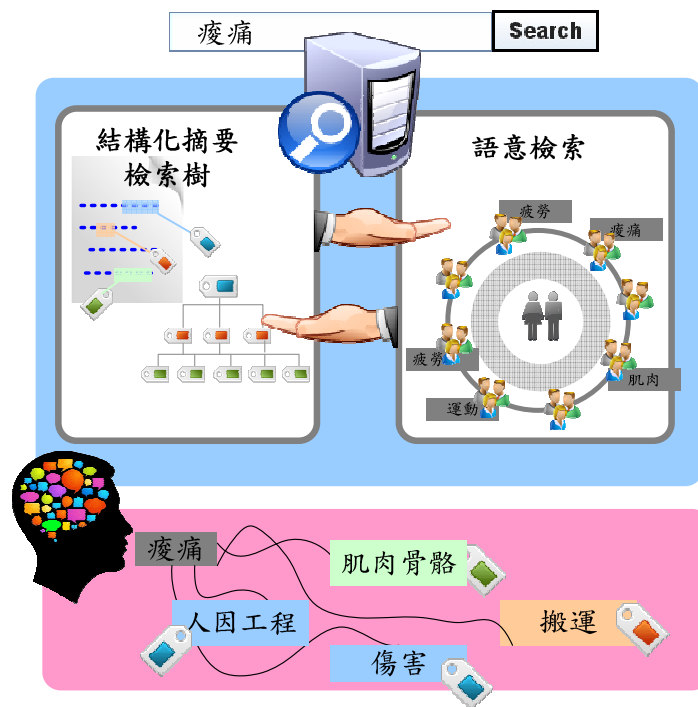


圖 3.3、問答解析示意圖

## 符號定義

$QW_i$	第 $i$ 個詢問句
$QW_{i,j}$	第 $i$ 個詢問句中第 $j$ 個詞彙，以下稱為簡稱詢問詞
$AW_k$	第 $i$ 個回覆句
$AW_{k,m}$	第 $k$ 個回覆句中第 $m$ 個詞彙，以下簡稱回覆詞
$C(QW_{i,j})$	第 $i$ 個詢問句中出現第 $j$ 個詢問詞之次數
$C(AW_{k,m})$	第 $k$ 個回覆句中出現第 $m$ 個回覆詞之次數
$P(AW_{k,m} QW_{i,j})$	第 $k$ 個回覆句之第 $m$ 個回覆詞出現於第 $i$ 個詢問句第 $j$ 個詢問詞之機率值
$QWAW_k$	第 $k$ 個詢問詞與回覆句配對組合，即回覆集合
$D_{i,q}$	第 $i$ 份文件包含 $q$ 個關鍵字詞
$D^o_q$	第 $q$ 個文件關鍵字集合向量
$QWAW^o_k$	第 $k$ 個回覆集合向量
$Sim(D_q QWAW_k)$	第 $q$ 個文件與第 $k$ 個回覆集合之相似值
$\omega(D_q, QWAW_k)$	文件與回覆集合之篩選門檻值（若回覆集合與文件相似值 $Sim(D_q QWAW_k)$ 大於門檻值即為候選文件
$Q$	整體語意類別之第三四分位數指標
$CandidateDoc\_Set_d$	第 $d$ 個具有專業語彙連結之候選文件集合

### 3.2.1 詢問詞隱含目的判斷

本步驟以詢問 (Question) 與回覆答案 (Answer) 方式從中取得詢問詞所隱含目的與相關詞，以訓練詢問詞與相關語意詞之相似機率。如公式(3.21)所示，當知識搜尋者輸入詢問句 ( $QW_i$ ) 經由斷詞取出詞性序列，即可從詢問句透過字詞拆解以擷取有效詢問詞 ( $QW_{i,j}$ )，根據勞工安全衛生研究所制定之雙語詞彙進而取得相關語意詞，即根據專有名詞「擦傷」可根據附註說明「覆蓋於身體的表面組織（如皮膚或黏膜）被擦掉或刮掉」，以取得相關語意詞「皮膚」或「黏膜」等，並藉由語意詞作為詢問詞以取得之回覆句 ( $AW_k$ ) 亦由多個回覆詞 ( $AW_{k,m}$ ) 所組成。藉由公式(3.22)計算詢問句 ( $QW_{i,j}$ ) 與回覆詞 ( $AW_{k,m}$ ) 之詞彙出現機率，以取得相關性詞彙並形成詢問詞集合並將各詢問



詞相對之回覆詞彙整於表 3.3。最後如公式(3.23)所示，藉由詢問詞相對於回覆詞語意門  
 檻值  $\omega(AW, QW)$  篩選與該回覆詞句有相關性之詢問詞以成集合。

$$\begin{aligned} QW_i &= \{QW_{i,1}, QW_{i,2}, QW_{i,3}, \dots, QW_{i,j}, \dots\} \\ AW_k &= \{AW_{k,1}, AW_{k,2}, AW_{k,3}, \dots, AW_{k,m}, \dots\} \end{aligned} \quad (3.21)$$

$$\begin{aligned} &P(AW_{k,m} | QW_{i,j}) \\ &= \frac{C(AW_{k,1} | QW_{i,1}) + C(AW_{k,1} | QW_{i,2}) \dots + C(AW_{k,1} | QW_{i,j})}{C(QW_{i,j})} \cdot C(AW_{k,1}) \end{aligned} \quad (3.22)$$

$$IF P(AW_{k,m}, QW_{i,j}) > \varpi(AW, QW) \text{ Then } AW_{k,m}, QW_{i,j} \in QWAW^{\varpi}_k \quad (3.23)$$

### 3.2.2 詞彙類別相似度判斷

本研究乃透過目標文件解析模組所提出之表達項目為基以建立文件關鍵字集  
 ( $D_{i,q}$ )，並如公式(3.24)將每個表達項目之表達語彙定義為專業語彙。本步驟以向量空  
 間模型 (Vector Space Model, VSM) 之餘弦函數 (Cosine) 計算該文件與回覆詞集合之  
 相似度，並以公式(3.25)判斷該回覆詞集合與文件之相似程度  $\text{Sim}(D_q | QWAW_k)$ ，若相似  
 度大於門檻值  $\omega$  且值越趨近於 1，即亦表示該群組越具有文件之解釋意義。

$$D_i = \{D_{i,1}, D_{i,2}, D_{i,3}, \dots, D_{i,q}, \dots\} \quad (3.24)$$

$$\begin{aligned} D_q^{\omega} &= [w_1, w_2, \dots, w_q]^T \\ QWAW^{\omega}_k &= [w_1, w_2, \dots, w_k]^T \\ \text{Sim}(D_q | QWAW_k) &= \frac{D_q^{\omega} \cdot QWAW_k^{\omega}}{|D_q^{\omega}| \cdot |QWAW_k^{\omega}|} \end{aligned} \quad (3.25)$$

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文件，其門檻  
 值  $\omega(D_q, QWAW_k)$  設定方式分別為平均值、中位數、四分位與門檻值直接定義等方式，  
 亦表示若通過門檻值  $\omega(D_q, QWAW_k)$  即表示專業語彙與該文件具有連結性，並將通過文  
 件放置候選文件集合 CandidateDoc\_Setd，其各門檻值設定方式如下：

(a)以平均值作為門檻值

如公式(3.26)所示，首先加總專業語彙比對次數，並除以總數求得比對平均值，進而將平均值作為門檻值  $\omega(D_q, QWAW_k)$ ，若相似度  $\text{Sim}(D_q | QWAW_k)$  低於整體平均值，即表示該專業語彙與文件不具連結，反之若高於整體平均值即表示該文件與專業語彙具有連結性，並放置候選文件集合以提供知識搜尋者。

$$\omega(D_q, QWAW_k) = \frac{\sum \text{Sim}(D_q | QWAW_k)}{N(CT_n)} \quad (3.26)$$

IF  $\text{Sum}(D_i) \geq \omega(CT, D_i)$  Then  $D_i \in \text{CondidateDoc\_Set}_d$

(b)以中位數作為門檻值

為避免比對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  $\omega(CT, D_i)$ 。如公式(3.27)所示，若專業語彙與文件比對加總  $\text{Sum}(D_i)$  高於（或相等）中位數，即表示該專業語彙與文件屬於高度關聯性，並放置候選文件集合以提供知識搜尋者。

$$\omega(CT, D_i) = \begin{cases} \frac{\text{Sum}(D_{\frac{i+1}{2}})}{2} & \text{IF } N(CT_n) \text{ is odd} \\ \frac{1}{2} \times \left( \frac{\text{Sum}(D_{\frac{i+1}{2}}) + \text{Sum}(D_{\frac{i+1}{2}+1})}{2} \right) & \text{IF } N(CT_n) \text{ is even} \end{cases} \quad (3.27)$$

and IF  $\text{Sum}(D_i) \geq \omega(CT, D_i)$  Then  $D_i \in \text{CondidateDoc\_Set}_d$

(c)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結果，於此乃制定整體筆數之四分位數作為門檻值  $\omega(CT, D_i)$ ，即以第三四分位數進行篩選。如公式(3.28)所示，首先計算第三四分位數指標  $Q$ ，以界定門檻值  $\omega(CT, D_i)$ ，進而挑選詞彙線性組合，當中，若專業語彙與文件比對加總  $\text{Sum}(D_i)$  大於四分位數則表示該專業語彙與文件屬於高度關聯性，並放置候選文件集合以提供知識搜尋者。

$$Q = N(CT_n) \times 75\%$$

$$\varpi(CT, D_i) = \begin{cases} \text{Sum}(D_{i \times 75\%}) & \text{IF } Q \notin \{X : |X| \in N\} \\ \frac{1}{2} \times (\text{Sum}(D_{i \times 75\%}) + \text{Sum}(D_{i \times 75\% + 1})) & \text{IF } Q \in \{X : |X| \in N\} \end{cases} \quad (3.28)$$

and IF  $\text{Sum}(D_i) \geq \varpi(CT, D_i)$  Then  $D_i \in \text{CondidateDoc\_Set}_d$

(d)直接定義門檻值

使用者亦可自行制訂門檻值，以進行篩選專業語彙與文件之關聯性。如公式(3.29)所示，若專業語彙與文件比對加總  $\text{Sum}(D_i)$  大於自行定義門檻值  $\varpi(CT, D_i)$  則表示該專業語彙與文件屬於高度關聯性，並放置候選文件集合以提供知識搜尋者。

$$\text{IF } \text{Sum}(D_i) \geq \varpi(CT, D_i) \quad \text{Then } D_i \in \text{CondidateDoc\_Set}_d \quad (3.29)$$

「問答解析」模組乃藉由詢問詞隱含目的判斷以取得該詢問詞之目的與相關回覆詞，再以回覆詞間接取得詢問詞與專業語彙語意關聯性。故當知識搜尋者輸入搜尋字串，此模組先行解析目標文件之主要詢問目的，在與文件相似性比對以取得相關文件。

### 3.3 結構化摘要模組

透過觀念性語句擷取階段本研究已得具有觀念性之語句，但考量摘要字數之限制以及摘要描述之完整性。在此階段仍需建立結構化摘要之法則，乃確保其觀念性語句具有該文件之代表性。考量知識搜尋者需求以及文章敘述完整性等因素，本研究乃將上述集合畫分為二部分，分別為簡略部分以及詳述部份。本研究乃期望藉由簡略結構化摘要與詳述結構化摘要之推論，即可協助使用者進行文件篩選。使用者可先行透過簡略結構化摘要快速瀏覽以得之文件之動機與目的等重點描述，若對該份文件有興趣即可再以詳述結構化摘要了解文件細部內容，即可更了解文件內容進而判斷文件之重要性。透過結構化摘要方式，知識搜尋者即可避免自由型式呈現摘要（如字數、表格呈現、文章結構與描述性等），進而影響文件篩選之準則。因此，本研究先行取得文件中之重點區塊與核心內容，再針對重點與核心部分進行結構化摘要之推論，並分為簡略摘要以及詳述摘要等兩種方式呈現。

針對簡略摘要推論部分，主要以簡短描述方式表達文件之動機與目的。因此，於簡略摘要階段，乃擷取最具有觀念性與代表性之語句，因而計算語句之向心性程度，同時

亦考量文章閱讀性且納入語句之結構強度作為準則判斷，進而取得簡略結構化摘要之最主要描述內容。針對詳述摘要推論部分，主要進行文件內容之細部描述，而詳述摘要與簡略摘要之不同處在於簡略摘要乃擷取文件之主要重點與核心，然詳述摘要主要呈現文件之細部內容。於此，詳述結構化摘要主要加入詞彙鏈概念，以表達文件考量文件流程性與時序性之描述。

### 3.3.1 簡略部分建立

結構化摘要之簡略部分主要乃計算語句之向心性程度，再進行語句之結構完整性之篩選。假設該語句同時存在多種觀念性語彙即表示該語句具有代表性，並將其語句列為候選語句以進行語句之結構強度計算。語句之結構強度計算乃考量摘要閱讀之順暢性，因而計算該語句之結構強度（即表示語句之主詞、受詞與動詞之間之關聯程度）。在此所擷取之語句集合包含：作業名稱集合、傷害成因集合以及改善方法之集合。簡略階段法則所使用符號定義如下：

#### 符號定義

CS <sub>i</sub>	目標文件 D <sub>T</sub> 之第 i 個具有人因工程工作場所研究報告觀念之語句
W <sub>1</sub>	具有作業名稱之觀念性語句且存在作業目的語彙之條件代碼
W <sub>2</sub>	具有作業名稱之觀念性語句且存在作業工具語彙之條件代碼
W <sub>3</sub>	具有傷害成因之觀念性語句且存在專業動詞語彙之條件代碼
W <sub>4</sub>	具有傷害成因之觀念性語句且存在作業工具語彙之條件代碼
W <sub>5</sub>	具有改善方法之觀念性語句且存在改善目的語彙之條件代碼
W <sub>6</sub>	具有改善方法之觀念性語句且存在作業工具語彙之條件代碼
W <sub>7</sub>	具有改善方法之觀念性語句且存在改善動詞語彙之條件代碼
ScoreON(CS)	該作業名稱語句之條件加總分數
ScoreIC(CS)	該傷害成因語句之條件加總分數
ScoreIM(CS)	該改善方法語句之條件加總分數
CandidateON	具作業名稱觀念之候選語彙集合
CandidateIC	具傷害成因觀念之候選語彙集合
CandidateIM	具改善方法觀念之候選語彙集合

TS	候選語彙之主詞
TV	候選語彙之動詞
TO	候選語彙之受詞
Score(TS,TV,TO)	該語句之結構加總分數
WS <sub>1</sub>	語句之向心性權重值
WS <sub>2</sub>	語句之結構性權重值
LowerLimit	擷取語句之最低門檻值
UppweLimit	擷取語句之最高門檻值
Total ScoreON(CS)	作業名稱語句之總分數
Total ScoreIC(CS)	傷害成因語句之總分數
Total ScoreIM(CS)	改善方法語句之總分數
SelectON	具作業名稱觀念之選擇語彙集合
SelectIC	具傷害成因觀念之選擇語彙集合
SelectM	具改善方法觀念之選擇語彙集合

#### 步驟(A1) 具有觀念性語句之向心性計算

根據簡略階段法則之定義擷取作業名稱、傷害成因與改善方法之觀念性語句乃為首要動作。在此步驟乃先計算作業名稱、傷害成因與改善方法之觀念性語句向心性，首先將其各表達項目之具有觀念性語句（即ON(CS<sub>i</sub>)、IC(CS<sub>i</sub>)與IM(CS<sub>i</sub>)）拆解成二至六字詞以形成集合表示（即ON(CS<sub>i,j</sub>)、IC(CS<sub>i,j</sub>)與IM(CS<sub>i,j</sub>)）。各表達項目之向心性判斷方式如公式(3.30)至公式(3.32)所示，首先以字詞進行各項語彙比對，並以條件代碼為1若不存在則為0方式累計向心性分數，最後乃以Score ON(CS)、ScoreIC(CS)與ScoreTM(CS)呈現觀念語句之向心性分數，並將具有向心性語彙放置於各表達項目之候選語彙集合。

$$\begin{aligned}
& \text{ON}(\text{CS}_i) = \{ \text{CS}_{i,1}, \text{CS}_{i,2}, \text{CS}_{i,3}, \dots, \text{CS}_{i,j}, \dots \} \\
& \text{IF ON}(\text{CS}_{i,j}) \text{ exist in OG}(\text{CS}) \forall j \\
& \text{Then } W_1 = 1, \text{ Otherwise } W_1 = 0 \\
& \text{IF ON}(\text{CS}_{i,j}) \text{ exist in OT}(\text{CS}) \forall j \\
& \text{Then } W_2 = 1, \text{ Otherwise } W_2 = 0 \\
& \text{Score ON}(\text{CS}_i) = W_1 + W_2 \\
& \text{IF } 1 < \text{ScoreON}(\text{CS}_i) \leq 2 \text{ Then ON}(\text{CS}_i) \\
& \in \text{CandidateON}
\end{aligned} \tag{3.30}$$

$$\begin{aligned}
& \text{IC}(\text{CS}_i) = \{ \text{CS}_{i,1}, \text{CS}_{i,2}, \text{CS}_{i,3} \dots, \text{CS}_{i,j}, \dots \} \\
& \text{IF IC}(\text{CS}_{i,j}) \text{ exist in PV}(\text{CS}) \forall j \\
& \text{Then } W_3 = 1, \text{ Otherwise } W_3 = 0 \\
& \text{IF IC}(\text{CS}_{i,j}) \text{ exist in B}(\text{CS}) \forall j \\
& \text{Then } W_4 = 1, \text{ Otherwise } W_4 = 0 \\
& \text{Score IC}(\text{CS}_i) = W_3 + W_4 \\
& \text{IF } 1 < \text{ScoreIC}(\text{CS}_i) \leq 2 \text{ Then IC}(\text{CS}_i) \\
& \in \text{CandidateIC}
\end{aligned} \tag{3.31}$$

$$\begin{aligned}
& \text{IM}(\text{CS}_i) = \{ \text{CS}_{i,1}, \text{CS}_{i,2}, \text{CS}_{i,3} \dots, \text{CS}_{i,j}, \dots \} \\
& \text{IF IM}(\text{CS}_{i,j}) \text{ exist in IG}(\text{CS}) \forall j \\
& \text{Then } W_5 = 1, \text{ Otherwise } W_5 = 0 \\
& \text{IF IM}(\text{CS}_{i,j}) \text{ exist in OT}(\text{CS}) \forall j \\
& \text{Then } W_6 = 1, \text{ Otherwise } W_6 = 0 \\
& \text{IF IM}(\text{CS}_{i,j}) \text{ exist in RV}(\text{CS}) \forall j \\
& \text{Then } W_7 = 1, \text{ Otherwise } W_7 = 0 \\
& \text{Score IM}(\text{CS}_i) = W_5 + W_6 + W_7 \\
& \text{IF } 2 < \text{ScoreIM}(\text{CS}_i) \leq 3 \text{ Then IM}(\text{CS}_i) \\
& \in \text{CandidateIM}
\end{aligned} \tag{3.32}$$

#### 步驟(A2) 具有觀念性語句之結構強度計算

經由候選語彙集合從中計算集合各語句之結構強度Score (TS, TV, TO)。針對各表達項目之主詞 (TS)、動詞 (TV) 以及受詞 (TO) 呈現不同，可分為三種語句結構。第一、完整結構語句，即表示該語句具三種結構元素之結合，其該語句之結構之分數Score (TS, TV, TO) 為 2。第二、半結構語句，假設該語句具有動詞但結合受詞或是主詞，即表示該語句乃須連結其他語句，其該語句之結構之分數Score (TS, TV, TO) 為 1。第三、

不完全結構語句，即表示該語句不具有動詞，同時亦表示該語句難以進行前後語句之連結，其該語句之結構之分數 Score (TS, TV, TO) 為 0，各表達項目之結構強度計算方式如公式(3.33)至公式(3.41)所示。

- 如公式(3.33)至公式(3.35)所示，針對作業名稱之觀念性語句之主詞 (TS) 部分可以作業名稱語彙 ON(CS) 判斷；動詞 (TV) 部分則以專業動詞語彙 PV(CS)、一般動詞語彙 GV(CS) 等判斷等之；受詞 (TO) 部分可以作業工具語彙 OT(CS) 判斷之。在此乃先判斷該語句是否具有動詞 (TV) 之結構，其次乃判斷受詞 (TO) 以及主詞 (TS) 部分。

$$\begin{aligned}
 ONCS_i &= \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
 IFON(CS_{i,j}) &\text{ existin } (PV(CS)) \text{ OR } (GV(CS)) \forall j \\
 \text{And } ON(CS_{i,j}) &\text{ existin } (OT(CS)) \forall j \\
 \text{And } IFON(CS_{i,j}) &\text{ existin } (ON(CS)) \forall j \\
 \text{Then } (Score(TS, TV, TO)) &= 2
 \end{aligned} \tag{3.33}$$

$$\begin{aligned}
 ONCS_j &= \{ CS_{j,1}, CS_{j,2}, CS_{j,3}, \dots, CS_{j,i}, \dots \} \\
 IFON(CS_j) &\text{ existin } (PV(CS)) \text{ OR } (GV(CS)) \forall j \\
 \text{And } ON(CS_j) &\text{ existin } (OT(CS)) \forall j \\
 \text{Or } IFON(CS_j) &\text{ existin } (ON(CS)) \forall j \\
 \text{Then } (Score(TS, TV, TO)) &= 1
 \end{aligned} \tag{3.34}$$

$$\begin{aligned}
 ONCS_i &= \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
 IFON(CS_{i,j}) &\text{ not exist in } (PV(CS)) \\
 \text{OR } (GV(CS)) &\forall j \\
 \text{Then } (Score(TS, TV, TO)) &= 0
 \end{aligned} \tag{3.35}$$

- 如公式(3.36)至公式(3.38)所示，針對傷害成因之觀念性語句之主詞 (TS) 部分可以傷害成因語彙 IC(CS) 判斷；動詞 (TV) 部分則以專業動詞語彙 PV(CS)、一般動詞語彙 GV(CS) 等判斷等之；受詞 (TO) 部分可以傷害成因語彙 IC(CS) 判斷之。

$$\begin{aligned}
& ICCS_i = \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
& IF IC(CS_{i,j}) \text{ exist in } (PV(CS)) \text{ OR } (GV(CS)) \forall j \\
& \text{And } IC(CS_{i,j}) \text{ exist in } (IC(CS)) \forall j \\
& \text{And IF } IC(CS_{i,j}) \text{ exist in } (IC(CS)) \forall j \\
& \text{Then } (Score(TS, TV, TO)) = 2
\end{aligned} \tag{3.36}$$

$$\begin{aligned}
& ICCS_i = \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
& IF IC(CS_{i,j}) \text{ exist in } (PV(CS)) \text{ OR } (GV(CS)) \forall j \\
& \text{And } IC(CS_{i,j}) \text{ exist in } (IC(CS)) \forall j \\
& \text{OR IF } IC(CS_{i,j}) \text{ exist in } (IC(CS)) \forall j \\
& \text{Then } (Score(TS, TV, TO)) = 1
\end{aligned} \tag{3.37}$$

$$\begin{aligned}
& ICCS_i = \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
& IF IC(CS_{i,j}) \text{ not exist in } (PV(CS)) \\
& \quad \text{OR } (GV(CS)) \forall j \\
& \text{Then } (Score(TS, TV, TO)) = 0
\end{aligned} \tag{3.38}$$

- 如公式(3.39)至公式(3.41)所示，針對改善方法之觀念性語句之主詞 (TS) 部分可以作業身分語彙ORT(CS)判斷；動詞 (TV) 部分則以專業動詞語彙PV(CS)、一般動詞語彙GV(CS)、評估動詞語彙RV(CS)等判斷等之；受詞 (TO) 部分可以作業工具語彙OT(CS)判斷之。

$$\begin{aligned}
& IMCS_i = \{ CS_{i,1}, CS_{i,2}, CS_{i,3}, \dots, CS_{i,j}, \dots \} \\
& IF IM(CS_{i,j}) \\
& \text{exist in } PV(CS) \text{ OR } GV(CS) \text{ OR } RV(CS) \forall j \\
& \text{And } IM(CS_{i,j}) \text{ exist in } (OT(CS)) \forall j \\
& \text{And IF } IC(CS_{i,j}) \text{ exist in } (ORT(CS)) \forall j \\
& \text{Then } (Score(TS, TV, TO)) = 2
\end{aligned} \tag{3.39}$$



$$\begin{aligned}
& \text{IMCS}_i = \{ \text{CS}_{i,1}, \text{CS}_{i,2}, \text{CS}_{i,3}, \dots, \text{CS}_{i,j}, \dots \} \\
& \text{IF IM}(\text{CS}_{i,j}) \text{ exist in PV}(\text{CS}) \\
& \text{OR GV}(\text{CS}) \text{ OR RV}(\text{CS}) \forall j \\
& \text{And IM}(\text{CS}_{i,j}) \text{ exist in (OT}(\text{CS})) \forall j \\
& \text{And IF IC}(\text{CS}_{i,j}) \text{ exist in (ORT}(\text{CS})) \forall j \\
& \text{Then (Score(TS, TV, TO))} = 1
\end{aligned} \tag{3.40}$$

$$\begin{aligned}
& \text{IMCS}_i = \{ \text{CS}_{i,1}, \text{CS}_{i,2}, \text{CS}_{i,3}, \dots, \text{CS}_{i,j}, \dots \} \\
& \text{IF IM}(\text{CS}_{i,j}) \text{ not exist in PV}(\text{CS}) \text{ OR GV}(\text{CS}) \\
& \quad \text{OR RV}(\text{CS}) \forall j \\
& \text{Then (Score(TS, TV, TO))} = 0
\end{aligned} \tag{3.41}$$

### 步驟(A3) 具有觀念性語句之權重值計算

考量結構化摘要之簡略部分主旨乃在於使知識搜尋者以短時間內篩選文件，為突顯語句向心性之重要性，本研究對其向心性與結構性分數各別加予權重值 $WS_1$ 、 $WS_2$ ，欲表示簡略性部分之摘要乃以語句之觀念向心性為主要考量，其次為語句之結構性，最後將其具有向心性分數並完整結構之語句擷取至各表達項目之選擇語彙集合，各表達項目之權重計算方式如公式(3.42)至公式(3.44)所示。

$$\begin{aligned}
& \text{TotalScore ON}(\text{CS}_i) = (\text{Score ON}(\text{CS}_i) * \text{WS}_1) \\
& + ((\text{Score(TS, TV, TO)}) * \text{WS}_2) \\
& \text{IF LowerLimit} < \text{TotalScoreON}(\text{CS}_i) \\
& \leq \text{UpperLimit Then ON}(\text{CS}_i) \in \text{SelectON}
\end{aligned} \tag{3.42}$$

$$\begin{aligned}
& \text{TotalScore IC}(\text{CS}_i) = (\text{Score IC}(\text{CS}_i) * \text{WS}_1) \\
& + (\text{Score}(\text{TS, TV, TO}) * \text{WS}_2) \\
& \text{IF LowerLimit} < \text{TotalScoreIC}(\text{CS}_i) \\
& \leq \text{UpperLimit Then IC}(\text{CS}_i) \in \text{SelectON}
\end{aligned} \tag{3.43}$$

$$\begin{aligned}
& \text{TotalScore IM}(\text{CS}_i) = (\text{Score IM}(\text{CS}_i) * \text{WS}_1) \\
& + ((\text{Score(TS, TV, TO)}) * \text{WS}_2) \\
& \text{IF LowerLimit} < \text{TotalScoreIC}(\text{CS}_i) \\
& \leq \text{UpperLimit Then IC}(\text{CS}_i) \in \text{SelectON}
\end{aligned} \tag{3.44}$$

呈上述步驟其觀念乃將具有作業名稱、傷害成因、改善方法等觀念性語句權重之計算與擷取。根據語句之向心性以及語句之結構完整性作為簡略部份之條件，並以條件之權重分數篩選以及擷取語句，以形成簡略部分結構性摘要。

### 3.3.2 詳述部分建立

於詳述部分之內容乃為描述部分以及評估部分，其內容多為細項描述。根據上述，針對描述部分所擷取之語句集合包含：作業行為集合以及作業環境集合，而評估部分所擷取集合乃以改善方法集合為主。本階段擷取語句方式乃以下三項為主要步驟，分別為語彙相依程度計算、連結詞之連接語句以及標點符號以判斷結束語句。詳述階段法則所使用之符號定義如下。

#### 符號定義

EndMark_Set	結尾符號之集合，其標點符號包含：。、!、?等
WOD <sub>i,j</sub>	具作業描述之第 i 個語彙於第 j 個語彙庫出現之權重值
TFOD <sub>i,j</sub>	具作業描述之第 i 個語彙於第 j 個語彙庫出現之頻率
NumSet	具觀念性語彙庫之數目
ODF <sub>i</sub>	具作業描述之第 i 個語彙於語彙庫之數目
ResSet	具有觀念性語彙候選語句之集合
Res(S)	具有觀念性語彙之候選語句
P <sub>i</sub>	人因工程工作場所研究報告第 i 個段落
CW <sub>k</sub> [P <sub>i</sub> ]	第 i 個段落中第 k 個語彙
I[i,k]	標記第 i 個段落之第 k 個語彙及第 k+1 個語彙是否為具有順序關係(1 代表具此順序關係，0 則無關係)
F[CW <sub>k</sub> ,CW <sub>k+1</sub> ]	人因工程工作場所改善告段落中，出現具有觀念性語彙之順序即語彙 k 至語彙 k+1 之發生頻率
P <sup>From</sup> (k)	以語彙 k 為關係之起頻率總和
P <sup>To</sup> (k)	以語彙 k 為關係之迄頻率總和
R[CW <sub>k</sub> ]	判定第 k 個語彙先後關係之比例值
SCW	語彙之整體順序趨勢

$P(\text{Res}(S_i) \text{Res}(S_{i-1}))$	第 i 個候選語句與第 i-1 個候選語句之連結機率值
$\text{WOD}_{i,j}$	具作業描述之第 i 個語彙於第 j 個語彙庫出現之權重值
$\text{Mk}_{i,j}$	標記第 i 個候選語句是否符合結構化摘要之第 j 個建立條件（1 代表符合條件，0 則無符合）
$\text{MaxR}[\text{CW}_k]$	判定第 k 個語彙與該段落存在全部語彙之最大比例值
$\text{MinR}[\text{CW}_k]$	判定第 k 個語彙與該段落存在全部語彙之最小比例值
$\text{Sum1}(\text{Res}(S_i))$	加總第 i 個候選語句是否符合結構化摘要之第 2 個與第 3 個建立條件之標記，即 $\text{Mk}_{i,2}$ 與 $\text{Mk}_{i,3}$ 之加總
$\text{Sum2}(\text{Res}(S_i))$	加總第 i 個候選語句是否符合結構化摘要之第 5 個與第 6 個建立條件之標記，即 $\text{Mk}_{i,5}$ 與 $\text{Mk}_{i,6}$ 之加總

#### 步驟(B1) 擷取具有作業描述之觀念性語句

依據階段部分可得欲擷取語句集合包含：作業行為集合以及作業環境集合。由於作業行為乃有多種語彙集合而成，在此本研究乃以作業描述語彙為基進行各語彙之比對。首先乃將作業行為觀念性語彙以及各隱含語彙進行TF-ISF字詞相似度計算，再以統計機率方式計算語彙與語彙之間相似性判斷，即可擷取出最具相關性之觀念性語彙。TF-ISF字詞權重計算之方法如公式(3.45)所示，其方式乃以作業描述（OD）為判定語彙，並與其他潛在語彙進行詞頻計算乃求得作業描述於各潛在語彙之相似關係，並取得 $\text{WOD}_{i,j}$ 較高分數之語彙作為候選語句 $\text{Res}(S_i)$ 並儲存於候選語句之集合 $\text{ResSet}$ 中。從候選語句之集合中，將其候選語句拆解成二至六字詞。如公式(3.46)所示，經由字詞拆解以形成集合並以 $\text{Res}(S_{i,j})$ 表示之，以進行語彙之判斷比對。

$$\text{WOD}_{i,j} = \text{TFOD}_{i,j} \cdot \log \frac{\text{NumSet}}{\text{ODF}_i} \quad (3.45)$$

$$\text{Res}(S_i) = \{S_{i,1}, S_{i,2}, S_{i,3}, \dots, S_{i,j}, \dots\} \quad (3.46)$$

#### 步驟(B2) 計算具有觀念性語彙之相依程度

為避免前後文語意之問題，本研究乃以語彙與語彙出現機率關係為基，透過語彙與語彙出現順序之關係，作為語句連結之依據以預測下一順序之語句。即候選語句  $\text{Res}(S_i)$

與後續連接語句  $\text{Res}(S_{i-1})$  之連結機率性並以  $P(\text{Res}(S_i)|\text{Res}(S_{i-1}))$  表示之，假設兩語句與具有關聯性即表示連結機率值越大；反之，假設若不具有相關性或不符語意其機率值越小。於此，本步驟乃先進行語彙之出現順序推導，並以段落  $P_i$  為一單位計算該段落出現之語彙（即  $CW_k[P_i]$ ）與語彙間出現順序關係。

假設在第  $i$  個段落中存在第  $k$  個語彙  $CW_k[P_i]$  之語彙具有作業目的或專業動詞等觀念性（以  $A(CS)$  表示之），且下一個出現語彙  $CS_{k+1}[P_i]$  亦具有其觀念性（以  $B(CS)$  表示之），則標記  $I[i,k]$  為 1 表示具有順序關係，若為 0 則無關係（如公式(3.47)所示）。待標記完成後，再將段落中出現之語彙順序關係進行加總並以  $F[CW_k, CW_{k+1}]$  表示之，根據公式(3.48)即可計算語彙  $k$  至語彙  $k+1$  之發生頻率。

$$I[i,k] = \begin{cases} 1, & \text{IF } CW_k[P_i] \in A(CS), CS_{k+1}[P_i] \in B(CS) \\ 0, & \text{Otherwise} \end{cases} \quad (3.47)$$

$$F[CW_k, CW_{k+1}] = \sum_{all i, all k} I[i,k] \quad (3.48)$$

上述動作即可得知語彙  $k$  至語彙  $k+1$  之發生頻率，另加入該語彙發生於段落之起始與結束之頻率以判定語彙整體先後順序關係，並如公式(3.49)至公式(3.51)所示以  $P^{From}(k)$  與  $P^{To}(k)$  之比值  $R[CW_k]$  表示。根據比值  $R[CW_k]$  排序，即可得知語彙  $k$  至語彙  $k+1$  之發生先後順序，並根據比值大小進行整體趨勢順序排序 SCSW，如公式(3.52)所示。

$$P^{From}(k) = \sum_{all k+n} F[CW_k, CW_{k+n}] \quad (3.49)$$

$$P^{To}(k) = \sum_{all k} F[CW_{k+n}, CW_k] \quad (3.50)$$

$$R[CW_k] = \frac{P^{From}(k)}{P^{To}(k)} \quad (3.51)$$

$$\begin{aligned} \text{SCSW} &= CW'_{k-1} \rightarrow CW'_k \rightarrow \dots \rightarrow CW'_{k+n} \rightarrow \dots \\ \text{Where } &R[CW'_{k-1}] \geq R[CW'_k] \geq \dots \geq R[CW'_{k+n}] \geq \dots \end{aligned} \quad (3.52)$$

透過整體趨勢排序 SCSW 即得知語彙與語彙間相依程度關係，根據語彙之間相依程度即可篩選具有依性之語彙關係，並以此為擷取觀念性語句為摘要之主要依據。

#### 步驟(B3) 定義具有連結語彙之觀念性語句

針對連結語彙之觀念性語句擷取方式，本步驟乃以「知識文件專業語彙庫」之連結語彙C(CS)進行篩選，透過連結語彙即可判斷該語句 $Res(S_i)$ 是否具有連結語句之功能。於連結語彙中可得連結之字詞，例如：「即」、「以」等，當中亦可在細分連結字詞之邏輯順序，例如連結字詞「由於」以及「進而」為比較其順序關係，根據語意邏輯可知「進而」之位置為「由於」之後。於此，本研究在此步驟將連結語彙進行細分出銜接語彙LinkC(CS)，以進行摘要結構之順暢性。由於具有連結字詞乃為語句之第一個字詞，於此本步驟即判斷該語句之第一個字詞是否為連結語彙或是銜接語彙即可確定並擷取具有連結語彙之觀念性語句。透過二至六字詞拆解，即判斷該語句之第一個具有意義之字詞是否存在於連結語彙C(CS)或是銜接語彙LinkC(CS)。

#### 步驟(B4) 定義開頭以及結束觀念性語句

本步驟乃透過步驟(B2)得知語彙迄總和  $P^{To}(K)$ 、起總和  $P^{From}(K)$ 以及起始與結束頻率比值  $R[CW_k]$ ，從中即可判定該語彙乃為開頭起始語彙或是結束語彙。結構化摘要之起始語彙多為表示目的，例如：作業目的語彙、作業名稱語彙等，並數  $R[CW_k]$ 與該段落存在語彙之比值，進行判定該語彙是否為起始語彙，若其比值為最大值  $MaxR[CW_k]$ 即表示該語彙為起之語彙。於此方式，亦可以比值判定該語彙若為最小值  $MinR[CW_k]$ 即有結束語彙之可能性，此外，針對結束語彙亦可以判斷該候選語句存在結尾符號及表示該語句為結束語彙。

#### 步驟(B5) 計算與擷取觀念性語句

經由上述步驟已可取得結構化摘要建立之原則，於本步驟乃將上述原則進行統整。經由字詞拆解，於本步驟乃以  $Res(S_{i,j})$ 判斷該語句是否符合結構化摘要建立之條件，並以結構化摘要符合標記值  $Mk_{i,j}$  表示之，其原則以及相對應之公式如公式(3.53)至公式(3.58)所示：

- 若候選語句  $Res(S_i)$ 具有整體趨勢順序排序 SCSW 之語彙  $k$ ，則結構化摘要符合標記值  $Mk_{i,1}=1$ 。

- 若候選語句  $Res(S_i)$  中存在連結語彙  $C(CS)$ ，則結構化摘要符合標記值  $Mk_{i,2}=1$ 。
- 若候選語句  $Res(S_i)$  中存在銜接語彙  $LinkC(CS)$ ，則結構化摘要符合標記值  $Mk_{i,3}=1$ 。
- 若候選語句  $Res(S_i)$  中存在語彙之評斷係數  $R[CW_k]$  與出現語彙總數量之比值為最大值，則結構化摘要符合標記值  $Mk_{i,4}=1$ 。
- 若候選語句  $Res(S_i)$  中存在語彙之評斷係數  $R[CW_k]$  與出現語彙總數量之比值為最小值，則結構化摘要符合標記值  $Mk_{i,5}=1$ 。
- 若候選語句  $Res(S_i)$  中存在結尾符號集中  $EndMark\_Set$ ，則結構化摘要符合標記值  $Mk_{i,6}=1$ 。

$$MK_{i,1} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \text{ exit in } SCSW \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.53)$$

$$MK_{i,2} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \\ & \text{exit in } C(CS) \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.54)$$

$$MK_{i,3} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \\ & \text{exit in } LinkC(CS) \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.55)$$

$$MK_{i,4} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \\ & \text{exit in } CW_k \\ & \text{and } R[CW_k] = MaxR[CW_k] \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.56)$$

$$MK_{i,5} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \text{ exit in } CW_k \\ & \text{and } R[CW_k] = MinR[CW_k] \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.57)$$

$$MK_{i,6} = \begin{cases} 1, & \text{IF } Res(S_{i,j}) \\ & \text{exit in } EndMark\_Set \forall j \\ 0, & \text{Otherwise} \end{cases} \quad (3.58)$$

綜合上述，於最後可將候選語句  $Res(S_i)$  之對應結構化摘要符合標記值  $Mk_{i,j}$  彙整如公式(3.59)。根據結構化摘要符合標記值  $Mk_{i,j}$  即作為擷取候選語句之考量，同時亦可確認該候選語句之放置位置。假設該候選語句之  $Mk_{i,4}$  為 1 即表示具有摘要之開頭，透過公式(3.60)與公式(3.61)以  $Mk_{i,2}$  與  $Mk_{i,3}$  之加總  $Sum1(Res(S_i))$  作為描述階段之篩選，最後則以  $Mk_{i,5}$  與  $Mk_{i,6}$  之加總  $Sum2(Res(S_i))$  作為評估階段之擷取根據。

$$MK = \begin{bmatrix} MK_{1,1} & MK_{2,1} & \dots & MK_{i,1} \\ MK_{1,2} & MK_{2,2} & \dots & MK_{i,2} \\ MK_{1,3} & MK_{2,3} & \dots & MK_{i,3} \\ MK_{1,4} & MK_{2,4} & \dots & \dots \\ MK_{1,5} & MK_{2,5} & \dots & MK_{i,5} \\ MK_{1,6} & MK_{2,6} & \dots & MK_{i,6} \end{bmatrix} \quad (3.59)$$

$$Sum1(Res(S_i)) = \sum_{j=2}^3 MK_{i,j} \quad (3.60)$$

$$Sum2(Res(S_i)) = \sum_{j=5}^6 MK_{i,j} \quad (3.61)$$

根據步驟(B2)語彙相依程度計算之結果，即可取得語彙整體趨勢順序排序SCSW。本研究乃以公式(3.62)SCSW結果作為主要依據，並根據其他結構化摘要建立之條件（即連結語彙或是屬於開頭、結束語彙），作為為摘要歸屬位置之輔佐。其詳述部份結構化摘要擷取語句以及各語句歸屬位置。

$$SCSW = CW'_{k-1} \rightarrow CW'_k \rightarrow \dots \rightarrow CW'_{k+n} \rightarrow \dots$$

$$\text{Where } R[CW'_{k-1}] \geq R[CW'_k] \geq \dots$$

$$\geq R[CW'_{k+n}] \geq \dots \quad (3.62)$$

### 3.4 結論

針對知識搜尋方式，大多乃以關鍵字、文件分類等方式加強資訊取得之效率，但於特定領域知識過於專業化而提高搜尋障礙與時間，因此本研究乃建立一套結合語意解析與文件摘要技術之知識文件檢索模式，透過文件解析已得知文件表達項目並建立「知識語彙庫」以及語彙法則，以進行詢問詞隱含目的判斷，進而取得具有相關性文件，亦可透過語彙法則作為結構化摘要之條件，期望於文件中擷取具有代表性語彙並作為判定依據，進而建立結構化摘要，以加強知識搜尋者搜尋方式。





## 第四章、系統架構規劃

針對前一章節所發展之方法論與模式，本研究乃開發一套結合語意解析與文件摘要技術之知識文件檢索系統，以確認方法論與模式之可行性。此系統之功能重點乃透過使用者上傳知識文件，系統管理者先行設定系統參數與新增詢問詞與回覆詞，進而執行知識文件表達項目分析、問答解析與結構化摘要等主功能。本章即針對本研究所提之「結合語意解析與文件摘要技術之知識文件檢索系統」，分別以系統核心架構、系統功能架構、資料模式定義及開發工具等主題進行深入說明。

### 4.1 結合語意解析與文件摘要技術之知識文件檢索系統之流程架構

本研究所開發之「結合語意解析與文件摘要技術之知識文件檢索系統」依其運作流程可分為「知識文件上傳」、「知識文件表達項目解析」、「知識文件問答解析」及「知識文件結構化摘要推論」等四大階段，此系統之運作流程架構如圖4.1所示，各功能層次之詳細流程說明如下。

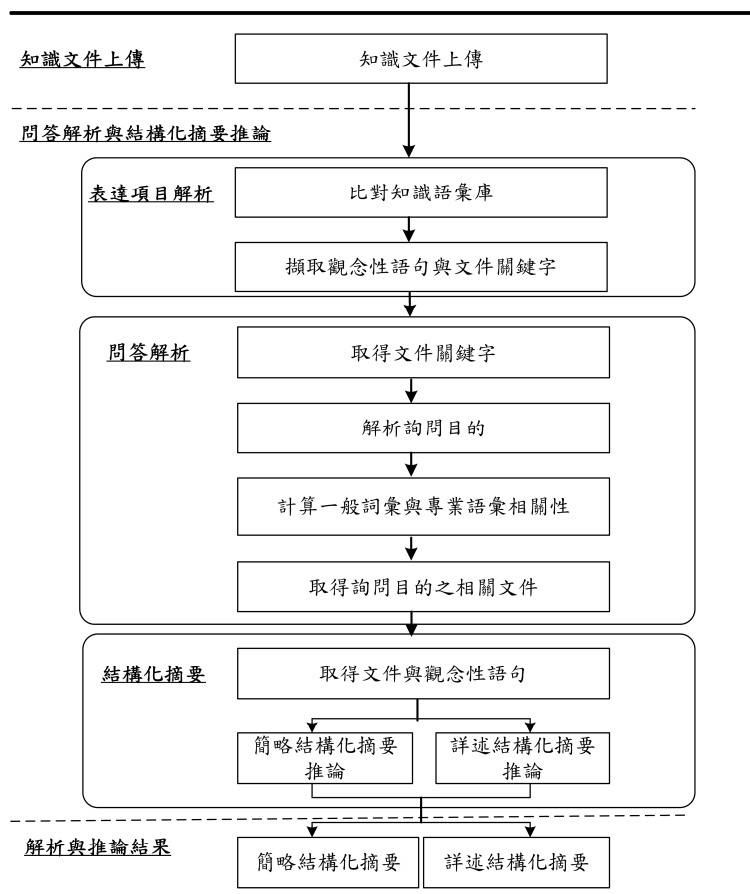


圖 4.1、結合語意解析與文件摘要技術之知識文件檢索系統之流程架構

➤ 知識文件上傳

使用者可將尚未進行文件表達項目解析與結構化摘要之知識文件新增至系統，系統即擷取文件內容，並由後續「知識文件表達項目解析」、「知識文件問答解析」及「知識文件結構化摘要推論」等模組判定該目標文件之表達項目與文件關鍵字詞並建立結構化摘要。

➤ 知識文件表達項目解析

此部份乃將目標文件進行解析，以擷取具有表達項目之觀念性語句同時計算該文件之關鍵字詞以作為後續問答解析與結構化摘要之依據。首先乃將文件斷句取得文件各語句，再透過語彙法則比對知識語彙庫並篩選出各表達項目（如：作業身分、作業行為、作業環境等）之觀念性語句，同時亦計算文件之關鍵字詞。

➤ 知識文件問答解析

此部分主要功能為解析搜尋字串以取得搜尋字串之詢問目的，同時連結相關語意詞彙，進而與文件關鍵字詞計算關聯性，並以此為根據進行相關文件搜尋。於此部分乃先需解析搜尋字串之主要詢問目的，並根據詢問目的詞彙連結相關回覆詞，並以回覆詞與文件之專業語彙進行相似性分析，以取得該回覆詞之於該專業類別解釋語意能力，即以詢問詞、回覆詞與專業語彙三者進行連結以取得相關文件。

➤ 知識文件結構化摘要推論

此部分乃推論目標文件之簡略結構化摘要與詳述結構化摘要，因此於此部份可在細分為「簡略結構化摘要推論」、「詳述結構化摘要推論」，簡略結構化摘要為簡單描述文件之動機、目的與結果，詳述結構化摘要則進行細部描述。

## 4.2 系統功能架構

本研究所建置結合語意解析與文件摘要技術之知識文件檢索系統乃架構於網際網路環境下。使用者可透過網際網路登入本系統，並使用本系統所提供之各項功能。當使用者登入系統後，系統即根據使用者帳號判斷該使用者於系統中之功能權限。

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

## 一般使用者

1. 可新增上傳未建立結構化摘要之知識文件上傳至系統
2. 可瀏覽符合使用者權限之所有知識文件
3. 可瀏覽符合使用者權限之詢問詞與回覆詞配對組合

## 系統管理者

1. 可上傳未建立結構化摘要之知識文件上傳至系統
2. 可瀏覽/編輯系統資料庫內之所有知識文件
3. 可查詢、新增、修改或刪除詢問詞與回覆詞配對組合內容
4. 可修改/查詢系統參數與門檻值
5. 可執行知識文件結構化摘要推論
6. 可執行知識文件問答解析推論

本系統所開發之重點模組共有「知識文件表達項目解析模組」、「知識文件問答解析模組」、「知識文件結構化摘要推論模組」、「知識文件問答維護模組」、「知識文件管理模組」及「系統參數設定」等六大模組；圖4.2即表示結合語意解析與文件摘要技術之知識文件檢索系統之核心模組架構。

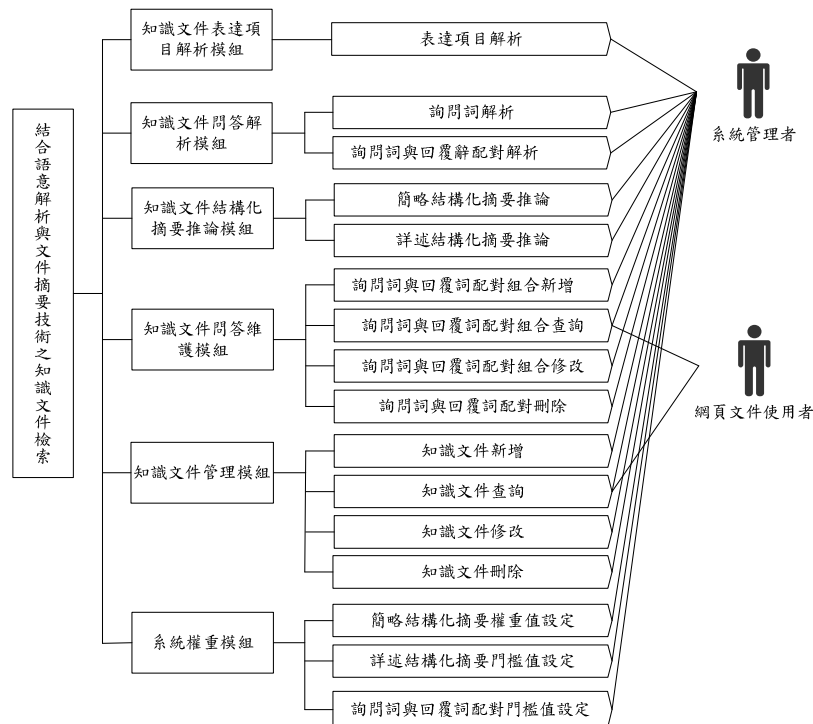


圖 4.2、結合語意解析與文件摘要技術之知識文件檢索系統之功能架構

針對上述系統架構所包含之基本功能模組說明如下：

**(一) 知識文件表達項目解析模組**

- 表達項目解析：系統管理者先輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後即可選擇需要判定之文件，系統將擷取知識文件之七種表達項目（即作業名稱、作業身分、作業環境、作業行為、作業時間、傷害成因與改善方法等）之觀念性語句，以待後續模組使用。

**(二) 知識文件問答解析模組**

- 詢問詞解析功能與文件篩選：系統管理者可輸入查詢字串，以取得該字串之主要詢問詞目的，以連結具有相關性之回覆詞，進而透過回覆詞與文件關鍵字相似性分析，以篩選文件。

**(三) 知識文件結構化摘要模組**

- 簡略結構化摘要推論：系統管理者可輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後即可選擇需要判定之文件，即可進行簡略結構化摘要推論。
- 詳述結構化摘要推論：系統管理者可輸入查詢字串、選取查詢條件、選擇上傳時間範圍及點選邏輯運算子，待查詢完畢後即可選擇需要判定之文件，即可進行詳述結構化摘要推論。

**(四) 知識文件問答維護模組**

- 詢問詞與回覆詞配對組合新增：提供系統管理者將詢問詞與回覆詞配對組合資料匯入並維護於系統資料表中。
- 詢問詞與回覆詞配對組合查詢：提供系統管理者與一般使用者進行查詢已維護之配對組合資料。
- 詢問詞與回覆詞配對組合修改：提供系統管理者修改錯誤之配對組合資料。
- 詢問詞與回覆詞配對刪除：提供系統管理者刪除錯誤之配對組合資料。

**(五) 知識文件管理模組**

- 知識文件新增功能：提供使用者與系統管理者透過上傳功能，將知識文件訓練樣本上傳至系統中。
- 知識文件查詢功能：提供使用者進行查詢已新增之知識文件。
- 知識文件修改功能：提供系統管理者修改錯誤之知識文件資料。
- 知識文件刪除功能：提供系統管理者刪除錯誤之知識文件資料。

#### (六) 系統參數設定模組

- 系統門檻值設定：提供權限內使用者進行修改與維護各模組之門檻如文件關鍵字篩選門檻、詢問詞與回覆詞配對組合篩選門檻等設定以加強系統準確性。
- 系統權重值設定：提供權限內使用者進行修改與維護簡略結構化摘要推論、詳述結構化摘要推論之權重質設定，進而保持系統參數之正確性。

本系統之使用者可分為一般使用者與系統管理者，並依權限而有不同執行權力。針對一般使用者可執行知識文件上傳新增、查詢等功能（如圖4.3之一般使用者所示），而系統管理者則可進行知識文件上傳新增、系統參數設定、知識文件問答解析推論、知識文件結構化摘要推論、知識文件問答解析維護與知識文件管理等（如圖4.3之系統管理者所示）。當使用者上傳未推論之知識文件後，系統管理者乃需完成參數設定以及新增詢問詞與回覆詞詞彙，待設定與新增完畢後，系統管理者即可執行知識文件表達項目解析模組以取得該文件之各表達項目之觀念性語句與文件之關鍵字詞。系統管理者再藉由知識文件問答解析模組判定搜尋字串之詢問目的進而篩選相關文件，亦可透過知識文件結構化摘要模組將文件進行簡略結構化摘要與詳述結構化摘要，最後使用者即由知識文件管理模組之查詢功能檢視系統之知識文件與判定結果，亦可透過知識文件問答維護模組檢視詢問詞與回覆詞配對組合結果。

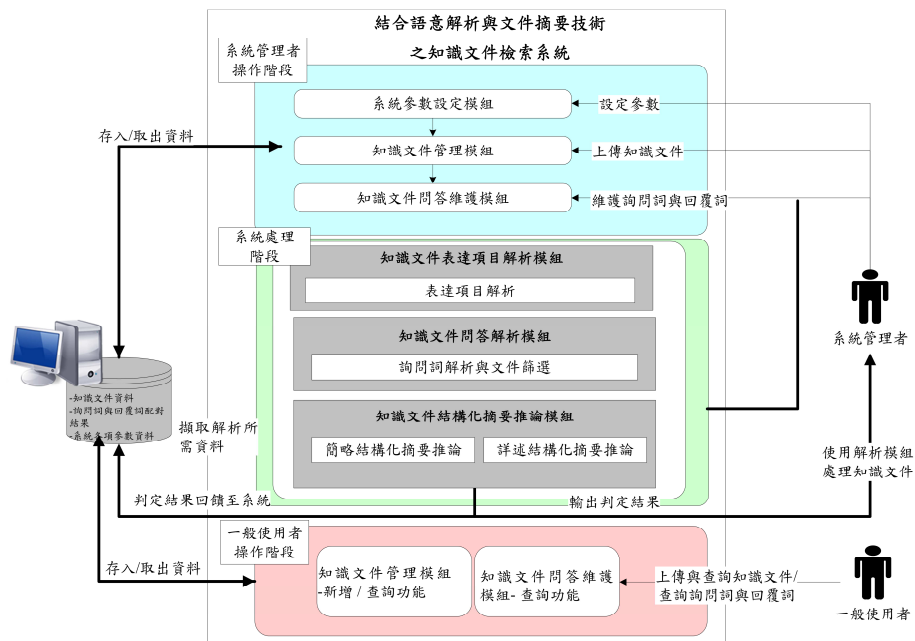


圖 4.3、結合語意解析與文件摘要技術之知識文件檢索系統運作架構

### 4.3 資料模式定義

本研究所發展之結合語意解析與文件摘要技術之知識文件檢索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，期使知識文件管理、知識文件問答解析與結構化摘要推論等任務可即時完成。依據系統運作之需要，將結合語意解析與文件摘要技術之知識文件檢索系統之資料分為「表達項目類別資料」、「表達項目資料」、「詞彙關聯資料」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) 表達項目類別資料

此資料之目的乃記錄知識文件各表達項目之類別資料，以有效分類知識文件表達項目；其所屬之資料表及其相關定義說明如下：

- 作業行為類別資料表：記錄作業行為表達項目所屬類別名稱。
- 作業環境類別資料表：記錄作業環境表達項目所屬類別名稱。
- 作業時間類別資料表：記錄作業時間表達項目所屬類別名稱。
- 改善方法類別資料表：記錄改善方法表達項目所屬類別名稱。
- 傷害成因類別資料表：記錄傷害成因表達項目所屬類別名稱。

#### (二) 表達項目資料

此資料之目的乃記錄知識文件各表達項目資料，以維護知識文件之表達項目資料；

其所屬之資料表及其相關定義說明如下：

- 作業行為資料表：記錄作業行為表達項目之語彙，如：作業行為編號、作業行為語彙以及所屬表達項目類別等，以作為知識文件表達項目解析模組應用之基礎。
- 作業環境資料表：記錄作業環境表達項目之語彙，如：作業環境編號、作業環境語彙以及所屬表達項目類別等，以作為知識文件表達項目解析模組應用之基礎。
- 作業時間資料表：記錄作業時間表達項目之語彙，如：作業時間編號、作業時間語彙以及所屬表達項目類別等，以作為知識文件表達項目解析模組應用之基礎。
- 改善方法資料表：記錄改善方法表達項目之語彙，如：改善方法編號、改善方法語彙以及所屬表達項目類別等，以作為知識文件表達項目解析模組應用之基礎。
- 傷害成因資料表：記錄傷害成因表達項目之語彙，如：傷害成因編號、傷害成因語彙以及所屬表達項目類別等，以作為知識文件表達項目解析模組應用之基礎。

### (三) 詞彙關聯資料

此資料之目的乃記錄知識文件資料內容與解析時所需關鍵資訊，以有效進行知識文件管理、問答解析與結構化摘要推論；其所屬之資料表及其相關定義說明如下：

- 文件資料表：記錄文件之基本資料，如文件編號、文件檔案名稱、文件內容與上傳等資訊。
- 詞彙鏈資料表：記錄詞彙鏈編號、詞彙鏈之詞彙、該詞彙所屬段落以及文件來源等，以作為知識文件結構化摘要模組應用之基礎。
- 詢問詞與回覆詞配對資料表：記錄詢問詞與回覆詞配對組合編號、配對組合之詢問詞與回覆詞，以作為知識文件問答解析模組應用之基礎。
- 回覆詞與專業語彙配對資料表：記錄回覆詞與專業語彙配對組合編號、配對組合之回覆詞與專業語彙，以作為知識文件問答解析模組應用之基礎。
- 標點符號資料表：定義知識文件中可能使用之各種符號。
- 非關鍵字基本資料表：定義不同字數之參照用非關鍵字集合，以作為知識文件結構化摘要模組應用之基礎。

### (四) 系統參數資料

此資料之目的乃記錄系統參數之資料，如系統門檻值、權重值等資料，透過設定以有效提高問答解析、結構化摘要推論之準確率；其所屬資料表及其相關定義如下：

- 系統參數名稱資料表：記錄系統參數名稱與參數敘述等資料。
- 系統參數值：記錄系統參數之數值，以影響判定模組之數據結果準確率。

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，使本研究所發展之結合語意解析與文件摘要技術之知識文件檢索系統可方便地進行文件與資料控管，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。

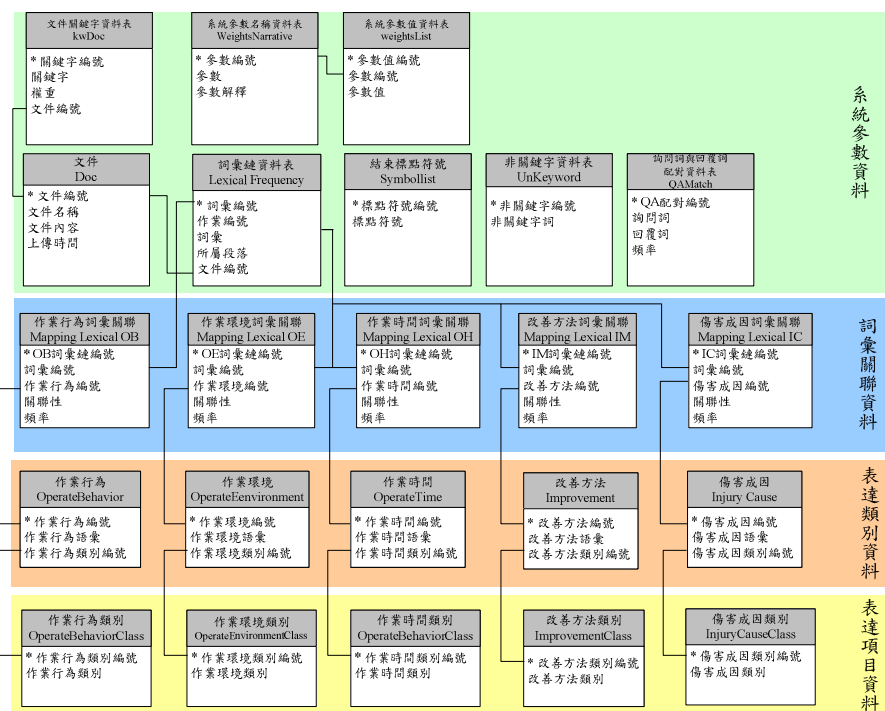


圖 4.4、結合語意解析與文件摘要技術之知識文件檢索系統之資料關聯

## 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「知識文件表達項目



解析模組」、「知識文件問答解析模組」、「知識文件結構化摘要推論模組」、「知識文件問答維護模組」、「知識文件管理模組」及「系統參數設定」等六大模組，以下即說明各系統功能之流程規劃。

### 知識文件表達項目解析模組

此模組之「表達項目解析」功能可供一般使用者與系統人員先行查詢欲解析之知識文件，取得該份文件並解析。系統先進行文件與知識語彙庫比對，進而取得各表達項目之觀念性語句，以作為問答解析模組中專業語彙連結資料與知識文件結構化摘要推論資料；其流程設計概念如圖 4.5 所示。

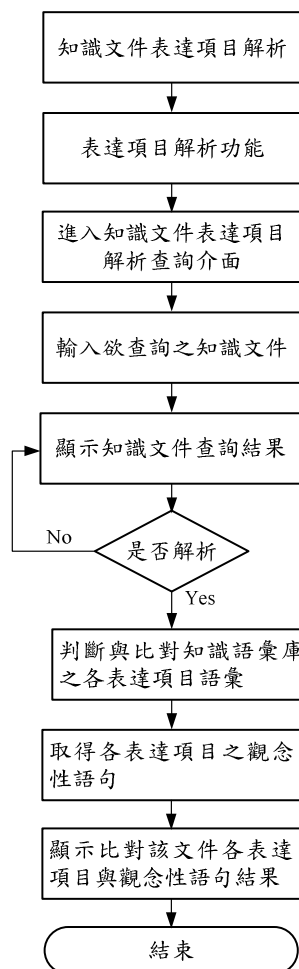


圖 4.5、「知識文件表達項目解析模組」功能流程

### 知識文件問答解析模組

系統管理者於此模組可進行「詢問詞解析與文件篩選」功能。於「詢問詞解析與文件篩選」功能中，系統管理者可輸入詢問句，系統即根據詢問句進行字詞拆解，同時計

算各字詞出現頻率並將高頻率之字詞作為主要詢問詞，藉由詢問詞解析以取得主要詢問目的。於「詢問詞與回覆詞配對解析」功能中，一般使用者與系統管理者必須先輸入詢問詞、回覆句與回覆詞，系統即可計算回覆詞於回覆句出現機率，同時取得較高出現機率之回覆詞與詢問詞配對組合。系統亦將配對組合儲存於資料庫中，以進行「知識文件問答維護」模組，進而保持知識文件問答解析之正確性。其流程設計概念如圖 4.6 所示。

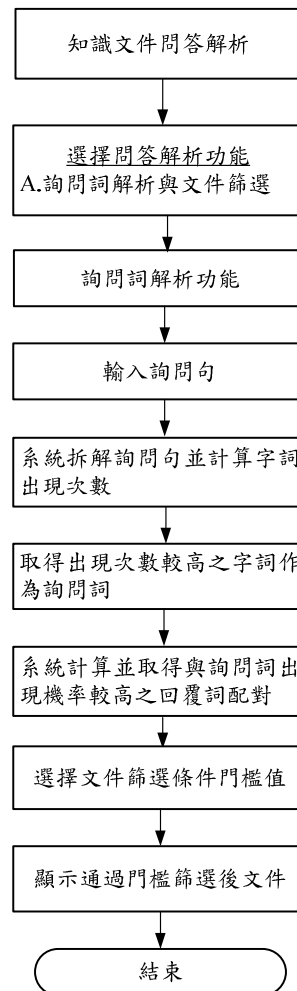


圖 4.6、「知識文件問答解析模組」功能流程

### 知識文件結構化摘要推論模組

一般使用者與系統管理者於此模組可進行「簡略結構化摘要推論」功能與「詳述結構化摘要推論」功能，於「簡略結構化摘要推論」功能中，一般使用者與系統管理者先行查詢欲解析之文件並決定推論，系統則先取得該文件之觀念性語句，再計算各語句之向心性分數與結構性分數，再分別乘上各權重值並進行加總，即可取得各語句之總分數並以此為依據排序、篩選且彙整各表達項目所表示之語句即完成簡略結構化摘要推論。

於「詳述結構化摘要推論」功能中，一般使用者與系統管理者亦以查詢方式取得欲解析之文件並決定推論，系統取得知識文件之觀念性語句後即計算該文件出現字詞之起訖總和，並以起訖總和為基取得詞彙鏈判斷係數進而建立詞彙鏈，透過詞彙鏈即可重新組合文件語句，最後則彙整各表達項目所表示之語句以完成詳述結構化摘要推論。其流程設計概念如圖 4.7 所示。

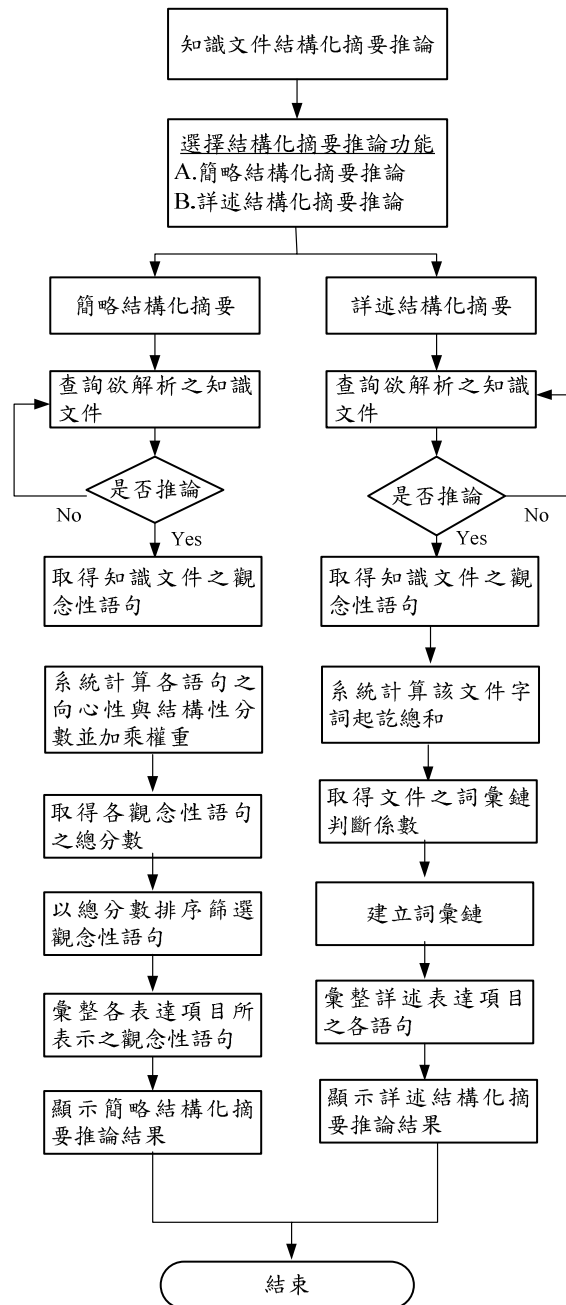


圖 4.7、「知識文件結構化摘要推論模組」功能流程

## 知識文件問答維護模組

此模組乃提供系統管理者新增、查詢、刪除及修改回覆詞與詢問詞。知識文件問答維護模組包含「詢問詞與回覆詞新增」、「詢問詞與回覆詞查詢」、「詢問詞與回覆詞刪除」與「詢問詞與回覆詞修改」等四大功能；其中，「詢問詞與回覆詞新增」功能乃提供系統管理者將詢問詞與回覆詞匯入系統資料庫內。「詢問詞與回覆詞查詢」功能乃提供系統管理者查詢詢問詞與回覆詞配對組合。此外，「詢問詞與回覆詞修改」與「詢問詞與回覆詞刪除」功能乃提供系統管理者進行修改與維護錯誤之問答配對，進而保持知識文件問答解析之正確性；其流程設計概念如圖 4.8 所示。

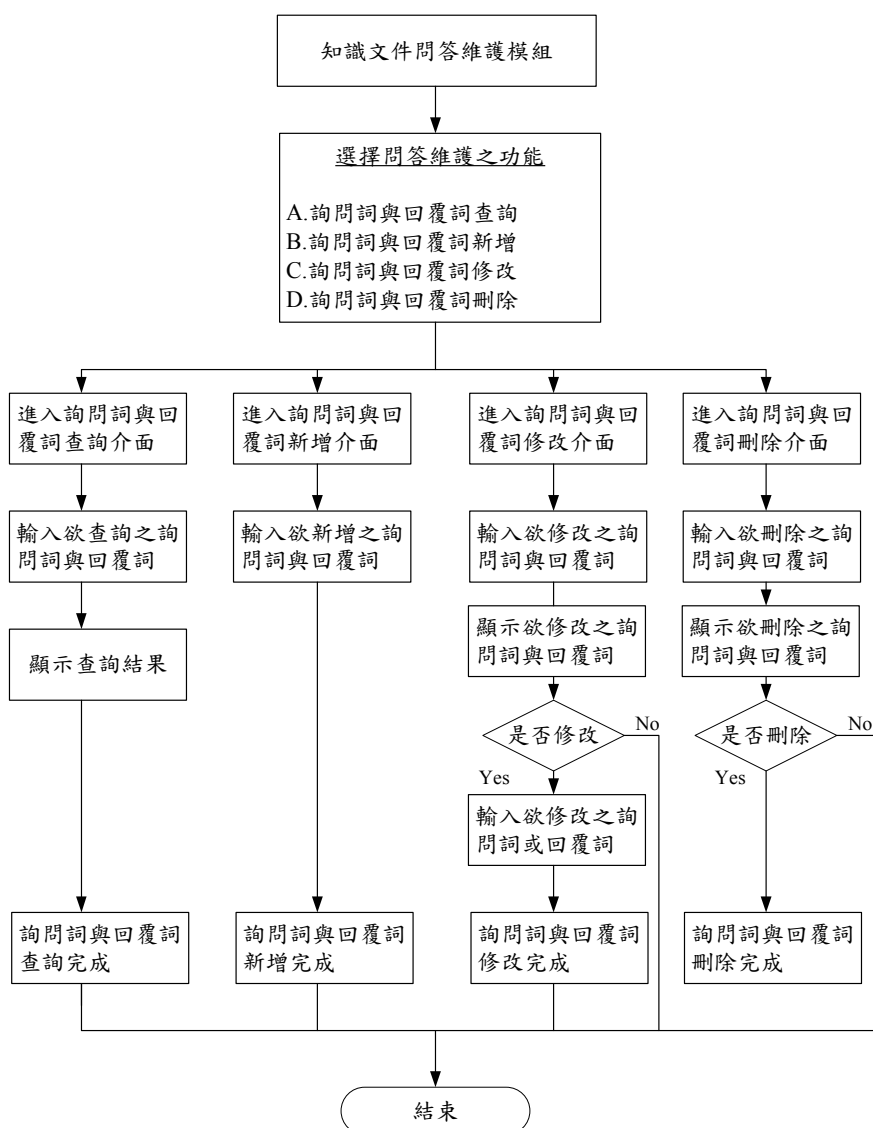


圖 4.8、「知識文件問答維護模組」功能流程

## 知識文件管理模組

此模組乃提供一般使用者與系統管理者新增、查詢知識文件之功能，此外系統管理者可再進行刪除及修改知識文件等功能。其中，「知識文件新增」功能乃提供一般使用者與系統管理者將知識文件匯入系統資料庫內。「知識文件查詢」功能乃提供一般使用者與系統管理者進行系統資料庫內之知識文件。當中，「知識文件修改」與「知識文件刪除」功能乃提供系統管理者進行修改與維護錯誤之知識文件，進而保持知識文件之正確性；其流程設計概念如圖 4.9 所示。

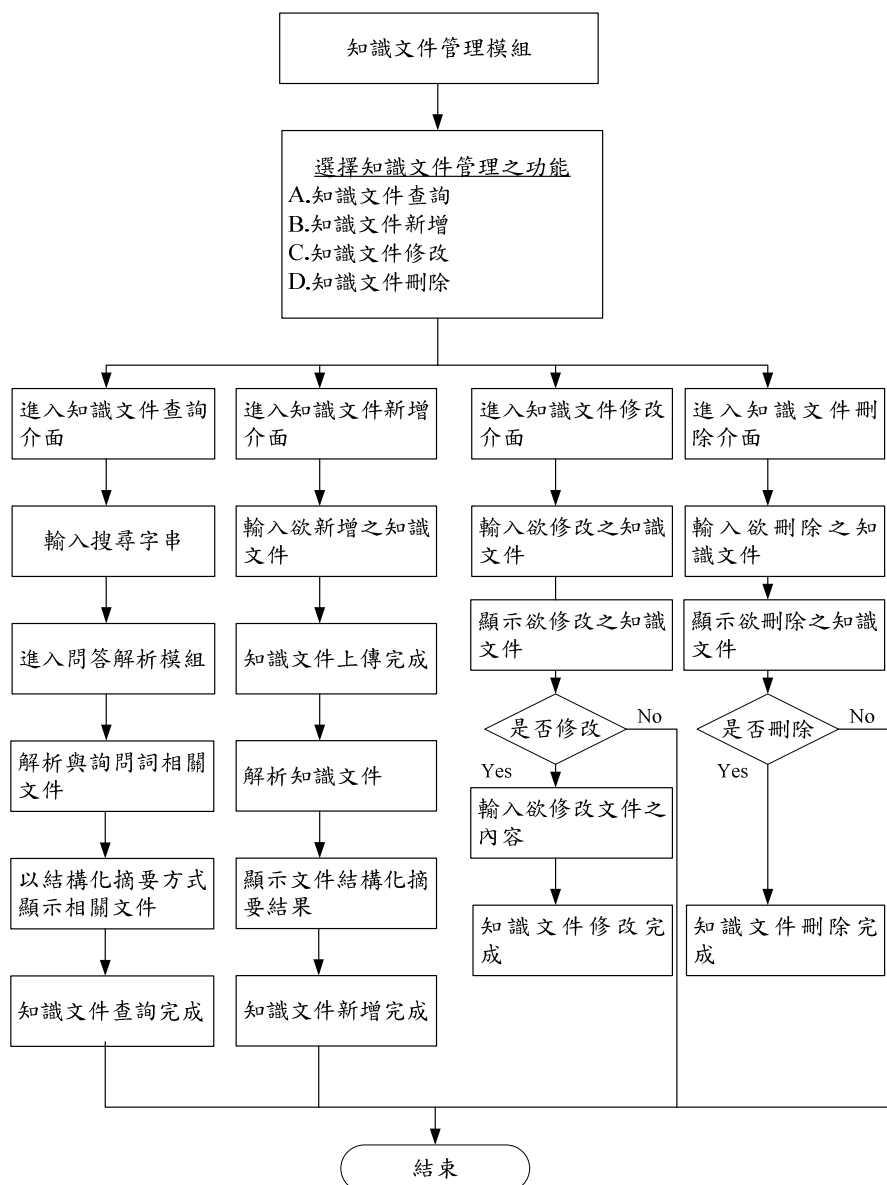


圖 4.9、「知識文件管理模組」功能流程

## 系統參數設定模組

為使系統管理者方便維護各系統相關資料，此模組乃提供系統管理者於線上修改各系統參數資料，包含「簡略結構化摘要權重值設定」、「詳述結構化摘要門檻值設定」及「詢問詞與回覆詞配對門檻值設定」等三功能；其中，「詳述結構化摘要門檻值設定」及「詢問詞與回覆詞配對門檻值設定」功能乃提供系統管理者進行修改與維護錯誤之系統參數，進而保持系統門檻值之正確性；「簡略結構化摘要權重值設定」功能乃提供系統管理者進行修改錯誤之權重值，進而保持系統權重值與預選等級級數之正確性，其流程設計概念如圖 4.10 所示。

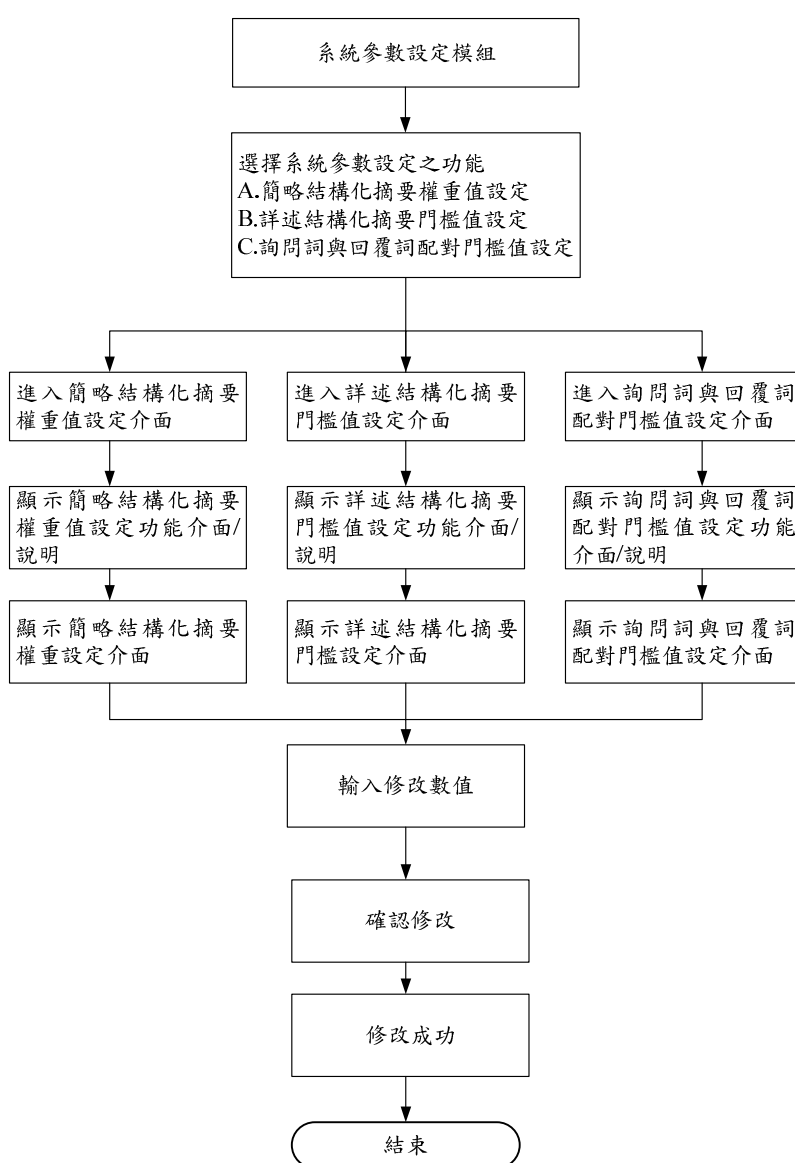


圖 4.10、「系統參數設定模組」功能流程

#### 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知識文件匯入系統，系統即根據系統管理者所設定問答解析之門檻值、簡略結構化摘要權重值與詳述結構化摘要門檻值等各參數設定完成，即可分別進行問答解析與結構化摘要推論，以取得詢問詞與回覆詞配對組合與知識文件結構化摘要，並存入系統資料庫內。上述步驟完成後，使用者即以自我解讀方式進行字串搜尋，透過問答解析即取得相關專業語彙並取得具有關聯性之知識文件，系統乃先解析知識文件之各表達項目並擷取該文件之觀念性語句，並根據系統門檻值與參數推論結構化摘要，以建立簡略結構化摘要與詳述結構化摘要，並透過結構化摘要方式為加強使用者閱讀文件之效果，其系統相關資料之存取與傳遞情形如圖4.11所示。

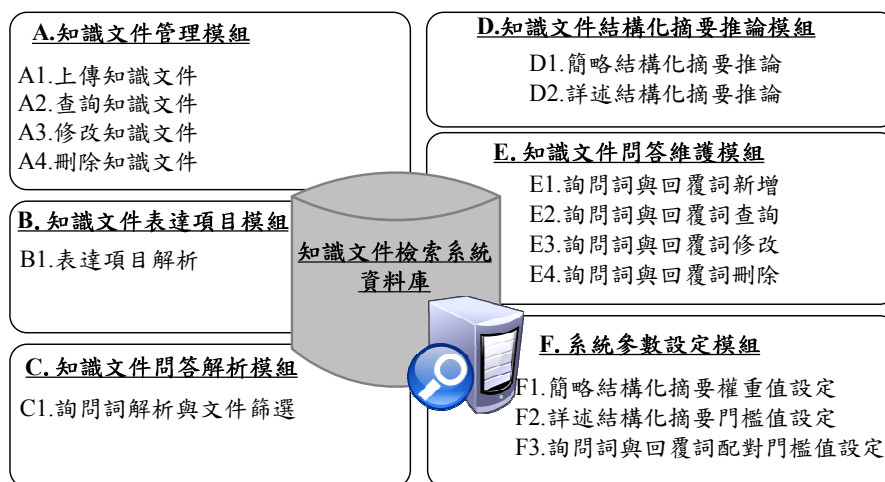


圖 4.11、系統資料流程

#### 4.5 系統開發工具

本系統乃建置於 Microsoft Windows XP作業系統上，並以Microsoft SQL Server2005 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 2005來存取資料庫，輔助系統進行知識文件分析。以下即分別介紹系統開發時所使用之工具。

##### ➤ JSP (Java Server Pages)

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

當使用者透過瀏覽器向伺服器端 (Server) 要求開啟JSP網頁時，架設於伺服器端之JSP引擎乃先將JSP網頁轉譯為Servlet程式，其次再將JSP執行後所產生之文件資料傳送至用戶端 (Client)，並同時顯示執行結果於瀏覽器上。此外，JSP還具有下列特性 (詹權恩，2004)：

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

#### ➤ 關聯式資料庫-Microsoft SQL Server 2005

Microsoft SQL Server 2005為一種關聯式資料庫管理系統，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，這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。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

根據第四章所提出之雛形架構與規劃，本研究乃發展一套結合語意解析與文件摘要技術之知識文件檢索系統，並針對系統中各權限使用者可執行之功能模組詳細介紹，各功能模組之操作說明乃彙整於附錄。本章以勞工安全衛生知識網之改善報告為案例驗證樣本，以分析本研究所提出之方法論與雛形系統的可行性。於第5.1節中以勞工安全衛生知識網之改善報告為案例，闡述系統之應用實效，並於第5.2節中檢驗與評估本知識文件檢索系統之類別推論績效。

### 5.1 系統案例之應用流程

為驗證知識文件結構化摘要系統於實務應用之可行性，本研究乃以勞工安全衛生知識網之改善報告為案例驗證樣本，並以知識文件結構化摘要系統之核心功能模組（包含「知識文件表達項目解析模組」、「知識文件問答解析模組」、「知識文件結構化摘要推論模組」等推論模組），以評估本研究所發展之方法論與開發系統之可行性。首先，系統管理者先將本系統之參數設定完畢，接著，一般使用者乃上傳一份尚未進行文件表達項目解析之知識文件於系統並進行分析，待分析完畢，系統即將此知識文件之簡略結構化摘要與詳述結構化摘要輸出予管理者，管理者再將此次結構化摘要建立結果回饋至系統中，最後使用者即可根據問答解析進行知識文件專業語彙與詢問詞連結，以取得相關性之研究報告，亦可透過知識文件結構化摘要加強文件篩選。完整運作架構如圖5.1所示，以下即進行系統應用情境之詳細說明。

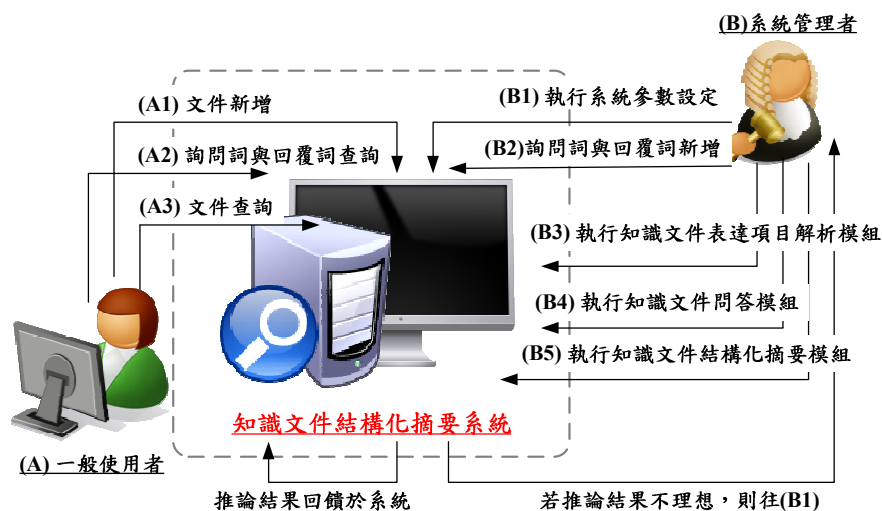


圖 5.1、知識文件結構化摘要系統之應用流程

➤ **系統管理者設定系統參數及語彙庫建立**

系統管理者必須先設定系統內之各項參數與門檻值，以保持系統主功能判斷之正確性。首先，管理者必須設定系統參數之門檻值，之後透過知識文件語彙維護模組，以設定各表達項目之專業語彙，最後管理者即以知識文件維護模組之知識文件新增功能，上傳勞工安全衛生知識網之研究報告訓練樣本至系統中，即完成系統參數設定及語彙庫建立之步驟。

➤ **一般使用者蒐集勞工安全衛生知識網之研究報告**

一般使用者於使用本系統前，先行蒐集未建立結構化摘要之研究報告，之後乃利用本系統幫助其建立研究報告之結構化摘要。使用者於使用本系統之前，蒐集了從勞工安全衛生知識網所提供之知識文件，如研究報告、技術叢書等（如圖 5.2 所示），並儲存文件與彙整如表 5.1 乃列出部分使用者所蒐集之研究報告。



圖 5.2、勞工安全衛生知識網之研究報告

表 5.1、使用者所蒐集之研究報告

文件編號	文件名稱
1	人因工程工作場所改善方法
2	半導體工廠員工肩頸疼痛之研究
3	高齡勞動作業環境適應性研究
4	中高齡職業駕駛工作相關體能評估

## ➤ (A1) 一般使用者上傳尚未推論之知識文件

透過知識文件管理模組之文件新增功能，一般使用者可選擇一份尚未推論判定之文件名稱「點燈測試作業」、文件作者「游志雲」、上傳時間「2012年6月17日」與文件內容「點燈測試作業是作業員將直徑極細小的電燈管放置桌面的兩電極上檢驗燈管是否可點亮之測試。…」等輸入相關資訊，同時瀏覽欲上傳之文件並點選「上傳」按鈕（如圖 5.3 所示）。待上傳完畢後，即呈現資料新增成功訊息，並輸出文件之基本資料如文件名稱、文件作者與文件內容等資訊（如圖 5.4 所示），以提供後續知識文件表達項目分析模組、知識文件問答解析模組與知識文件結構化摘要使用。

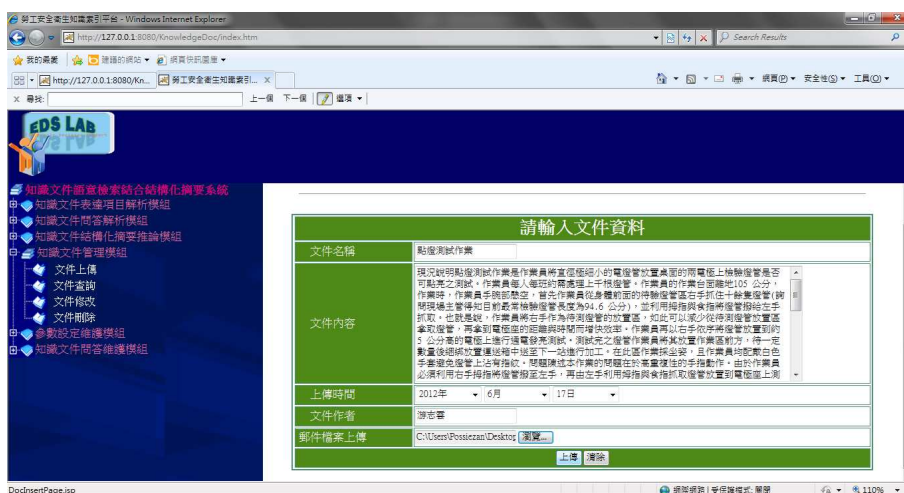


圖 5.3、知識文件管理模組-輸入文件資料



圖 5.4、知識文件管理模組-文件新增成功

➤ **(B1) 系統管理者設定系統參數**

待使用者新增文件後，系統管理者則需設定系統內之各項參數與門檻值，以保持系統主功能判斷之正確性。管理者必須點選系統參數設定模組以設定系統參數之門檻值與權重值，即完成系統參數設定步驟（如圖 5.5 與圖 5.6 所示）。



圖 5.5、系統參數設定模組-系統參數檢視



圖 5.6、系統參數設定模組-系統參數設定

➤ **(B2) 系統管理者新增詢問詞與回覆詞配對組合**

系統管理者透過「知識文件問答維護模組」新增系統內之詢問詞與回覆詞配對組合，以保持系統問答解析推論之正確性。經由「詢問詞與回覆詞新增」功能輸入詢問詞「肌肉」並根據肌肉搜尋取得回覆句「一般抬舉重物對於腰椎會形成…」以及該回覆句

主要回覆詞「人工抬舉」、「腰椎」、「搬運」，待輸入五筆回覆句以及十五個回覆詞等相關資訊，並點選「確定」按鈕（如圖 5.7 所示），以進行配對分析結果與詢問詞「肌肉」具有相關回覆詞包含「搬運」、「人工抬舉」、「腰椎」等詞彙（如圖 5.8 與圖 5.9 所示），系統管理者即可點選「新增配對結果」將上述配對組合以完成新增作業，匯入詞彙於系統資料庫內（如圖 5.10 所示），使系統精確判定詞彙語意達到問答效果。



圖 5.7、詢問詞與回覆詞輸入畫面



圖 5.8、詢問詞與回覆詞配對結果畫面(1)



圖 5.9、詢問詞與回覆詞配對結果畫面(2)



圖 5.10、詢問詞與回覆詞新增成功畫面

➤ (B3) 系統管理者執行知識文件表達項目解析

當使用者完成文件上傳動作，且系統管理者亦完成系統參數設定與詢問詞回覆詞新增，系統管理者即可執行主功能以進行知識網之問答解析與建立該研究報告之結構化摘要。首先系統管理者於知識文件表達項目解析模組先以「作業」為查詢條件，進行文件篩選並取得相關文件(如圖 5.11 所示)，隨後即可勾選「庫房搬運作業」之「是否解析」，並點選判定按鈕以進行知識文件表達項目解析(如圖 5.12 所示)，系統即顯示該研究報告內容以及各表達項目之觀念性語句結果，如圖 5.13 至圖 5.15 所示，表達項目「作業身分」之觀念性語句有「因此需要兩個作業人員合力抬舉方能搬動儲存箱。」、「對於作業人員的肌肉骨骼，」等；表達項目「作業環境」之觀念性語句有「每個推車約堆置五

到六個儲存箱以增加庫房存放空間利用率，」、「此時便需要兩個作業人員合力將所有上層的儲存箱抬舉並放置在旁側地板上，」等；表達項目「作業行為」之觀念性語句有「負責此區的作業人員在一個工作天之內可能搬運總重達一噸以上的物料，」等；表達項目「傷害成因」之觀念性語句有「問題陳述此作業區人因工程的主要問題乃在於重複搬運的動作。」、「關於前述重複搬運動作所造成的人因工程問題，」等；表達項目「改善方法」之觀念性語句有「可以很大程度地減少作業人員的重複搬運動作，」、「同時作業人員也可不必以彎腰的不良工作姿勢置放或抬舉暫時擺放於地板的上層儲存箱，」等。



圖 5.11、知識文件表達項目解析查詢介面



圖 5.12、知識文件表達項目解析查詢介面



圖 5.13、知識文件表達項目解析結果(1)

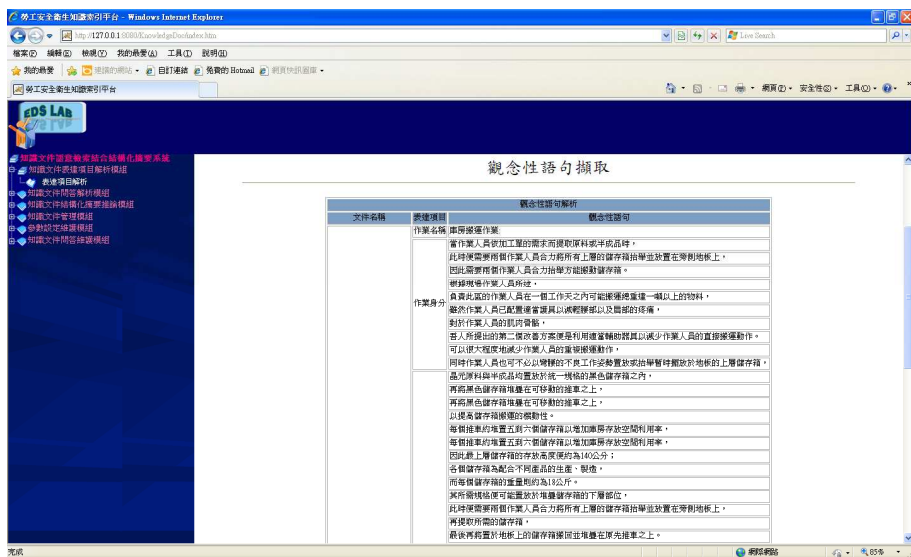


圖 5.14、知識文件表達項目解析結果(2)





圖 5.15、知識文件表達項目解析結果(3)

➤ **(B4) 系統管理者執行知識文件問答解析**

系統管理者可於知識問答解析模組之問答解析推論中，輸入詢問句「我手腳很酸痛」作為搜尋字串進行查詢，透過詢問詞解析與篩選文件功能中即可取得該詢問句之主要詢問目的為「酸痛」，根據詢問詞「酸痛」經過詢問詞、回覆句與回覆詞配對解析，以取得「酸痛」之配對組合包含「營造業」、「年紀」、「長時間」等（如圖 5.16 與圖 5.17）。藉由詢問詞與回覆詞配對組合「酸痛：營造業、年紀、長時間」，即可與文件之關鍵字詞進行餘弦相似性分析並根據門檻值（平均數、中位數、四分位數與自行定義）以篩選並下載文件（如圖 5.18 與圖 5.19）。



圖 5.16、知識文件問答解析-詢問詞解析目的

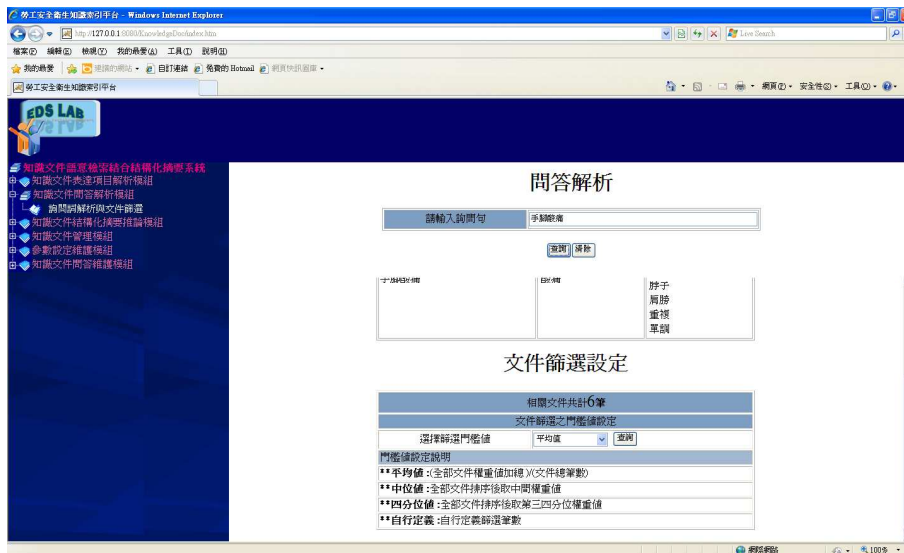


圖 5.18、知識文件問答解析-文件篩選結果

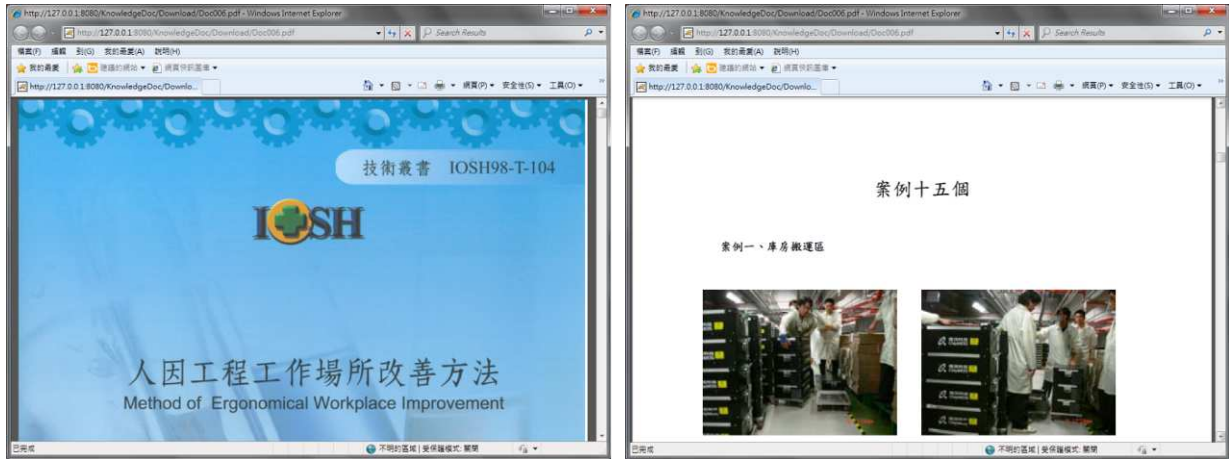


圖 5.19、知識文件問答解析-文件下載

### ➤ (B5) 系統管理者執行知識文件結構化摘要推論

系統管理者於結構化摘要推論中可進行簡略結構化摘要功能與詳述結構化摘要功能，首先於知識文件解析模組輸入研究報告之基本資料例如「人因」，查詢該文件是否存在（如圖 5.20 所示），隨後即可勾選其「是否解析」，並點選判定按鈕以進行知識文件結構化摘要推論（如圖 5.21 所示），以取得該文件中各觀念性語句之向心性分數與結構性分數，如圖 5.22 所示，「同時作業人員也可不必以彎腰…」之向心性分數為「3」、結構性分數為「3」，最後根據加權取得加總分數為「12」，進而根據權重值根據「文件名稱」、「作業名稱」、「作業行為」、「傷害成因」與「改善方法」等表達項目建立簡略結構化摘要（如圖 5.23 所示）。於詳述結構化摘要推論，則取得於文件中各詞彙之起迄計算結果，如圖 5.24 左半部所示，「庫房」字詞起之總和為「295」、迄之總和為「89」，以取得詞彙鏈判斷係數為「3.315」；「主要功能」字詞起之總和為「91」、迄之總和為「5」，以取得詞彙鏈判斷係數為「18.2」等，並以進行字詞組合形成詞彙鏈先後順序為「庫房」、「搬運」、「作業」、「主要功能」、「...」、「主要問題」、「重複搬運」等進而根據「作業描述」、「問題描述」以及「改善方式」取得詳述結構化摘要（如圖 5.24 右半部所示）。最後，管理者即可將本次分析紀錄儲存至系統中以供使用者查詢。

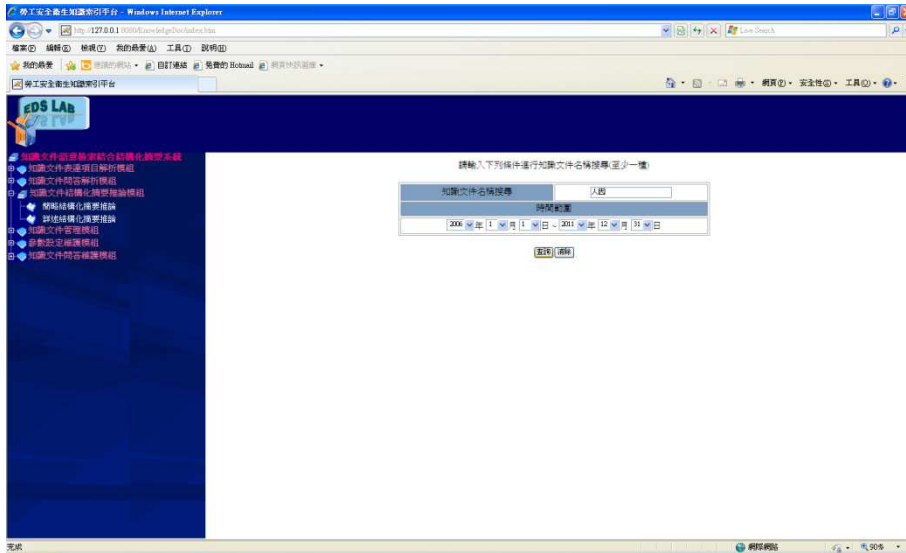


圖 5.20、知識文件結構化摘要-查詢文件



圖 5.21、知識文件結構化摘要-文件查詢結果



圖 5.22、知識文件結構化摘要-簡略結構化摘要解析



圖 5.23、知識文件結構化摘要-簡略結構化摘要呈現



圖 5.24、知識文件結構化摘要-詳述結構化摘要呈現

➤ (A2) 一般使用者查詢詢問詞與回覆詞

使用者可於知識文件問答維護模組之「詢問詞與回覆詞查詢」功能，查詢已新增之詢問詞與回覆詞。系統乃提供條件查詢欄位於使用者輸入查詢條件「肌肉」並以「詢問詞」為查詢字串，同時限制文件時間範圍於「2007年1月1日」至「2012年12月31日」之間進行篩選(如圖 5.25 所示)，即可取得詢問詞「肌肉」以及相關回覆詞「搬運」、「抬舉」等六筆詢問詞與回覆詞組合，以及各組合係數「1.0」、「0.27」、「0.27」等(如圖 5.26 所示)。

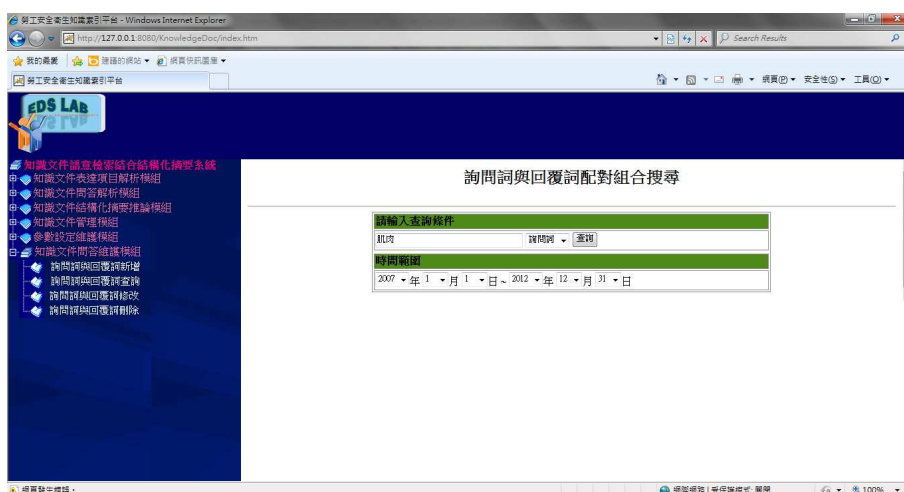


圖 5.25、知識文件問答維護-詢問詞與回覆詞查詢畫面



圖 5.26、知識文件問答維護-詢問詞與回覆詞查詢結果畫面

### ➤ (A3) 一般使用者查詢知識文件

使用者可透過知識文件管理模組查詢已上傳之文件內容，使用者可輸入查詢條件「人因」並以「文件名稱」為查詢字串，同時限制文件時間範圍於「2007年1月1日」至「2012年12月31日」之間進行篩選，即可取得「人因工程工作場所改善方法-庫房搬運作業」、「人因工程工作場所改善方法-車椅組裝室作業」與「人因工程工作場所改善方法-氣泡料攪拌作業」等六筆文件（如圖 5.27 與圖 5.28 所示）。當該文件已完成推論之判定，使用者即可點選「已分析」連結功能，並以彈跳視窗方式顯示該文件之文件表達項目分析結果、簡略結構化摘要結果與詳述結構化摘要推論結果（如圖 5.29 與圖 5.30 所示）。



圖 5.27、知識文件管理模組-文件查詢



圖 5.28、知識文件管理模組-文件查詢結果(1)

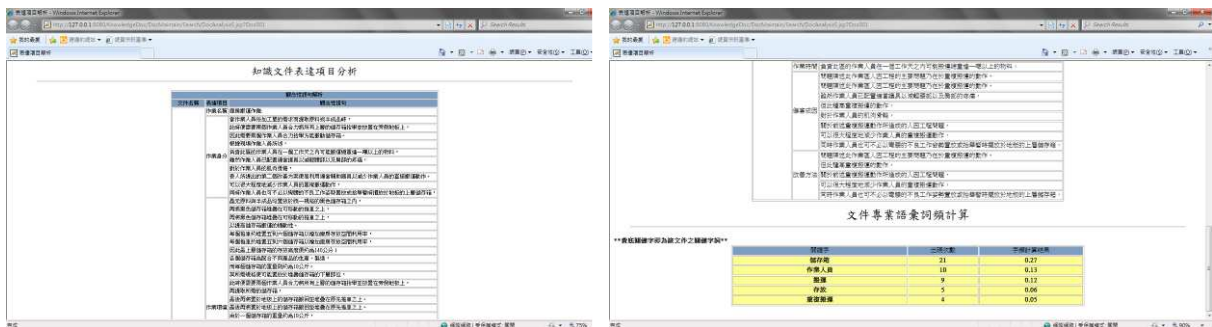


圖 5.29、知識文件管理模組-文件查詢結果(2)

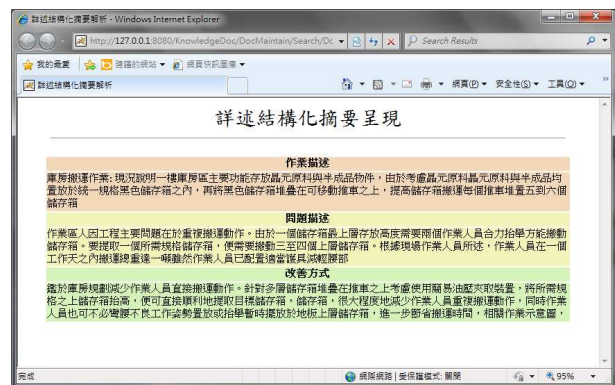
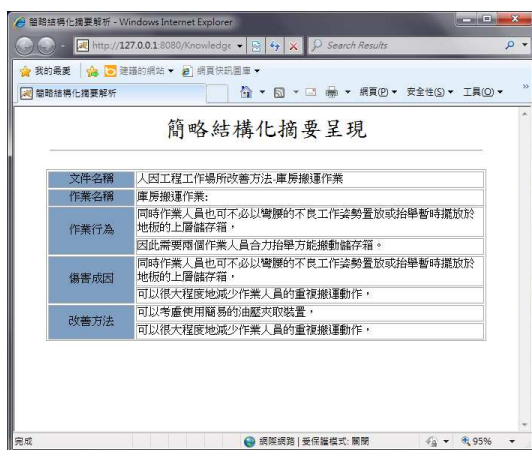


圖 5.30、知識文件管理模組-文件查詢結果(3)



## 5.2 系統驗證與評估

本研究乃以「結合語意解析與文件摘要技術之知識文件檢索推論」為基，開發一套「結合語意解析與文件摘要技術之知識文件檢索系統」，並以勞工安全衛生知識網作為應用情境，將研究報告與技術叢書作為知識文件並以「知識文件問答解析」與「知識文件結構化摘要」等主題進行系統績效驗證。本研究所提出之系統乃協助知識搜尋者能以搜尋字串方式進行文件搜尋，並可以結構化摘要呈現方式輔佐知識搜尋者篩選文件。是故，於驗證之前，本研究乃先行勞工安全衛生知識網取得研究報告與技術叢書之案例作業，並解析各文件之構成要素（即表達項目）與文件內容主要表達順序性等，以作為系統驗證與分析之基礎。以下乃針對本系統各主題之驗證方式（即「驗證資料說明」、「驗證方式說明」與「驗證指標定義」）依序說明，以驗證「知識文件檢索系統」績效與準確性。

### 驗證資料取得

針對「知識文件檢索系統」之驗證，本研究乃先行勞工安全衛生研究所（勞研所）之網際論壇取得知識搜尋者之詢問句，以作為知識文件問答解析之驗證資料，同時亦至勞工安全衛生知識網取得研究報告與技術叢書，以作為知識文件結構化摘要之驗證資料。其各驗證資料乃說明如下：

針對知識文件問答解析之驗證資料，透過勞研所之網際論壇，知識搜尋者可提出問題，並透過該網站檢索方式以取得相關回覆句。因此，本研究乃針對勞研所取得知識搜尋者之詢問句，並彙整成「知識搜尋者詢問句資料表」（如表5.2所示）。此外，勞研所亦針對專有名詞乃建立雙語詞彙，並說明該詞彙之解釋與意義，本研究即以勞研所所提供之雙語詞彙之解釋作為驗證實際檢索回覆詞（如表5.3所示）。

表 5.2、知識搜尋者詢問句資料表

題號	詢問句	知識搜尋者	詢問時間
1	作業作業環境測定項目	佛曰不可說...	2009/12/4
2	關於升降機	japaneseky	2007/10/17
3	請問有關緊急照明的問題	何志強	2010/10/11
4	請問有沖身洗眼器的檢查表嗎	loveven.tw	2010/9/9
5	想請問各位先進,關於[惰性氣體]和[易燃氣體],	流星	2012/2/1
6	為避免局部振動危害,在工程方面有那些方法	water	2009/11/16
7	請問三公噸以下固定式起重機	Feather	2010/3/11
8	根本找不到母索...安全帶要何用...	Isandy5300oh	2009/4/5
9	請問:事故調查之原因分析	ken	2009/7/30
10	如果與鄰地有落差...該設置欄杆,有法規...	隨遇而安	2012/5/18

表 5.3、勞工安全衛生研究所定義之雙語詞彙資料表

類別	中文詞彙	English	簡稱	附註說明	出處	相關詞
A	擦傷	ABRASION	--	覆蓋於身體的表面組織（如皮膚或黏膜）被擦掉或刮掉	美國安全工程師協會	表面組織、皮膚、黏膜、擦掉、刮掉
A	容許負載	ALLOWABLE LOAD	--	請參考 Load Weight（承載重量）。	美國安全工程師協會	承載重量
M	機械防護	MACHINE GUARDING	--	在機械裝上設備或機具以消除操作機械所產生的危害。參考 Barrier Guard（防護圍柵）。	美國安全工程師協會	機具、防護圍柵

針對知識文件結構化摘要之驗證資料，本研究乃先解析勞工安全衛生知識網各報告或技術叢書中案例作業，以維護於系統資料庫中以完成系統驗證前資料建置作。透過研究報告與技術叢書之案例作業以解析各文件之構成要素（即表達項目）與文件內容主要表達順序性等，藉由案例作業可發現，其內容乃先描述「現況說明」即作業行為、作業環境與作業運作流程細述，進而「問題陳述」導致傷害成因，最後再提出有效之「改善方法」，藉由案例作業分析即可將各表達項目與相關語彙彙整成「知識文件表達項目資料表」（如表5.4所示）。最後乃將上述資料表作為「知識檢索系統」之訓練資料，其驗證資料分析流程如圖5.31。

表 5.4、知識文件表達項目資料表

表達細項		礦泉水運送作業	庫房搬運作業
作業領域		--	--
作業名稱		礦泉水運送作業	庫房搬運作業
作業身分	作業員性別	--	--
	作業員年齡	--	--
作業環境	設備佈置	輸送帶	--
		--	--
	工具介紹	拖車、木箱、水桶	黑色儲存箱、推車
作業行為	作業目的	主要工作是將輸送帶上已完成填灌作業的礦泉水桶搬運至拖車上的木箱中	主要功能為存放晶元原料與半成品物件，以為後續製造、加工之用
	作業描述	礦泉水運送作業的主要工作是將輸送帶上已完成填灌作業的礦泉水桶搬運至拖車上的木箱中，直至拖車上的所有木箱均已裝滿礦泉水桶，再將拖車推上貨車運送至各部門	將黑色儲存箱堆疊在可移動的推車之上，以提高儲存箱搬運的機動性。一般而言，每個推車約堆置五到六個儲存箱以增加庫房存放空間利用率
	作業姿勢	抬舉、搬運、高舉、彎腰	搬運、抬舉
	專業動詞	裝滿礦泉水的水桶重量約為 20 公斤	儲存箱的重量約為 18 公斤一個工作天之內可能搬運總重達一噸以上的物料
作業時間	作業次數/天	一天必須至少搬運 200 桶、有時候則可能多達 500 桶	
	作業時間/次	--	--
	作業距離/次	輸送帶與拖車之間的最近距離為 1.5 公尺、輸送帶與拖車上木箱的各別距離便在 1.7 公尺到 3.2 公尺之間	--
傷害成因	傷害因素	施力過大	重複搬運
	傷痛部位	手臂部分、背部	腰部、肩部
	改善目的	有效地舉高或放下礦泉水桶並改變其水平位置，平順地將礦泉水桶準確放置於拖車上的木箱中	大程度地減少作業人員的重複搬運動作，同時作業人員也可不必以彎腰的不良工作姿勢置放或抬舉暫時擺放於地板的上層儲存箱，可以進一步節省搬運時間，增加工作效率
改善方法	改善工具	象鼻子真空省力裝置	油壓夾取裝置
	改善流程	只要以此裝置的吸物盤口含納礦泉水桶的頸部出口處，作業人員便可以手指頭調整氣閥的流量	考慮使用簡易的油壓夾取裝置，將所需規格之上的儲存箱抬高，便可直接順利地提取目標儲存箱，完全避免搬動額外的儲存箱
	改善評估	負荷明顯降低部位	負荷明顯降低部位
	改善費用	--	--

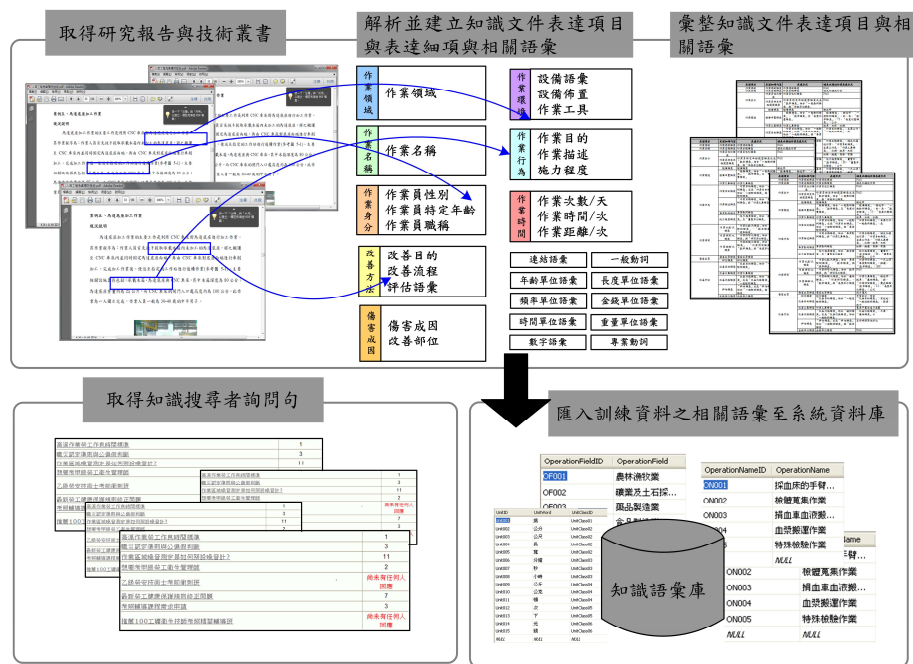


圖 5.31、驗證資料蒐集與建置

### 5.2.1 知識文件檢索系統整體驗證方式

針對「知識文件檢索系統」本研究乃提出二個課題以進行系統績效驗證，其課題分別為「(A)知識文件問答解析」與「(B)知識文件結構化摘要推論」等並分別設計指標以進行系統驗證分析。以下乃針對各課題分別細述與說明驗證設計、指標定義及驗證績效。

#### (A) 識文件問答解析驗證方式說明

於「知識文件問答解析」乃進行搜尋字串與知識文件語意關聯檢索，其目的透過詢問句以取得相關文件。於此課題本研究乃以勞工安全衛生研究所之網際論壇中，以知識搜尋者之詢問句作為測試資料，並以勞研所制定之雙語詞彙作為實際檢索回覆詞，以驗證知識文件問答解析之績效。此外，本研究乃規劃兩階段之系統驗證，於系統驗證第一階段乃於800筆詢問句與回覆句組合中，隨機選取200筆組合作為訓練資料，並利用訓練階段所取得之文件相關語意詞，再從中隨機挑選10句詢問句，以作為系統測試資料，藉以觀察系統問答解析是否符合實際相關字詞，以確認本研究提方法論之正確性。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（即系統測試階段），於此階段乃分6個週期（每週期皆匯入100筆不重複之詢問句與回覆句配對組合，共計600筆組合），藉由持續匯入以分析系統於不同訓練測試資料下之解析效果，於各週期中乃

利用前述10句詢問句重新進行知識文件問答解析，以分析系統之長期學期趨勢。

### (A-1) 知識文件問答解析評估指標定義

藉由本系統所推論之回覆詞與實際檢索回覆詞進行驗證系統準確性，於此階段績效指標包含「回覆詞召回率」與「回覆詞準確率」，其參數定義與計算方式如公式(5.1)至公式(5.2)所示：

$R_i(QA)$	第i筆問答解析之回覆詞召回率
$A_i(QA)$	第i筆問答解析之回覆詞準確率
$N_i(QA)$	第i筆實際所含之回覆詞數量
$T_i(QA)$	第i筆經系統解析所得之回覆詞數量
$C_i(QA)$	第i筆實際所含之回覆詞與第i筆經系統解析所得之回覆詞交集數目

問答解析之回覆詞召回率 ( $R_i(QA)$ ) 即表示經由系統問答解析後能取得正確回覆詞之解析能力，亦為實際問答所含之回覆詞數量 ( $N_i(QA)$ ) 與經系統解析所得之交集數目與系統解析所得之回覆詞之比例 ( $C_i(QA)$ )。

$$R_i(QA) = \frac{C_i(QA)}{N_i(QA)} \quad (5.1)$$

問答解析之回覆詞正確率 ( $A_i(QA)$ ) 即表示經由問答解析後能取得正確問答回覆詞之正確能力，亦為經系統解析所得之回覆詞數量 ( $T_i(QA)$ ) 與經系統解析所得之交集數目與系統解析所得之回覆詞之比例 ( $C_i(QA)$ )。

$$A_i(QA) = \frac{C_i(QA)}{T_i(QA)} \quad (5.2)$$

### (A-2) 知識文件問答解析驗證結果與分析

本研究乃將知識文件問答解析驗證結果分為「第一階段驗證結果分析」與「第二階段驗證結果分析」兩大項目。於「第一階段驗證結果分析」項目中，乃測試詢問句與回覆句以測試系統進行「知識文件問答解析」之正確性，以瞭解系統初期知識文件驗證問

答解析之相關績效。而於「第二階段驗證結果分析」項目中，本研究乃於各測試週期新增訓練知識文件（即訓練詢問句與回覆句組合），以評估不同訓練資料量下系統進行「知識文件問答解析」功能之績效。以下即以「第一階段驗證結果分析」、「第二階段驗證結果分析」與「驗證結果整體分析」等三主題說明本研究之驗證結果。

於第一階段系統驗證中，針對知識文件問答解析部分乃由 800 句詢問詞與回覆句中，隨機取 200 份作為問答解析之訓練資料，並逐一匯入系統中，以作為第一階段驗證之基礎訓練資料。以下即針對各項指標說明系統驗證過程，並分析系統驗證之結果。

### (A-I) 知識文件問答解析第一階段驗證結果分析（10 句詢問句）

在 800 份文件作為訓練資料之基礎下，將其問答資料包含受測詢問句、透過詢問句取得之回覆句等資料匯入系統（如表 5.5 所示），再隨機抽取 10 句詢問句作為測試資料，於第一階段測試結果乃彙整於表 5.6 所示。以表 5.6 為例，根據詢問句「請問有沖身洗眼器的檢查表嗎」，根據勞研所定義之相關回覆詞為「照度」與「光通量」等相關詞，而於系統推論之相關詞數共兩筆，其正確推論相關詞數為一筆，因此回覆詞之召回率與正確率分別各為 50%。此外，回覆詞召回率與回覆詞正確率之分佈趨勢如圖 5.32 與圖 5.33 所示。

表 5.5、知識文件問答解析之詢問句與回覆句組合

詢問句	相關回覆句	回覆句數
台灣目前製造業意外事故調查步驟	近年來發生多起建築工程斜籬安裝時作業勞工墜落的意外事故，	5
	安裝時勞工常會站到斜籬上施作而無適當防護，墜落事故時有所聞。	
	提供適當之安全防護措施，並督促勞工佩帶個人安全防護具。	
	因腳踩踏於斜籬骨架，而骨架發生鬆脫致罹災者發生墜落，罹災者雖繫安全帶但並未勾掛於母索上。	
	應設置防止物體飛落之設備，並供給安全帽等防護具，使勞工戴用，其中斜籬即為防止物體飛落之設備。	
疲勞程度規範	疲勞是當前各種行業非常重要的職業安全與衛生議題之一，疲勞可能影響一位勞工在工作時注意安全的的能力，嚴重削弱勞工本身的判斷能力，進而造成產能的損失或是意外災害的發生	3
	許多人經常認為司機在開車時入睡即表示其疲勞。	
	疲勞通常是用來描述類似「想睡」、「疲倦」或者「筋疲力盡」的一種經驗術語，也是一種生理和心理的經驗感受疲勞的產生對職業駕駛是特別危險的，因為疲勞症狀之一就是本身會對於自己疲倦的程度失去判斷能力	
酒精濃度含量多少會導致意外發生	當血液中酒精含量達 0.1% 時，人的動作協調、視覺、言談及平衡 會受損，出現中毒現象。	3
	當血液中酒精含量達 0.5% 時，神經生理平衡會嚴重受損而且失去意識，廚師罹患高比率的口腔癌 與罹患高比率的咽喉癌可能是由於高量飲酒所引起。	

表 5.6、知識文件問答解析第一階段之受測詢問句結果分析

受測詢問句	相關詞 1	相關詞 2	相關詞 3	實際 相關 詞數	推論 相關 詞數	正確 推論 數	召回率	正確率
關於升降機	工作情況	工作場所	-	2	1	0	0.00%	0.00%
請問有關緊急照明的問題	安全法規	電扶梯	-	2	3	1	50.00%	33.33%
請問有沖身洗眼器的檢查表嗎	照度	光通量	-	2	2	1	50.00%	50.00%
想請問各位先進,關於[惰性氣體]和[易燃氣體],	沖洗水	沖洗軟管	-	2	2	1	50.00%	50.00%
為避免局部振動危害,在工程方面有那些方法	不活性	氦、氖、 氬、氪、 氙	-	2	1	0	0.00%	0.00%
請問三公噸以下固定式起重機	壓路機	重型營建車輛	-	2	2	1	50.00%	50.00%
根本找不到母索...安全帶要何用...	鉸鏈	手搖桿	動力裝置	3	2	1	33.30%	100.00%
請問：事故調查之原因分析	固定	懸吊	碰撞	3	2	1	33.00%	50.00%
如果與鄰地有落差,該設置欄杆,有法規規定高差多大以上該設置欄杆嗎?	作業場所	實際原因	近似原因	3	1	0	0.00%	0.00%
如果與鄰地有落差,該設置欄杆,有法規規定高差多大以上該設置欄杆嗎?	護欄	扶手	-	2	1	1	50.00%	100.00%
平均值							31.63%	43.33%

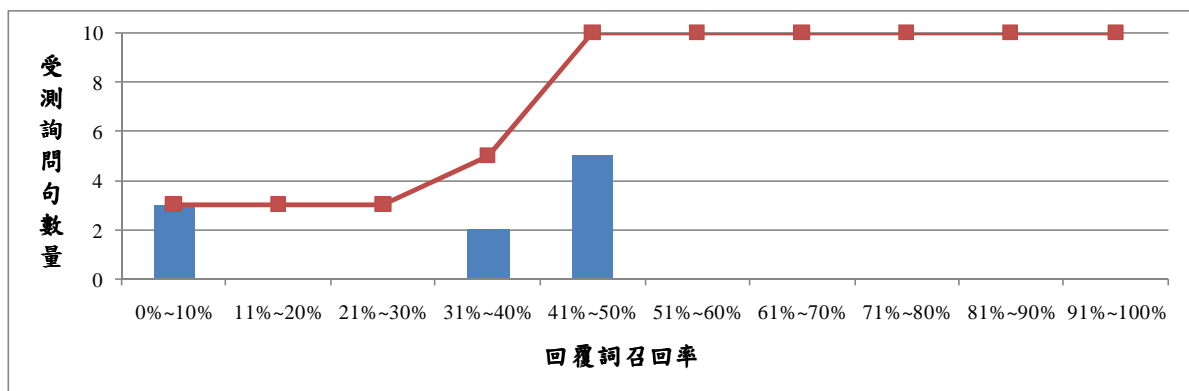


圖 5.32、知識文件問答解析第一階段之回覆詞召回率之分布趨勢

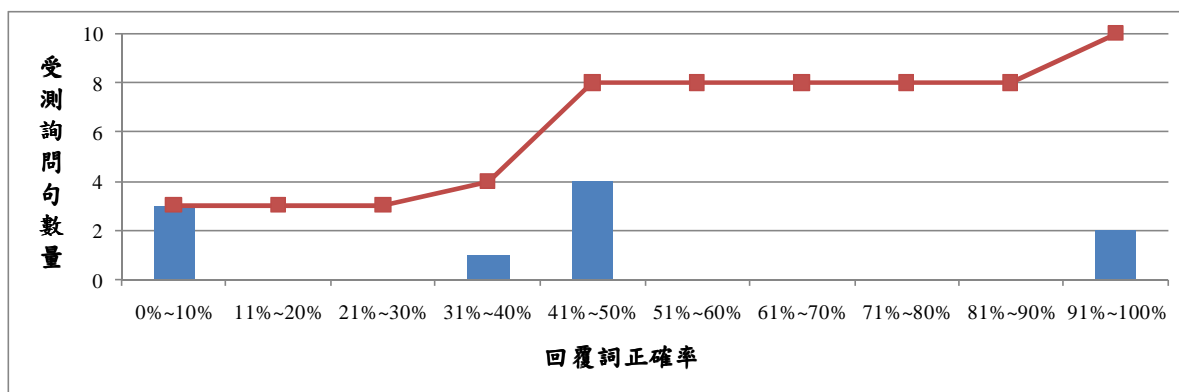


圖 5.33、知識文件問答解析第一階段之回覆詞正確率之分布趨勢

透過表 5.6、圖 5.32 與圖 5.33 統計結果以得知，系統於第一階段之系統驗證結果中，回覆詞召回率約落於 33.30% 與 50% 間，其平均為 31.63%，而回覆詞正確率約落於 33.33%、50% 與 100% 間，其平均為 43.33%。整體而言，其回覆詞召回率與正確率績效，尚無法準確判斷知識文件之問答。

#### (A-II) 知識文件問答解析二階段驗證結果分析 (10 句詢問句)

第二階段系統驗證之作法乃以第一階段驗證中，針對知識文件問答解析部分乃將第二階段乃分 6 個週期(每週期皆匯入 100 筆不重複之詢問句與回覆句配對組合，共計 600 筆組合)，並以第一階段驗證時所選取之 10 筆詢問句與回覆句組合重新進行系統績效測試，以瞭解系統於不同訓練資料數量基礎下進行知識文件之問答解析及結構化摘要之績效變化趨勢，進而分析本系統之學習成長能力。此階段中各週期驗證知識文件問答解析指標受測詢問句之資料相關結果可整理如表 5.7，而各驗證週期之績效分佈趨勢如圖 5.34 所示。



表 5.7、知識文件問答解析績效彙整

知識文件問答解析		各週期知識文件問答解析驗證—詢問句與回覆句組合數量							平均
		第一階段	第二階段						
		第一週期	第二週期	第三週期	第四週期	第五週期	第六週期	第七週期	
		200 筆	300 筆	400 筆	500 筆	600 筆	700 筆	800 組	
召回率	平均值	31.63%	48.27%	61.63%	70.00%	78.33%	86.67%	86.67%	66.17%
	標準差	22.84%	9.58%	22.34%	21.94%	23.64%	21.94%	21.94%	20.60%
	成長率	-	16.64%	13.36%	8.37%	8.33%	8.34%	0%	9.17%
正確率	平均值	43.33%	50.00%	58.33%	66.67%	71.67%	84.17%	84.17%	65.56%
	標準差	37.02%	7.86%	16.20%	19.25%	20.86%	21.68%	21.68%	20.63%
	成長率	-	6.67%	8.88%	7.79%	5.00%	12.50%	0%	6.81%

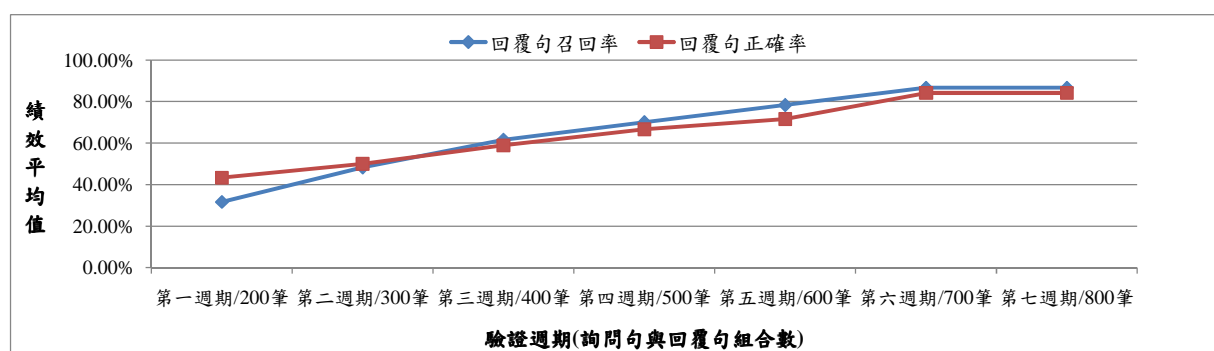


圖 5.34、知識文件問答解析各驗證週期之績效分佈趨勢

由表 5.7 與圖 5.34 可得知，以每週期增加 100 份訓練用詢問與回覆句組合為單位，平均每週期知識文件問答解析召回率與正確率之整體平均成長率分別為 9.17% 及 6.81%；故可知知識文件問答解析專業語彙與一般詞彙關聯性能力良好，且知識文件問答解析乃具學習能力。

### (A-III) 知識文件問答解析整體驗證結果分析

在知識文件問答解析第一階段驗證中，當訓練詢問句與回覆句組合筆數為 200 筆時，其「正確率」與「召回率」分別為 31.63% 及 43.33%，由數據可得知於第一階段知識文件問答解析之績效欠佳，然而，於第二階段驗證中，系統乃隨著不同週期及訓練組合句之增加下，各項數據呈現顯著之成長，於第七週期時，知識文件問答解析「正確率」與「召回率」可提升至 86.67% 及 84.17%。有鑑於上述之數據，可證實本研究所建置之「知識文件問答解析模組」乃實具可行性，並可推斷若未來不斷訓練更多詢問句與回覆句組合時，其績效亦將持續成長。

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5.8所示。由表5.8整理結果可得知，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六週期內呈現收斂狀態，因此，以本研究所選定之驗證資料（勞工安全衛生研究所網際論壇之詢問句與回覆句組合）為例，當系統使用約700筆訓練詢問句與回覆句組合時，可使系統各項推論指標績效提升至85%之水準。

表 5.8、知識文件問答解析綜合兩階段之驗證績效彙整

驗證指標	整體平均值	收斂週期	收斂前每週期平均成長率	整體每週期平均成長率
知識文件問答解析召回率	66.17%	第六週期	11.01%	9.17%
知識文件問答解析正確率	65.56%	第六週期	8.17%	6.81%

整體而言，由第一階段與第二階段之驗證分析中，可得知當使用者進行「知識文件問答解析」時，訓練資料與週期乃影響系統績效之主要原因，系統之績效將隨著不同週期及訓練量之增加，呈現持續成長之狀況，並於最終達到穩定且良好之績效水平。故本研究所建置之「知識文件問答解析模組」確實能幫助知識搜尋者取得專業語彙與一般詞彙之關聯性。

### (B) 知識文件結構化摘要驗證方式說明

於「知識文件結構化摘要效能分析」乃進行知識文件制式結構化摘要推論，其目的乃透過結構化摘要以呈現知識文件表達涵義。於此課題，本研究乃以每份知識文件之案例作業作為一份文件，並以二十位知識搜尋者作為測試者，以了解結構化摘要呈現滿意度與表達準確性。針對測試者部分，本研究乃邀請二十位大專院校學生（十位女性與十位男性）作為知識搜尋者以進行測試，當中受測同學皆屬於同一系所，並依照年級各取兩名，即大學一年級至四年級之男性與女性各一名，共計八名；研究所共取一名男性與一名女性，總計二十名進行測試。首先，本研究先行提供文件之原文內容以供受測者閱讀，藉由閱讀方式使測試者了解文件之內容與大綱。待閱讀完後，本研究再將本系統所推論之階段性成果給予測試者閱讀，根據測試者之閱讀主觀感受以進行滿意度評級。

對此，本研究乃規劃兩階段之系統驗證，於系統驗證第一階段乃於500份知識文件中，隨機選取100筆作為訓練資料，並利用訓練階段所取得之文件結構化摘要，再從中

隨機挑選15份知識文件（如表5.9及表5.10所示），以作為系統測試資料，藉以觀察系統結構化摘要推論之使用者滿意度與摘要呈現方式，以確認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（即系統測試階段），於此階段乃分5個週期（每週期皆匯入80筆不重複之知識文件，共計400筆組合），藉由持續匯入以分析系統於不同訓練測試資料下之解析效果，並從中了解知識文件結構化摘要之長期學習趨勢。

表 5.9、知識文件結構化摘要驗證之測試資料（其中 8 份）

文件名稱	作業名稱	文件內容	文件字數
人因工程現場不良工作姿勢改善績效評估研究-貝果運送作業	貝果運送作業	現況說明，此作業主要是將烘烤完成的貝果經由輸送帶運送到再加工區。輸送帶連接烘烤機和貝果再加工區，在輸送過程中，輸送帶必須先約升約 2~3 公尺之後再下降，以增加麵包冷卻的時間。然而，此輸送帶並非專為貝果生產線特別設計，在爬升運送時，貝果…	905
工作現場人因工程危害預防效益研究-氣泡料攪拌作業	氣泡料攪拌作業	氣泡料攪拌作業:現況說明泡料室攪拌作業，通常是一人獨立完成所有工作。泡料室面積約七坪左右，室內主要擺設為攪拌機 2 台（圖 9-1）及放置泡料粉的貨架（圖 9-2），攪拌機圓桶高 94 公分，直徑約 89 公分，空…	913
人因工程工作場所改善方法-切塊作業	切塊作業	切塊作業分切雞塊作業是作業員將經由輸送帶運送到面前的盤裝雞肉拿起，取出雞肉放在棧板上，接著再利用大刀將半雞分為三大塊，再逐一剝成小塊狀，整齊排列於盤內，完成後放入左後方的塑膠籃裡。首先，供料作…	847
作業場所人體尺寸圖譜在現場應用之探討	人工下料作業	人工下料之改善，這一站作業員需將兩部分原料倒進攪拌機，一部份原料需要控制重量，需使用磅秤，另一部份原料不需秤重，整包倒入攪拌機即可。待攪拌機攪拌均勻，將攪拌好的原料送進後面製程的機器及結束此…	528
職業肌肉骨骼傷害防治手冊	鋼捲包裝作業	現況說明，鋼捲包裝主要的工作內容是包裝已完成鍍鋅（鋅）的鋼捲，以利之後的運送作業。這個工作站的主要設備是一個放置於地板的鋼捲固定架，其高度為 10cm，作業所使用的器具包括拉緊器及固定器，兩者…	950
人因工程工作場所改善方法-礦泉水運送作業	礦泉水運送作業	現況說明，礦泉水運送作業的主要工作是將輸送帶上已完成填灌作業的礦泉水桶搬運至拖車上的木箱中，直至拖車上的所有木箱均已裝滿礦泉水桶，再將拖車推上貨車運送至各部門。在此運送作業的相關設施中，輸送帶…	708
人因工程現場不良工作姿勢改善績效評估研究-布捲吊掛作業	布捲吊掛作業	現況說明，布捲吊掛作業的主要工作內容是將布捲抬舉至吊掛架上。作業人員會先以一根圓柱形鐵條插入位於裁布桌下的布捲中心(該鐵條是為了將布捲固定於吊掛架上)，再由兩位人員合力彎腰將布捲抬起舉高並放入至吊掛架的叉鉤上。問題陳述，在這項作業…	934
工作現場人因工程危害預防效益研究-顏料桶上料作業	顏料桶上料作業	現況說明，上料指的是作業員利用長柄鏟子將顏料放置機器上，進行顏料裝桶的動作。首先作業員站在裝有顏料，直徑達 78 公分的顏料桶後方，利用長達 116 公分，重約 4 至 5 公斤的不銹鋼長鏟將顏料鏟起。再將長鏟橫越顏料桶上方，將顏料放置於機器的原料入。問題陳…	784

表 5.10、知識文件結構化摘要驗證之測試資料（其中 7 份）

文件名稱	作業名稱	文件內容	文件字數
工作現場人因工程危害預防效益研究-馬達底座加工作業	馬達底座加工作業	現況說明，馬達底座加工作業的主要工作是利用 CNC 車床對馬達底座進行加工作業。其作業程序為：作業人員首先徒手提取承載木箱內未加工的馬達底座，將之搬運至 CNC 車床內並同時固定馬達底座兩端，再由 CNC 車床對底座兩端進行車削加工。完成加工作業後，便送…	639
工作現場人因工程危害預防效益研究-模具加工作業	模具加工作業	現況說明，模具加工作業區的主要工作為進行模具零件的加工作業，包含組立、合模、去毛邊及研磨等作業，通常是一人以手工具獨力完成所有工作。其作業程序為作業人員首先將手推車送來的整組模具材料零件，逐一置放於工作台上，再以電動螺絲起子與扳…	788
作業場所人體尺寸圖譜在現場應用之探討-包裝重量檢驗作業	包裝重量檢驗作業	包裝重量檢驗站的作業中具有潛在危害，此作業之改善可分為兩個部分，以下分述。第一部分：料品由機器封裝，並經輸送帶送至自動秤重平台，符合規定重量之料品則運至下一站，不符合標準之料品則自動移下輸送…	586
工作現場人因工程危害預防效益研究-庫房搬運作業	庫房搬運作業	現況說明，一樓庫房區的主要功能為存放晶元原料與半成品物件，以為後續製造、加工之用。由於考慮晶元原料的特性，晶元原料與半成品均置放於統一規格的黑色儲□存箱之內，再將黑色儲存箱堆疊在可移動的推車之上，以提高儲存箱搬運的機動性。一般而言，每個推車…	930
裝潢木工職業危害分析評估改善與與空氣清淨裝置開發研究-裁切作業	裁切作業	本項作業已經進入室內裝潢最後階段，參與勞工主要為水電勞工，進行燈具等安裝時，必預使用電動或手動機具進行天花板鑿孔作業，勞工將直接暴露於鑿孔所散發之粉塵中，此時板材切割量相較於裝潢木工切割板材時為少，但因散發粉塵直接掉落至勞工之呼吸帶，所以佩…	261
職業肌肉骨骼傷害防治手冊-大理石補膠作業	大理石補膠作業	現況說明，補網作業是將大理石板上膠以補強強度及填補裂縫。作業場所為一寬 250 公分、高 90 公分的書，輸送帶是以橫向鋼條串起的滾輪所組成，輸送帶一處上方約 50 公分處有一捆與輸送帶同寬的紗網。首先，作業員將紗網於捲軸架上拉出，平鋪於長 250 公分和寬 180 公…	816
人因工程結合品管圈活動介入現場改善實例研究	座椅裝配作業	國瑞汽車觀音廠中，以生產休旅車與客貨兩用車為主，採混線方式生產，日產能合計為 60 部。本工作站名為後艙二班，有二位作業員，以裝配作業為主，包含前後保險桿、安全帶、油箱蓋和前後座椅組裝。作業員依據各車型之生產指示貼紙選擇適當之料件型號進行作業。…	598

### (B-1) 知識文件結構化摘要評估指標定義

針對結構化摘要呈現滿意度以及結構化摘要資訊準確性，本研究乃設計「摘要滿意度指標」以及「摘要文字呈現指標」作為驗證指標以檢視本系統結構化摘要推論效果。針對「摘要滿意度指標」乃驗證摘要呈現之成效，並針對知識搜尋者觀感以設計滿意度準則以輔佐知識搜尋者評級與考量，而「摘要文字呈現指標」乃驗證摘要之字數呈現，判斷摘要能否於範圍內有效表達文件內容。以下分別針對「摘要滿意度指標」以及「摘要文字呈現指標」進行說明。

「摘要滿意度指標」乃以使用者觀感進行績效驗證，本研究所設定使用者滿意度評級項目乃參考郭佩慧（2006），將文件重點（動機與目的明確）、文句結構（閱讀性）與字數（字數長度）作為閱讀摘要之評級標準。另外亦根據詹元智（2002）將評級記分採用 1、3 與 5 分，以保留彈性空間，使測試者可於確定問項內容後再予以訂定給分，並將不同程度結果以數字方式呈現表示等級。

針對本研究設計滿意度準則以輔佐知識搜尋者評級與考量，當中包含「字數長度符合閱讀」、「動機與目的明確」、「摘要閱讀性佳」等三種考量，並根據考量標準分別細述如下，並彙整如表 5.11 所示，其參數定義與計算方式如公式(5.3)所示：

- 字數長度符合閱讀：針對使用者閱讀摘要之觀感可分為合宜、略多或略少與過多或過少三種等級，使用者可針對摘要之字數進行觀感判定。當中，若使用者認為摘要字數呈現合宜則準則函數值為5、若字數略多或略少則準則函數值為3，最後字數過多或過少則準則函數值為1。
- 動機與目的明確：針對使用者閱讀摘要後之理解能力進行觀感分析，當中可分為非常明確、普通與不明確三種等級。若使用者閱讀摘要後認為該內容完整表達文件之動機與目的其為非常明確，準則函數值為5，若能大致了解者則準則函數值為3，最後不明確者則為1。
- 摘要閱讀性佳：針對使用者閱讀摘要之內容流暢性進行觀感分析，透過摘要內容流暢性分析即可評級摘要之閱讀性，當中可分為非常流暢、普通與不流暢三種等級。若使用者於閱讀中認為內容順暢，且文句對等、有順序性等，準則函數值為5，若內容流暢度尚可則準則函數值為3，最後不流暢者則為1。

表 5.11、評級項目與評級考量標準資料表

評級項目	得分	評級考量
字數長度符合閱讀	5	摘要字數呈現合宜
	3	摘要字數略多或略少
	1	摘要字數過多或過少
動機與目的明確	5	動機與目的非常明確
	3	動機與目的大致了解
	1	動機與目的不明確
摘要閱讀性佳	5	內容順暢、有順序性
	3	內容流暢度上可、有部分語意不清
	1	內容語意不清

- $S_i(\text{SAU})$  第*i*份知識文件之結構化摘要滿意度
- $I1_i(\text{SAU})$  第*i*份知識文件之結構化摘要之考量準則（字數長度），若知識搜尋者觀感為合宜則為5、略多或略少為3，若過多或過少則為1
- $I2_i(\text{SAU})$  第*i*份知識文件之結構化摘要之考量準則（動機與目的明確），若知識搜尋者針對摘要內容非常明確則為5、大致了解則為3，若不明確則為1
- $I3_i(\text{SAU})$  第*i*份知識文件之結構化摘要之考量準則（摘要閱讀性佳），若知識搜尋者認為內容順暢則為5、尚可則為3，若不順暢則為1

$$S_i(\text{SA}) = \frac{I1_i(\text{SA}) + I2_i(\text{SA}) + I3_i(\text{SA})}{3} \quad (5.3)$$

「摘要文字呈現指標」乃進行摘要呈現與表達性驗證，針對摘要之字數呈現可分為合宜、略多或略少與過多或過少三種等級，本研究乃以Kuo等人（2002）為基礎，將摘要文字最佳效果設定為原文百分之十至百分之十五，並作為簡略摘要之標準。若字數為標準範圍正負10字則為合宜（準則函數值為5）、若於標準之正負30字則為略多或略少（準則函數值為3），若超過標準之正負30字則為過多或過少（準則函數值為1）。另外，針對詳述結構化摘要部分，由於詳述結構化乃為簡化文件與細述文件之內容，其字數範圍設定為原文百分之五十，若字數為標準範圍正負10字則為合宜（準則函數值為5）、若於標準之正負30字則為略多或略少（準則函數值為3），若超過標準之正負30字則為過多或過少（準則函數值為1）。其各指標定義與計算方式如公式(5.4)至公式(5.5)所示：

- $DEZ_i(\text{SAS})$  第*i*份知識文件之簡略結構化摘要文字呈現指標
- $CnEZt_i(\text{D})$  第*i*份知識文件之簡略結構化摘要總字數
- $MaxEZS_i(\text{D})$  第*i*份知識文件之簡略結構化摘要之標準字數最大值
- $MinEZS_i(\text{D})$  第*i*份知識文件之簡略結構化摘要之標準字數最小值
- $CntEZA_i(\text{D})$  第*i*份知識文件之簡略結構化摘要之系統呈現字數
- $IIEZ_i(\text{SAS})$  第*i*份知識文件之簡略結構化摘要之文字呈現效能考量準則，若摘要字數於為標準值正負10字為合宜，其考量值為5、正負30字內為3，超過則為1
- $DT_i(\text{SAS})$  第*i*份知識文件之詳述結構化摘要文字呈現指標
- $CntDT_i(\text{D})$  第*i*份知識文件之詳述結構化摘要總字數
- $CntDTS_i(\text{D})$  第*i*份知識文件之詳述結構化摘要之標準字數

CntDTA<sub>i</sub>(D) 第i份知識文件之詳述結構化摘要之系統呈現字數

I1DT<sub>i</sub>(SAS) 第i份知識文件之簡略結構化摘要之文字呈現效能考量準則，若摘要字數於為標準值正負10字為合宜，其考量值為5、正負30字內為3，超過則為1

$$\begin{aligned} \text{MinEzS}_i(D) &= \text{CntEz}_i(D) \cdot 10\% \\ \text{MaxEzS}_i(D) &= \text{CntEz}_i(D) \cdot 15\% \\ \text{I1Ez}_i(\text{SAS}) &= \begin{cases} 5, & \text{if } \text{MinEzS}_i(D) - 10 < \text{CntEzA}_i(D) < \text{MaxEzS}_i(D) + 10 \\ 3, & \text{if } \text{MinEzS}_i(D) - 30 < \text{CntEzA}_i(D) < \text{MaxEzS}_i(D) + 30 \\ 1, & \text{if } \text{MinEzS}_i(D) - 30 > \text{CntEzA}_i(D) \\ & \text{OR } \text{MaxEzS}_i(D) + 30 < \text{CntEzA}_i(D) \end{cases} \end{aligned} \quad (5.4)$$

$$\begin{aligned} \text{CntDTS}_i(D) &= \text{CntDT}_i(D) \cdot 50\% \\ \text{I1DT}_i(\text{SAS}) &= \begin{cases} 5, & \text{if } \text{CntDTS}_i(D) - 10 < \text{CntDTA}_i(D) < \text{CntDTS}_i(D) + 10 \\ 3, & \text{if } \text{CntDTS}_i(D) - 30 < \text{CntDTA}_i(D) < \text{CntDTS}_i(D) + 30 \\ 1, & \text{if } \text{CntDTS}_i(D) - 30 > \text{CntDTA}_i(D) \\ & \text{OR } \text{CntDTS}_i(D) + 30 < \text{CntDTA}_i(D) \end{cases} \end{aligned} \quad (5.5)$$

## (B-2) 知識文件結構化摘要驗證結果與分析

於第一階段系統驗證中，針對知識文件結構化摘要部分則以 500 份知識文件中隨機挑選 100 份作為結構化摘要之訓練資料，並逐一匯入系統中，以作為第一階段驗證之基礎訓練資料。以下即針對各項指標說明系統驗證過程，並分析系統驗證之結果。

### (B-I) 知識文件結構化摘要第一階段驗證結果分析 (15 份知識文件)

針對知識文件結構化摘要推論部分，在 100 份文件作為訓練資料之基礎下，將其知識文件資料包含作業名稱、文件內容與上傳時間等匯入系統，再隨機抽取 15 份知識文件作為測試資料（如表 5.12 所示，以貝果運送作業、氣泡料攪拌作業、馬達底座加工作業、礦泉水運送作業與切塊作業為範例），於第一階段測試結果乃彙整於表 5.13 與表 5.14。以表 5.13 為例，於貝果運送作業其表達項目「作業行為」之描述為「在推送時作業員需從輸送帶的前端一路…，」與「輸送帶連接烘烤機和貝果再加工區，」、表達項目「傷害成因」之描述為「亦即頭頸前傾、背部直立、雙手位於肩下方、腿部為坐姿、重量在 5 公斤以下、而改善後的工作姿勢並沒有改變，」等，而表達項目「改善方法」之描述為「建議使用如圖 113 所示的檔板，」。此外，簡略摘要使用者滿意度與文件摘要呈現效果之分佈與趨勢彙整如表 5.14、表 5.15、圖 5.35、圖 5.36 與圖 5.37 所示。

表 5.12、知識文件結構化摘要驗證之測試資料（其中 5 份）

文件名稱	作業名稱	文件內容	文件字數
人因工程現場不良工作姿勢改善績效評估研究-貝果運送作業	貝果運送作業	現況說明，此作業主要是將烘烤完成的貝果經由輸送帶運送到再加工區。輸送帶連接烘烤機和貝果再加工區，在輸送過程中，輸送帶必須先約升約 2~3 公尺之後再下降，以增加麵包冷卻的時間。然而，此輸送帶並非專為貝果生產線特別設計，在爬升運送時，貝果…	905
工作現場人因工程危害預防效益研究-氣泡料攪拌作業	氣泡料攪拌作業	氣泡料攪拌作業:現況說明泡料室攪拌作業，通常是一人獨立完成所有工作。泡料室面積約七坪左右，室內主要擺設為攪拌機 2 台（圖 9-1）及放置泡料粉的貨架（圖 9-2），攪拌機圓桶高 94 公分，直徑約 89 公分，空…	913
工作現場人因工程危害預防效益研究-馬達底座加工作業	馬達底座加工作業	現況說明，馬達底座加工作業的主要工作是利用 CNC 車床對馬達底座進行加工作業。其作業程序為：作業人員首先徒手提取承載木箱內未加工的馬達底座，將之搬運至 CNC 車床內並同時固定馬達底座兩端，再由 CNC 車床對底座兩端進行車削加工。完成加工作業後，便送…	639
人因工程工作場所改善方法-礦泉水運送作業	礦泉水運送作業	現況說明，礦泉水運送作業的主要工作是將輸送帶上已完成填灌作業的礦泉水桶搬運至拖車上的木箱中，直至拖車上的所有木箱均已裝滿礦泉水桶，再將拖車推上貨車運送至各部門。在此運送作業的相關設施中，輸送帶…	708
人因工程工作場所改善方法-切塊作業	切塊作業	切塊作業分切雞塊作業是作業員將經由輸送帶運送到面前的盤裝雞肉拿起，取出雞肉放在棧板上，接著再利用大刀將半雞分為三大塊，再逐一剝成小塊狀，整齊排列於盤內，完成後放入左後方的塑膠籃裡。首先，供料作…	847

表 5.13、知識文件結構化摘要第一階段實際結果呈現（以 5 份為例）(1)

文件名稱	作業名稱	作業行為	傷害成因	改善方法
人因工程現場不良工作姿勢改善績效評估研究-貝果運送作業	貝果運送作業	在推送時作業員需從輸送帶的前端一路走到輸送帶高起較緩和處，	生物力學計算顯示改善前與改善後的頸、腰、肩、肘、腕、髖、膝、踝與腳底等九個關節的肌肉骨骼負荷程度如下表。	建議使用如圖 113 所示的檔板，
		輸送帶連接烘烤機和貝果再加工區，	亦即頭頸前傾、背部直立、雙手位於肩下方、腿部為坐姿、重量在 5 公斤以下、而改善後的工作姿勢並沒有改變，	建議使用如圖 113 所示的檔板，
人因工程工作場所改善方法-氣泡料攪拌作業	氣泡料攪拌作業	泡料粉都放在棧板上，	改善方案改善姿勢不良與過度施力的問題，	肌肉主動收縮而可以產生最大力量稱之。
		攪拌機圓桶高 94 公分	改善方案改善姿勢不良與過度施力的問題，	同時當上身向前屈曲時
人因工程工作場所改善方法-切塊作業	切塊作業	能夠以支架支撐代替人工，	使腿部與軀幹分離，	以目前的作業高度，
		置於作業台上，	供站立於作業台兩端的作業員進行肢解作業。	可以逕行檢討改變作業姿勢。



表 5.13、知識文件結構化摘要第一階段實際結果呈現（以 5 份為例）(2)

文件名稱	作業名稱	作業行為	傷害成因	改善方法
工作現場人因工程危害預防效益研究-馬達底座加工作業	馬達底座加工作業	而 CNC 車床的閘門入口處高度約為 100 公分。	改善方案要改善作業人員因徒手抬舉、搬運 25 公斤馬達底座所造成的過度施力問題，	盡量讓作業範圍保持在人員的上身範圍，
		馬達底座重量約為 25 公斤，	將造成腰、肩、肘與腕等部位的肌肉骨骼傷害。	以減少肉骨骼的疲勞與傷害發生。
人因工程工作場所改善方法-礦泉水運送作業	礦泉水運送作業	此裝置乃為象鼻子真空省力裝置	只要以此裝置的吸物盤口含納礦泉水桶的頸部出口處，	作業人員可以不必直接搬運礦泉水桶，
		而輸送帶與拖車上木箱的各別距離便在 1.	也會對腰部造成相當大的剪力。	有效地舉高或放下礦泉水桶並改變其水平位置，

表 5.14、第一階段知識文件之簡略摘要滿意度指標績效彙整

受測者	15 份測試資料														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00	1.00	1.67	1.00	1.67	1.00	1.67	1.67	1.00	1.67	1.67	1.67	1.67	1.67	1.00
2	1.00	1.67	1.67	1.00	1.00	1.67	1.67	1.67	1.67	1.00	1.67	1.67	1.67	1.67	3.00
3	1.67	1.67	3.67	2.33	1.67	1.00	1.67	1.00	1.67	1.00	1.67	1.67	1.00	1.00	1.00
4	2.33	1.67	1.67	1.67	2.33	2.33	1.67	1.00	1.67	1.67	1.00	1.67	1.67	2.33	1.67
5	1.67	1.67	1.00	1.67	2.33	1.67	1.00	1.00	1.00	1.67	1.00	1.00	1.67	1.00	1.67
6	1.67	1.67	1.67	1.00	1.67	1.67	1.00	1.00	1.67	1.00	1.67	1.00	1.67	1.67	1.00
7	3.67	1.00	1.67	2.33	2.33	1.67	1.00	1.67	1.00	1.00	1.00	1.00	1.67	1.67	1.00
8	1.67	1.00	3.00	1.00	1.67	1.67	1.00	1.00	1.00	1.67	1.00	1.67	1.67	1.00	1.00
9	1.00	1.67	1.67	1.67	2.33	1.67	1.00	1.00	1.67	1.67	1.67	1.67	1.67	2.33	1.67
10	2.33	1.00	1.00	1.67	1.00	1.67	1.00	1.67	1.67	1.00	1.00	1.67	1.00	1.67	3.00
11	1.00	1.00	1.67	1.00	1.00	1.00	1.00	1.67	1.00	1.00	1.67	1.67	1.00	1.00	1.00
12	1.00	1.67	1.67	1.67	1.00	1.00	1.00	1.67	1.67	1.67	1.00	1.00	1.00	1.67	1.00
13	1.67	1.67	2.33	2.33	1.67	1.00	1.00	2.33	1.67	1.67	1.67	1.67	1.00	1.00	1.00
14	1.67	1.67	1.67	1.67	1.00	1.67	1.67	1.67	1.67	1.00	1.67	1.67	1.67	1.67	1.67
15	1.67	1.00	1.67	1.00	1.67	1.67	1.67	1.00	2.33	1.67	1.00	1.00	1.00	1.00	1.67
16	3.67	1.67	1.67	1.00	1.67	1.67	1.00	3.00	1.67	1.67	1.67	1.67	2.33	1.67	2.33
17	1.00	1.67	1.67	1.67	1.67	1.67	2.33	2.33	1.67	1.67	1.00	1.67	1.67	1.00	1.00
18	1.67	1.00	1.00	1.00	1.67	2.33	1.67	1.67	1.00	1.67	1.67	3.67	1.67	1.00	1.00
19	2.33	1.00	1.00	1.67	1.67	1.00	1.67	1.67	1.67	1.67	1.67	1.67	1.00	1.67	1.67
20	1.67	1.00	3.00	1.67	2.33	1.67	1.00	1.67	1.67	1.00	1.00	1.67	1.00	1.67	1.67
平均	<b>1.77</b>	<b>1.37</b>	<b>1.80</b>	<b>1.53</b>	<b>1.67</b>	<b>1.54</b>	<b>1.33</b>	<b>1.57</b>	<b>1.50</b>	<b>1.40</b>	<b>1.37</b>	<b>1.60</b>	<b>1.44</b>	<b>1.47</b>	<b>1.50</b>

表 5.15、第一階段知識文件之摘要呈現效果指標績效彙整

文件作業名稱	原文件 字數	簡略摘要呈現效果指標績效			詳述摘要呈現效果指標績效			
		理想摘要 字數範圍	結構化 摘要字數	文字呈 現效能	理想摘要 字數	結構化 摘要字數	文字呈 現效能	
貝果運送作業	905	91	136	206	1	453	513	1
氣泡料攪拌作業	913	91	137	184	1	457	517	1
切塊作業	847	85	127	125	5	424	517	1
馬達底座加工作業	639	64	96	191	1	320	470	1
模具加工作業	788	79	118	157	1	394	576	1
包裝重量檢驗作業	586	59	88	132	1	293	354	1
顏料桶上料作業	784	78	118	210	1	392	468	1
裁切作業	261	26	39	143	1	131	149	3
人工下料作業	528	53	79	126	1	264	374	1
鋼捲包裝作業	950	9	143	194	1	475	632	1
礦泉水運送作業	708	71	106	197	1	354	474	1
布捲吊掛作業	934	93	140	178	1	467	589	1
庫房搬運作業	930	93	140	126	5	465	492	1
座椅裝配作業	598	60	90	108	3	299	424	1
大理石補膠作業	816	82	122	196	1	408	431	3
平均值 (簡略摘要)					<b>1.67</b>	平均值 (詳述摘要)		<b>1.27</b>

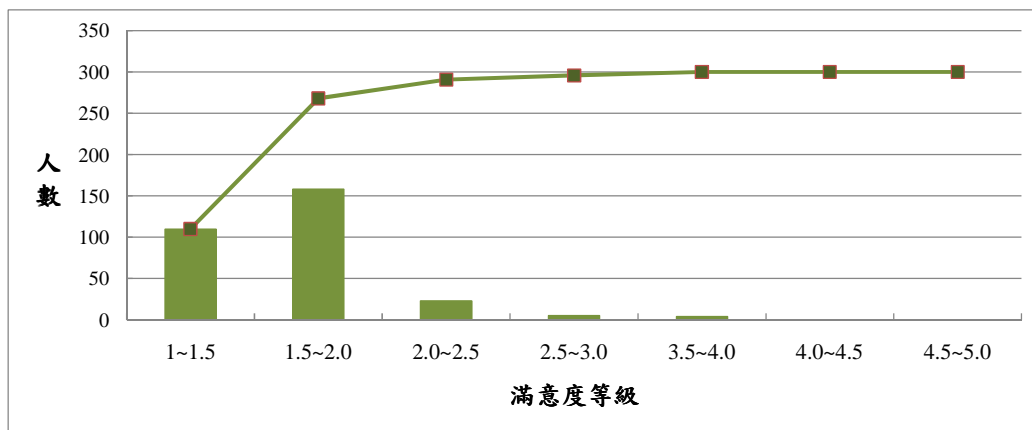


圖 5.35、第一階段知識文件之摘要滿意度之分佈趨勢

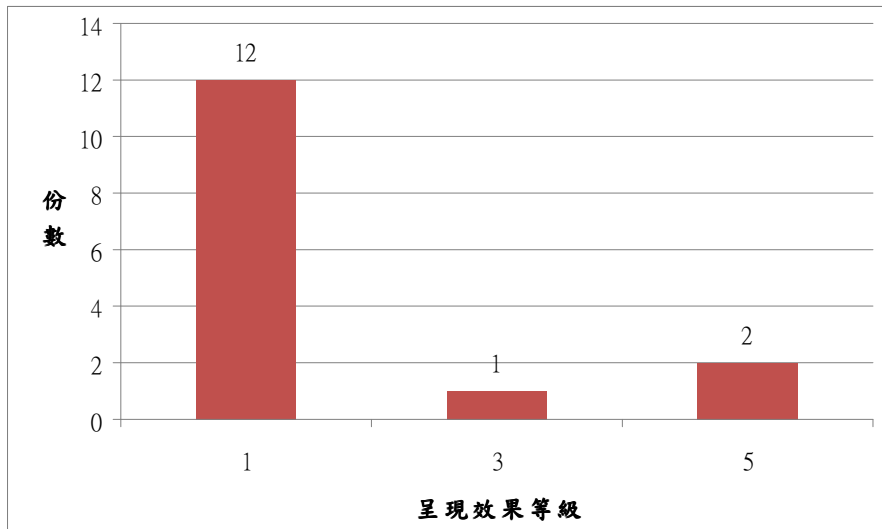


圖 5.36、第一階段知識文件之簡略摘要呈現效果之分佈趨勢

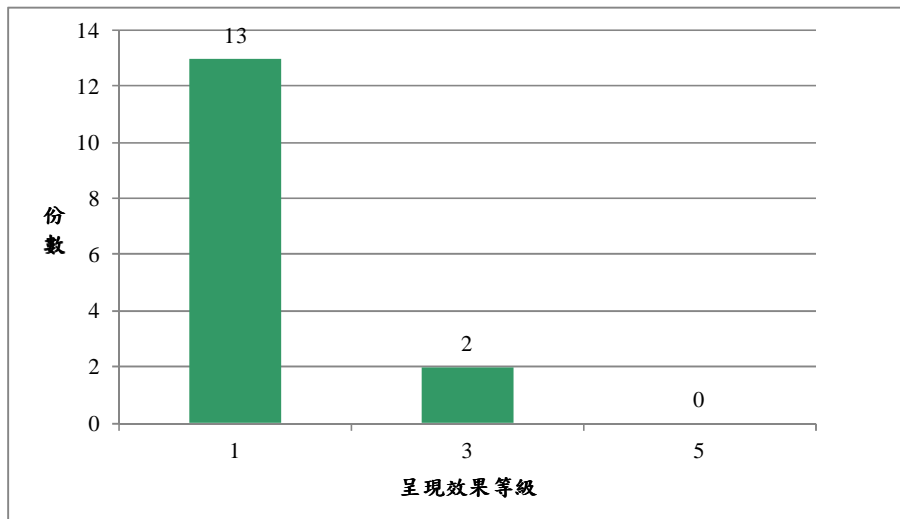


圖 5.37、第一階段知識文件之詳述摘要呈現效果之分佈趨勢

藉由表 5.14、表 5.15、圖 5.35、圖 5.36 與圖 5.37 統計結果以得知，於第一階段之系統驗證結果中其使用者滿意度約落於 1.00 與 1.67，其平均滿意度為 1.52。另外簡略結構化摘要文字呈現部分，有兩筆簡略結構化摘要符合標準字數正負 10 內，且有一筆文件為標準字數正負 30 內，共計三筆符合，其文字呈現效果平均為 1.67，而詳述結構化摘要文字呈現部分，僅有兩筆符合標準字數正負 30 內，其文字呈現效果平均為 1.27。整體而言，根據滿意度指標（滿意度 5、3、1）可得知無論是使用者滿意度，亦或是文字呈現效果皆低於滿意度之平均值。

## (B-II) 知識文件結構化摘要第二階段驗證結果分析 (15 份知識文件)

第二階段系統驗證之作法乃以第一階段驗證中，針對知識文件結構化摘要部分將第二階段分為分 5 個週期 (每週期皆匯入 80 筆不重複之知識文件，共計 400 筆組合)，並以第一階段驗證時所選取之 15 份知識文件重新進行系統績效測試，以瞭解系統於不同訓練資料數量基礎下進行知識文件之問答解析及結構化摘要之績效變化趨勢，進而分析本系統之學習成長能力。以下乃針對問答解析及結構化摘要之各項指標說明系統第二階段各週期驗證過程，並分析系統各週期之驗證結果。

針對知識文件結構化摘要之第二階段驗證，其最後測試(第五週期)之結果如表 5.16 所示 (以貝果運送作業、氣泡料攪拌作業、馬達底座加工作業、礦泉水運送作業與切塊作業為範例)，以表 5.16 為例，於貝果運送作業其表達項目「作業行為」之描述為「此作業主要是將烘烤完成的貝果經由輸送帶運送到再加工區。」與「作業員來回走動的頻率相當高，」、表達項目「傷害成因」之描述為「可避免作業員來回走動堆放貝果的情形，」等，而表達項目「改善方法」之描述為「作業員可站在輸送帶起始處將擋板插入輸送帶，」。而各驗證週期知識文件結構化摘要之「摘要滿意度」與「文字呈現效果」績效分佈趨勢如圖 5.37 與圖 5.38 所示。

表 5.16、第二階段第六週期實際結果呈現(1)

文件名稱	作業名稱	作業行為	傷害成因	改善方法
人因工程現場不良工作姿勢改善績效評估研究-貝果運送作業	貝果運送作業	此作業主要是將烘烤完成的貝果經由輸送帶運送到再加工區。	可避免作業員來回走動堆放貝果的情形，	作業員可站在輸送帶起始處將擋板插入輸送帶，
		此輸送帶並非專為貝果生產線特別設計，	作業員來回走動的頻率相當高，	建議使用如圖 113 所示的擋板，
人因工程工作場所改善方法-氣泡料攪拌作業	氣泡料攪拌作業	泡料室攪拌作業的過程是泡料粉經由板車推送至攪拌機的圓桶旁邊，	同時當上身向前屈曲時，	可以防止作業員徒手彎腰搬運 25 公斤的泡料粉，
		可以防止作業員徒手彎腰搬運 25 公斤的泡料粉，	作業員軀幹向前屈曲超過 60 度，	我們的建議是採用真空搬運機，
人因工程工作場所改善方法-切塊作業	切塊作業	供站立於作業台兩端的作業員進行肢解作業。	問題陳述這個作業的主要問題在於震動。	我們建議採用以外力取代中的「以支架支撐」的方式，
		切塊作業分切雜塊作業是作業員將經由輸送帶運送到面前的盤裝	問題陳述這個作業的主要問題在於震動。	則不必考慮「改變身體施力部位」，

表 5.16、第二階段第六週期實際結果呈現(2)

工作現場人因工程危害預防效益研究-馬達底座加工作業	馬達底座加工作業	而 CNC 車床的閘門入口處高度約為 100 公分。	改善方案要改善作業人員因徒手抬舉、搬運 25 公斤馬達底座所造成的過度施力問題，	盡量讓作業範圍保持在人員的上身範圍，
		馬達底座重量約為 25 公斤，	將造成腰、肩、肘與腕等部位的肌肉骨骼傷害。	以減少肉骨骼的疲勞與傷害發生。
人因工程工作場所改善方法-礦泉水運送作業	礦泉水運送作業	礦泉水運送作業的主要工作是將輸送帶上已完成填灌作業的礦泉水桶搬運至拖車上的木箱中，	要改善因抬舉、搬運重達 20 公斤的礦泉水桶所造成的施力過大問題，	可以使用一個輔助搬運裝置來協助作業人員搬運。
		每個作業人員一天必須至少搬運 200 桶，	此搬運重量已相當接近美國勞工安全衛生研究所的抬舉指引搬運重量上限 23 公斤	完全避免因施力過大所造成肌肉骨骼的疲累問題，

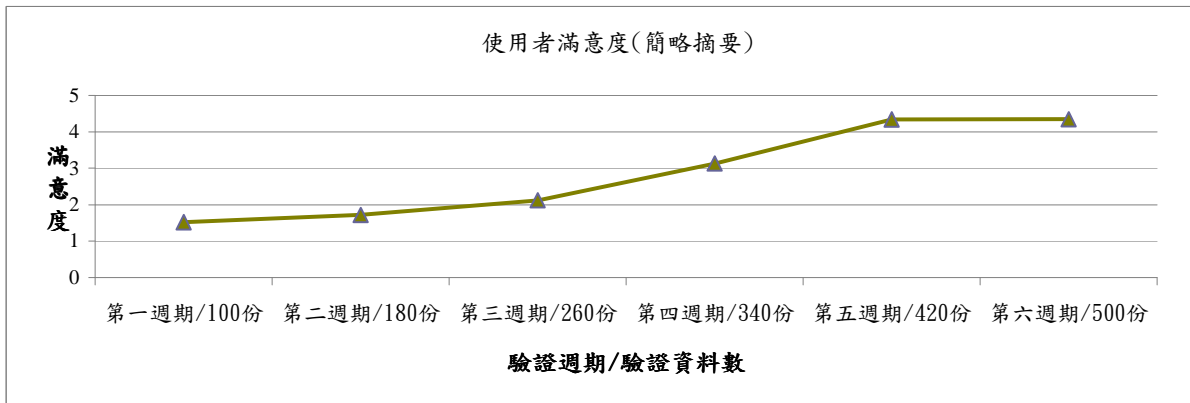


圖 5.38、各驗證週期之結構化摘要滿意度之分佈趨勢

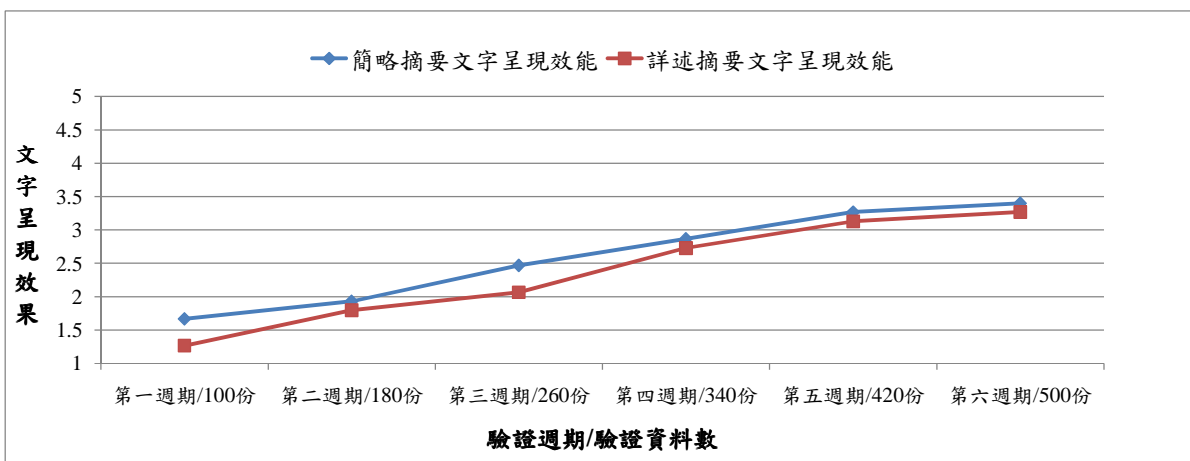


圖 5.39、各驗證週期之結構化摘要文字呈現效果之分佈趨勢

於圖 5.38 與圖 5.39 分佈趨勢以得知，於第二階段以每週增加 80 份知識文件為單位，平均每週使用者滿意度由 1.72 成長至 4.35，而文字呈現效果簡略部分由 1.93 成長至 3.40，符合標準字數之文件從 5 筆成長至 14 筆；詳述部分由 1.80 成長至 3.21，符合標準字數之文件從 6 筆成長至 14 筆故可得知本系統結構化摘要推論良好，且具學習能力。

### (B-III) 知識文件結構化摘要推論整體驗證結果分析

在知識文件結構化摘要推論第一階段驗證中，當訓練知識文件數量為 100 筆時，其使用者平均滿意度為 1.52。另外簡略結構化摘要文字呈現部分，其文字呈現效果平均為 1.67，而詳述結構化摘要文字呈現部分效果平均為 1.27。整體而言，由數據可得知於第一階段知識文件結構化摘要推論之績效欠佳，然而，於第二階段驗證中，系統乃隨著不同週期及訓練組合句之增加下，各項數據呈現顯著之成長，平均每週使用者滿意度由 1.72 成長至 4.35，而文字呈現效果簡略部分由 1.93 成長至 3.40。綜合上述數據，可證實本研究所建置之「知識文件結構化摘要推論模組」乃實具可行性。

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 5.17 所示。由表 5.17 整理結果可得知，各項驗證指標之「收斂前每週平均成長率」及「整體每週平均成長率」皆為正數成長，且各項驗證指標皆於第五週內呈現收斂狀態，因此，以本研究所選定之驗證資料（勞工安全衛生知識網知識文件）為例，當系統使用約 420 份知識文件時，可使系統各項推論指標績效提升至滿意度 3 以上之水準。

表 5.17、知識文件結構化摘要綜合兩階段之驗證績效彙整

驗證指標	整體平均值	收斂週期	收斂前每週平均成長率	整體每週平均成長率
簡略摘要使用者滿意度	2.86	第五週期	0.71%	0.57%
簡略摘要文字呈現效果	2.60	第五週期	0.4%	0.35%
詳述摘要文字呈現效果	2.38	第五週期	0.47%	0.4%

整體而言，由第一階段與第二階段之驗證分析中，可得知當使用者進行「知識文件結構化摘要推論」時，訓練資料與週期乃影響系統績效之主要原因，系統之績效將隨著不同週期及訓練量之增加，呈現持續成長之狀況，並於最終達到穩定且良好之績效水平。

### 5.3 小結

綜合上述，本研究乃針對結合語意解析與文件摘要技術之知識文件檢索系統，以進行知識文件問答解析與知識文件結構化摘要推論，並根據回覆詞召回率以及回覆詞正確率進行問答解析之驗證，並以簡略摘要使用者滿意度、簡略摘要文字呈現效果與詳述摘要文字呈現效果進行結構化摘要推論之驗證。當中各驗證皆以兩階段方式進行，且於第二階段分成週期匯入資料，以瞭解知識文件問答解析與結構化摘要推論之績效之影響，其針對兩驗證之分析結果彙整如下：

- 針對知識文件問答解析部分，由驗證結果以得知於訓練詢問句與回覆句組合筆數為200筆時，其「回覆詞正確率」與「回覆詞召回率」分別為31.63%及43.33%，待匯入約700筆組合時即可達到收斂，其「回覆詞正確率」與「回覆詞召回率」可提升至86.67%及84.17%。有鑑於上述之數據，即表示本系統之知識文件問答解析功能具可行性。
- 針對知識文件結構化摘要推論部分，由驗證結果以得知於訓練詢問句與回覆句組合筆數為100筆時，其使用者平均滿意度為1.52。另外簡略結構化摘要文字呈現部分，其文字呈現效果平均為1.67，而詳述結構化摘要文字呈現部分效果平均為1.27。待匯入約420筆知識文件時即可達到收斂，平均每週期使用者滿意度由1.72成長至4.35，而文字呈現效果簡略部分由1.93成長至3.40。綜合上述數據，即表示本系統之知識文件結構化摘要推論具可行性。

## 第六章、結論與未來發展

雖有特定領域網站作為資訊搜尋窗口以縮小搜尋範圍，但由於網路資源龐大，從中仍需進行文件分類、關鍵字查詢等方式以準確獲取所需文件。但若於特定知識領域可能因知識搜尋者無相關領域背景知識，導致以關鍵字搜尋方式無法確切取得回饋，或是知識搜尋者無法定義關鍵字詞反而以揣測方式進行搜尋，因而增加知識搜尋時間。除此之外，搜尋者必須透過文件標題先行篩選再以閱讀摘要方式判定文件取決與否，而該網頁之文件摘要呈現方式乃以作者自行撰寫，其自由形式呈現方式容易影響搜尋者判斷文件取決思考。

為解決上述問題，本研究乃先解析勞工安全衛生知識網之技術叢書與研究報告，以了解知識文件之表達方式、呈現內容以及相關性語彙，再以上述資料為基發展一套「結合語意解析與文件摘要技術之知識文件檢索」方法論，以將知識文件進行表達項目解析並取得知識文件關鍵字詞，同時建立專業語彙與一般語彙之語意關聯，再根據表達項目將知識文件轉成制式結構化摘要。最後，根據上述方法論乃建置一套「結合語意解析與文件摘要技術之知識文件檢索系統」，以協助知識搜尋者尋求問題解，並能透過結構化摘要呈現方式快速且有效取得資訊與知識。本研究期望藉由「結合語意解析與文件摘要技術之知識文件檢索系統」協助知識搜尋者進行知識文件，以減少搜尋文件之時間，進而加提高勞工安全衛生知識網知識分享之成效。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四項任務，以下乃分別說明各項任務之成果。

#### 1. 解析知識文件之表達項目

於此任務本研究乃先閱讀勞工安全衛生知識網之研究報告與技術叢書，並釐清知識文件之表達項目與內容，以歸類出八個表達項目（作業名稱、傷害成因、作業行為、作業人員、作業時間與改善方法等）。



## 2. 發展「結合語意解析與文件摘要技術之知識文件檢索」方法論

經由知識文件表達項目解析任務，即可取得知識文件之表達項目等組成知識文件之元素。本研究即以解析結果發展一套「結合語意解析與文件摘要技術之知識文件檢索」方法論。當中方法論可包含「知識文件表達項目解析模組」、「知識文件問答解析模組」與「知識文件結構化摘要推論模組」。透過「知識文件表達項目解析模組」以判斷並擷取知識文件之觀念性語句與代表性關鍵字詞。再由「知識文件問答解析模組」即進行知識文件專業語彙與一般詞彙之語意關聯性，藉由語意關聯性進而取得相關知識文件，最後透過「知識文件結構化摘要推論模組」將知識文件之轉為制式化摘要呈現，以有效表達知識文件之動機、目的或細部描述。

## 3. 建置「結合語意解析與文件摘要技術之知識文件檢索系統」

本研究乃以「結合語意解析與文件摘要技術之知識文件檢索」方法論為依據，運用 JSP (Java Server Pages) 語法進行開發，並利用 SQL Server 2005 來存取資料庫，以建置一套「結合語意解析與文件摘要技術之知識文件檢索系統」，並使系統之使用者能進行上傳與查詢知識文件、以詢問句查詢並取得相關文件，或以結構化摘要方式檢視文件以及設定系統中相關預設值等維護作業。

## 4. 方法論與系統績效驗證

為驗證本研究所提出之方法論與系統之正確性與有效性，本研究乃以「勞工安全衛生知識網」之研究報告與技術叢書中案例作業為例，分別針對「知識文件問答解析」與「知識文件結構化摘要」等兩議題進行驗證，並以「回覆詞召回率」與「回覆詞正確率」檢視系統問答解析績效與正確性，亦由「使用者滿意度」與「文字呈現效果」指標驗證與評估本研究之結構化摘要推論之實用性。

綜合上述內容，本研究所提出之「結合語意解析與文件摘要技術之知識文件檢索」方法論與「結合語意解析與文件摘要技術之知識文件檢索」系統可有效將知識文件解析並取得代表關鍵字詞，以進行語意搜尋，並能將知識文件轉為制式摘要呈現，以提高文件篩選之效能。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**

本研究乃以特定知識領域之知識文件為基礎，發展一套「結合語意解析與文件摘要技術之知識文件檢索」模式，以解析知識文件表達項目並進行專業語彙語意關聯以及結構化摘要呈現。此方法論之相關重點成效乃歸納如下：

- 本研究乃解析並取得知識文件之觀念性語句以及具有代表性之專業語彙，並將一般詞彙與專業語彙進行語意關聯，知識搜尋者即可以一般詞彙進行搜尋，並取得相關知識文件。
- 本研究提出之結構化摘要即可將自由形式知識文件轉為制式化之文件摘要，知識搜尋者能以制式呈現方式更容易審視知識文件，同時亦可避免個人閱讀偏好影響文件之篩選。

## **技術開發層面**

本研究乃以「結合語意解析與文件摘要技術之知識文件檢索」方法論為依據，並以JSP（Java Server Pages）以及SQL Server 2005等系統開發工具建置「結合語意解析與文件摘要技術之知識文件檢索」系統，其系統具體成效乃歸納如下：

- 本系統所開發之「結合語意解析與文件摘要技術之知識文件檢索」乃建置於網際網路中，並可執行「知識文件表達項目解析」、「知識文件問答解析」、「知識文件結構化摘要推論」以取得知識文件觀念性語句、關鍵字詞與相關語意詞，以及知識文件簡略結構化摘要與詳述結構化摘要呈現。
- 由驗證結果得知，本系統之各項驗證指標（即「知識文件問答解析」之回覆詞召回率、回覆詞正確率與「知識文件結構化摘要推論」使用者滿意度與文件字數呈現）皆具學習能力；當使用至一定訓練資料數量系統即可有效解析知識文件之觀念性語句、專業語彙與一般詞彙之語意關聯與建立結構化摘要。

## **實務應用層面**

本研究所建置之「結合語意解析與文件摘要技術之知識文件檢索」系統可有效分析知識文件，故此系統可應用於線上特定領域網站以解決知識文件搜尋等任務。其具體貢獻與成效乃歸納如下：

- 本系統可應用於勞工安全衛生知識網等相關領域網站，以有效表達知識文件之描述、並協助知識搜尋者尋求問題解答等搜尋任務。

- 本系統可應用於勞工安全衛生知識網等相關領域網站，以結構化摘要方式呈現知識文件，以協助知識搜尋者省力且快速方式獲得資訊，進而提高知識領域搜尋效能。

## 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「結合語意解析與文件摘要技術之知識文件檢索」方法論與「結合語意解析與文件摘要技術之知識文件檢索」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. 發展一套知識文件多文件摘要檢索模式

未來期望藉由多文件摘要技術發展出一套知識文件檢索技術，以協助知識搜尋者進多文件資料整合，以歸納相關性文件之重點。

### 2. 發展一套多知識文件問答系統

未來可針對問項類型（如詢問改善方式、詢問勞工安全法規等特定條件），以知識文件之主題、表達項目等條件方式進行問答系統，以協助知識搜尋者再縮小搜尋範圍，更能精確且有效取得知識與資訊。

### 3. 發展一套權重值自動推論法則

本研究所提出之權重值乃由系統管理者自行判斷設定而來，可能從結構化摘要推論之精神以得知主要以向心性權重為主而結構性權重為輔，由於人為判定權重時可能造成不同呈現結果，因此於未來可探討系統參數權重值對於系統推論績效之影響，並發展一套權重值自動推論法則，即藉由系統自動推論方式以避免人為判定，並有效決定系統參數之權重達到最佳效果。

## 參考文獻

1. 郭佩慧，2006，「中文閱讀摘要學習系統的發展與應用」，碩士論文（指導教授：洪碧霞／黃國禎），國立台南大學測驗統計研究所。
2. 詹元智，2002，「國小數學科實作評量之效度探討」，碩士論文（指導教授：張麗麗），屏東師範學院教育心理與輔導研究所。
3. Benedí, J. M. and Sánchez, J. A., 2005, “Estimation of stochastic context-free grammars and their use as language models” *Computer Speech & Language*, Vol. 19, No. 3, pp. 249-274.
4. Bollegala, D., Okazaki, N. and Ishizuka, M., 2010, “A bottom-up approach to sentence ordering for multi-document summarization,” *Information Processing and Management*, Vol. 46, No. 1, pp. 89-109.
5. Bouras, C., Pouloupoulos, V. and Tsogkas, V., 2008, “PeRSSonal’s core functionality evaluation: enhancing text labeling through personalized summaries,” *Data&Knowledge Engineering*, Vol. 64, No. 1, pp. 330-345.
6. Cao, Y. G., Liu, F., Simpson, P., Antieau, L., Bennett, A. Cimino, J. J., Ely, J. and Yu, H., 2011, “AskHERMES: An online question answering system for complex clinical questions,” *Journal of Biomedical Informatics*, Vol. 44, No. 2, pp. 277-288.
7. 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.
8. Dalmas, D. and Webber, B., 2007, “Answer comparison in automated question answering,” *Journal of Applied Logic*, Vol. 5, No. 1, pp. 104-210.
9. Dorr, B. J. and Gaasterland, T., 2007, “Exploiting aspectual features and connecting words for summarization-inspired temporal-relation extraction,” *Information Processing and Management*, Vol. 43, No. 6, pp. 1681-1704.
10. Dunlavy, D. M., O’Leary, D. P., Conroy, J. M. and Schlesinger, J. D., 2007, “QCS: A system for querying, clustering and summarizing documents,” *Information Processing and Management*, Vol. 43, No. 6, pp. 1588-1605.
11. Elhadad, N., Kan, M. Y., Klavans, J. L. and McKeown, K. R., 2005, “Customization in a

- unified framework for summarizing medical literature,” *Artificial Intelligence in Medicine*, Vol. 33, No. 2, pp. 179-198.
12. Erdogan, H., Sarikaya, R., Chen, S. F., Gao, Y. and Picheny, M., 2005, “Using semantic analysis to improve speech recognition performance,” *Computer Speech & Language*, Vol. 19, No. 3, pp. 321-343.
  13. Guo, Q. L. and Zhang, M., 2009, “Question answering based on pervasive agent ontology and semantic web,” *Knowledge-Based Systems*, Vol. 22, No.6, pp. 443-448.
  14. Hahu, U., Romacker, M. and Schulz, S., 2002, “MEDSYNDIKATE-a natural language system for the extraction of medical information from findings reports,” *International Journal of Medical Informatics*, Vol. 67, No. 1-3, pp. 63-74.
  15. Han, K. S., Song, Y. I., Kim, S. B. and Rim, H. C., 2007, “Answer extraction and ranking strategies for definitional question answering using linguistic features and definition terminology,” *Information Processing & Management*, Vol. 43, No. 2, pp. 353-364.
  16. He, Y. H., Hui, S. C. and Quan, T. T., 2009, “Automatic summary assessment for intelligent tutoring systems,” *Computers & Education*, Vol. 53, No. 3, pp. 890-899.
  17. Huang, M., Zhu, X. and Li, M., 2006, “A hybrid method for relation extraction from biomedical literature,” *International Journal of Medical Informatics*, Vol. 75, No. 6, pp. 443-455.
  18. Jones, M., Love, B. C., 2007, “Beyond common features: The role of roles in determining similarity,” *Cognitive Psychology*, Vol. 55, No. 3, pp. 196-231.
  19. Jung, W., Ko, Y. and Seo, J., 2005, “Automatic text summarization using two-step sentence extraction,” *Lecture Notes in Computer Science*, Vol. 3411, pp. 71-81.
  20. Ko, Y., Kim, K. and Seo, J., 2003, “Topic keyword identification for text summarization using lexical clustering,” *IEICE transactions on information and systems*, Vol. E86-D, No. 9, pp. 1695-1701
  21. Ko, Y., Park, J. and Seo, J., 2004, “Improving text categorization using the importance of sentences,” *Information Processing and Management*, Vol. 44, No. 1, pp. 65-79.
  22. 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.
23. Legara, E. F., Monterola, C. and Abundo, C., 2011, "Ranking of predictor variables based on effect size criterion provides an accurate means of automatically classifying opinion column articles," *Physica A: Statistical Mechanics and its Applications*, Vol. 390, No. 1, pp. 110-119.
  24. Li, Q. and Chen, Y. P., 2010, "Personalized text snippet extraction using statistical language models," *Pattern Recognition*, Vol. 43, No. 1, pp. 378-386.
  25. Lin, F. R. and Liang, C. H., 2008, "Storyline-based summarization for news topic retrospection," *Data & Knowledge Engineering*, Vol. 45, No. 3, pp. 473-490.
  26. Ling, X., Jiang, J., He, X., Mei, Q., Zhai, C. and Schatz, B., 2007, "Generating gene summaries from biomedical literature: A study of semi-structured summarization," *Information Processing and Management*, Vol. 43, No. 6, pp. 1777-1791.
  27. Lorch, R. F., Lorch, E. P., Ritchey, K., McGovern, L. and Coleman, D., 2001, "Effects of headings on text summarization," *Contemporary Educational Psychology*, Vol. 26, No. 2, pp. 171-191.
  28. Mingxin, W., 2011, "An architecture of information retrieval towards the semantic Web," *Energy Procedia*, Vol. 11, No. 10, pp. 4857-4861.
  29. Moens, M. F., 2007, "Summarizing court decisions," *Information Processing and Management*, Vol. 43, No. 6, pp. 1748-1764.
  30. Moens, M. F., Angheluta, R. and Dumortier, J., 2005, "Generic technologies for single- and multi-document summarization," *Information Processing and Management*, Vol. 41, No. 3, pp. 569-586.
  31. Moreda, P., Llorens, H., Saquete, E. and Palomar, M., 2011, "Combining semantic information in question answering systems," *Data & Knowledge Engineering*, Vol. 47, No. 6, pp. 870-885.
  32. Nomoto, T. and Matsumoto, Y., 2001, "A new approach to unsupervised text summarization," *Proceedings of the 24th International Conference on Research in Information Retrieval*, pp. 26-34.
  33. Oh, H. J., Myaeng, S. H. and Jang, M. G., 2012, "Effects of answer weight boosting in strategy-driven question answering," *Information Processing and Management*, Vol. 48,

No. 1, pp. 83-93.

34. Oh, H. J., Sung, K. Y., Jang, M. G. and Myaeng, S. H., 2011, "Compositional question answering: A divide and conquer approach," *Information Processing and Management*, Vol. 47, No. 6, pp. 808-824.
35. Oliva, J., Serrano, J. I. Castillo, M. D. and Iglesias, A., 2011, "SyMSS: A syntax-based measure for short-text semantic similarity," *Data&Knowledge Engineering*, Vol. 70, No. 4, pp. 390-405.
36. Ouyang, Y., Li, W., Li, S. and Lu, Q., 2011, "Applying regression models to query-focused multi-document summarization," *International Journal of Medical Informatics*. Vol. 47, No. 2, pp. 227-237.
37. Rindflesch, T. C. and Fiszman, M., 2003, "The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text," *Biomedical Informatics*, Vol. 36, No. 6, pp. 462-477.
38. Ruiz-Casado, M., Alfonseca, E. and Castells, P., 2007, "Automatising the learning of lexical patterns: An application to the enrichment of WordNet by extracting semantic relationships from Wikipedia," *Data & Knowledge Engineering*, Vol. 61, No. 3, pp. 484-499.
39. Scheffer, T., 2004, "Email answering assistance by semi-supervised text classification," *Intelligent Data Analysis*, Vol. 8, No. 5, pp. 481-493.
40. Steinberger, J., Poesio, M., Kabadjov, M. A. and Jezek, K., 2007, "Two uses of anaphora resolution in summarization," *Information Processing and Management*, Vol. 43, No. 6, pp. 1663-1680.
41. Sweeney, S., Crestani, F. and Losada, D. E., 2008, "'Show me more': Incremental length summarisation using novelty detection," *Information Processing and Management*, Vol. 44, No. 2, pp. 663-686.
42. Teng, C., Xiong, N., He, Y., Yang, L. T. and Liu, D., 2010, "A behavioural mode research on user-focus summarization," *Mathematical and Computer Modelling*, Vol. 51, No. 7-8, pp. 985-994.
43. Terol, R. M., Martínez-Barco, P. and Palomar, M., 2007, "A knowledge based method for the medical question answering problem," *Computers in Biology and Medicine*, Vol.

- 37, No. 10, pp. 1511-1521.
44. Uzuner, O., Mailoa, J., Ryan, Y. and Sibanda, T., 2010, "Semantic relations for problem-oriented medical records," *Artificial Intelligence in Medicine*, Vol. 50, No. 2, pp. 63-73.
  45. Vanderwende, L., Suzuki, H., Brockett, C. and Nenkova, A., 2007, "Beyond SumBasic: Task-focused summarization with sentence simplification and lexical expansion," *Information Processing and Management*, Vol. 43, No. 6, pp. 1606-1618.
  46. Xie, S., Liu, Y., 2010, "Improving supervised learning for meeting summarization using sampling and regression," *Computer Speech & Language*, Vol. 24, No. 3, pp. 495-514.
  47. Yang, C. C. and Wang, F. L., 2007, "An information delivery system with automatic summarization for mobile commerce," *Decision Support Systems*, Vol. 43, No. 1, pp. 46-61.
  48. Yangarber, R., Grishman, R., Tapanainen, P. and Huttunen, S., 2000, "Automatic acquisition of domain knowledge for information extraction," *Association for Computational Linguistics*, Vol. 2, pp. 940-946.
  49. Yangarber, R., 2003, "Counter-training in discovery of semantic patterns," *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 343-350.
  50. Ye, S., Chua, T. S., Kan, M. Y. and Qiu, L., 2007, "Document concept lattice for text understanding and summarization," *Information Processing and Management*, Vol. 43, No. 6, pp. 1643-1662.
  51. Yeh, J. Y., Ke, H. R., Yang, W. P. and Meng, I.H., 2005, "Text summarization using a trainable summarizer and latent semantic analysis," *Information Processing and Management*, Vol. 41, No. 1, pp. 79-95.
  52. Yin, X., Huang, J. X. and Li, Z., 2011, "Mining and modeling linkage information from citation context for improving biomedical literature retrieval," *Information Processing and Management*, Vol. 47, No. 1, pp.53-67.
  53. Zajic, D., Dorr, B. J., Lin, J. and Schwartz, R., 2007, "Multi-candidate reduction: Sentence compression as a tool for document summarization tasks," *Information Processing and Management*. Vol. 43, No. 6, pp. 1549-1570.



54. Zhou, L., Tao, Y., Cimino, J. J., Chen, E. S., Liu, H., Lussier, Y. A., Hripcsak, 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.

## 附錄、系統功能操作說明

本研究所發展之結合語意解析與文件摘要技術之知識文件檢索系統乃以本文中第4.4.1節所提出系統功能流程為依據，並開發系統之「知識文件表達項目解析模組」、「知識文件問答解析模組」、「知識文件結構化摘要推論模組」、「知識文件問答維護模組」、「知識文件管理模組」及「系統參數設定」等六大模組，以下分別介紹此六大模組中各功能詳細說明。

### A. 知識文件表達項目解析模組

本系統開發「知識文件表達項目解析模組」乃以知識文件專業語彙庫為基，並進行文件斷句、比對語彙法則以確實擷取觀念性語句。本程式乃提供權限內使用者進行知識文件表達項目解析功能。「知識文件表達項目解析」功能乃透過標點符號資料表將目標文件斷句並進行二至六字詞斷詞，以取得語句之組合詞彙並藉由斷詞與知識文件語彙法則比對，即可判定該詞是否具有知識文件之表達觀念，進而取得擷取具有知識文件之觀念性語句。

#### A.1 「知識文件表達項目解析」之使用說明

當權限內使用者選擇知識文件表達項目解析功能時，系統透過標點符號資料表進行目標文件斷句，以取得目標文件各語句，再以二至六字詞方式將每一語句進行斷詞，並透過斷詞與語彙法則比對，進而歸納具有知識文件表達之觀念性語句，並以表格方式彙整呈現。舉例而言，當使用者以文件名稱「T-104 人因工程危害預防技術」之知識文件進行解析，系統先進行斷句並取得「庫房搬運作業：」、「一樓庫房區的主要功能為存放晶元原料與半成品物件，」、「當作業人員依加工單...」等語句，再以二至六字詞方式進行斷詞以取得各語句中所含詞彙，即「庫房搬運作業」透過二至六字詞可斷詞為「庫房」、「庫房搬」、「庫房搬運」、「庫房搬運作」、「庫房搬運作業」等詞彙，藉由斷詞與語彙法則比對結果以得知「庫房搬運作業」符合作業名稱語彙法則，即表示「庫房搬運作業：」為有知識文件之觀念性語句，並於彙整表格中表達項目「作業名稱」之觀念性語句填入「庫房搬運作業：」（如圖 A.1 所示）。



圖 A.1、知識文件表達項目解析

## B. 知識文件問答解析模組

「知識文件問答解析模組」主要功能為解析詢問句之主要目的，並且計算與目的相關之回覆詞，再進行文件關鍵字相似性分析，同時透過平均值、中位數、四分位數與使用者自行定義門檻值方式以篩選文件，進而取得與該詢問目的相關之文件。於此功能為解析詢問句之主要詢問目的，並將主要詢問目的字詞作為詢問詞，再透過統計方式尋找詢問詞與回覆詞之配對組合。透過詢問詞與回覆詞配對組合即可跟文件之關鍵字詞進行相似性分析，根據相似性計算結果排序並以平均值、中位數、四分位數與使用者自行定義門檻值以篩選文件數量並取得相關文件。

### B.1 「詢問詞解析與文件篩選」之使用說明

於本研究所開發之「詢問詞解析與文件篩選」功能，使用者可輸入搜尋字串並進行查詢，並以語意關聯方式連結知識文件之專業語彙，進而搜尋具有相關性文件。當權限內使用者進行「詢問詞解析與文件篩選」功能時，先行透過詢問詞分析以取得主要詢問目的（即為詢問詞），再根據詢問詞與回覆詞配對組合以連結相關回覆詞並作為文件選取之依據（如圖 B.1 所示）。根據回覆詞與回覆句出現次數進行頻率計算，以得知於該回覆詞於此回覆句中出現機率，亦即可根據機率篩選出詢問詞與回覆詞配對組合（如圖 B.2 與圖 B.3 所示）。當取得詢問詞與回覆詞配對組合即可跟文件之關鍵字詞進行餘弦相似性分析，當數值越接近於 1 亦表示文件越符合詢問目的，並以相似值為基進行排序，且提供四種不同門檻值以篩選文件。其門檻值可分為平均數（全部文件之相似值加總除

以文件總筆數)、中位數(透過相似值排序後再取中間值)、四分位數(透過相似值排序後再取第三四位數值)與使用者自行定義(如圖 B.4)。使用者可使用下拉式選單方式選擇不同門檻值,當使用者選取使用者自行定義時,則跳出文字方塊以輸入篩選筆數(如圖 B.5),待門檻值選擇完畢即可進行查詢。其查詢結果乃先呈現未篩選前文件與詢問詞相似值排序,再根據門檻值顯示文件篩選表單,根據表單呈現即可取得篩選方式、篩選門檻數值、篩選過後之文件以及文件相關資訊(如圖 B.6 至圖 B.9),使用者可根據表單各欄位進行排序,以不同角度篩選文件(如圖 B.10),亦可點選詳細資料進行單一文件查詢,以取得該筆文件詳細內容加強篩選機制(如圖 B.11),若文件符合需求即可點選文件下載以下載文件(如圖 B.12 與圖 B.13)。

舉例而言,當使用者輸入詢問句「我手腳很酸痛」作為搜尋字串進行查詢,透過詢問詞解析與篩選文件功能中即可取得該詢問句之主要詢問目的為「酸痛」,根據詢問詞「酸痛」經過詢問詞、回覆句與回覆詞配對解析,以取得「酸痛」之配對組合包含「營造業」、「年紀」、「長時間」等(如圖 B.3)。藉由詢問詞與回覆詞配對組合「酸痛:營造業、年紀、長時間」,即可與文件之關鍵字詞進行餘弦相似性分析並根據門檻值(平均數、中位數、四分位數與自行定義)以篩選文件。若以平均數作為門檻值其門檻值為0.8619333,並取得最具相關性文件依序為「人因工程工作場所改善方法-馬達底座加工作業」、「人因工程工作場所改善方法-庫房搬運作業」與「人因工程工作場所改善方法-車椅組裝室作業」等,並提供各文件之上傳時間與相關關鍵字等資訊(如圖 B.6),亦可根據表單之相關關鍵字進行排序,以呈現不同角度之篩選結果(如圖 B.10),並可點選查詢以取得該筆文件內容、作者等詳細資訊(如圖 B.11),若該文件符合需求即可按下載連結以取得文件(如圖 B.12 至圖 B.13)。



圖 B.1、詢問詞解析與文件篩選(1)



圖 B.2、詢問詞解析與文件篩選(2)



圖 B.3、詢問詞解析與文件篩選(3)



圖 B.4、詢問詞解析與文件篩選(4)



圖 B.5、詢問詞解析與文件篩選(5)



圖 B.6、詢問詞解析與文件篩選(6)



圖 B.7、詢問詞解析與文件篩選(7)

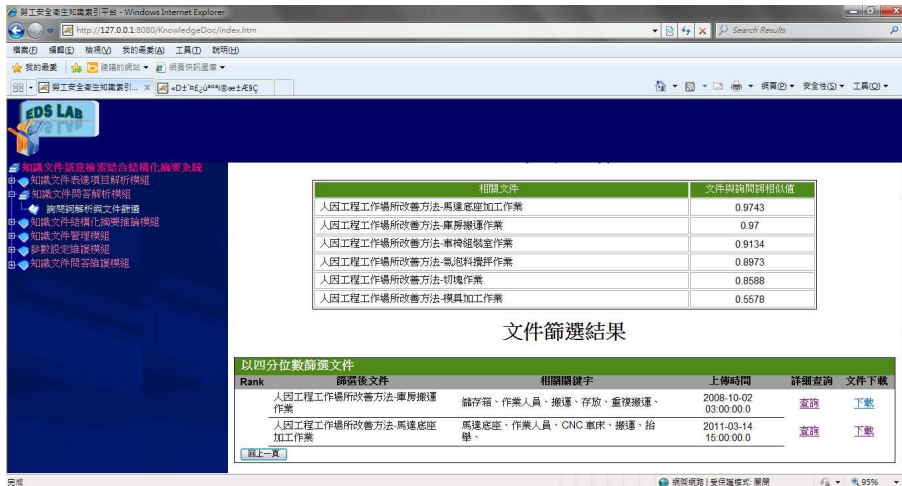


圖 B.8、詢問詞解析與文件篩選(8)



圖 B.9、詢問詞解析與文件篩選(9)

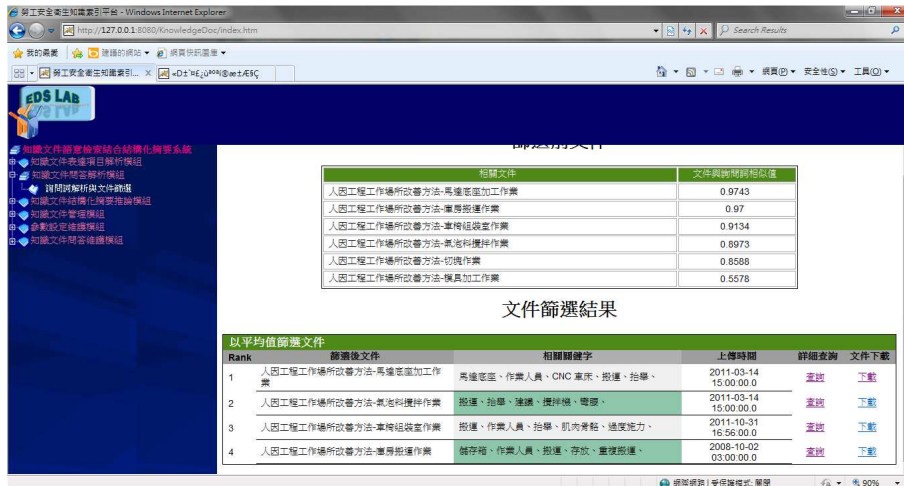


圖 B.10、詢問詞解析與文件篩選(10)

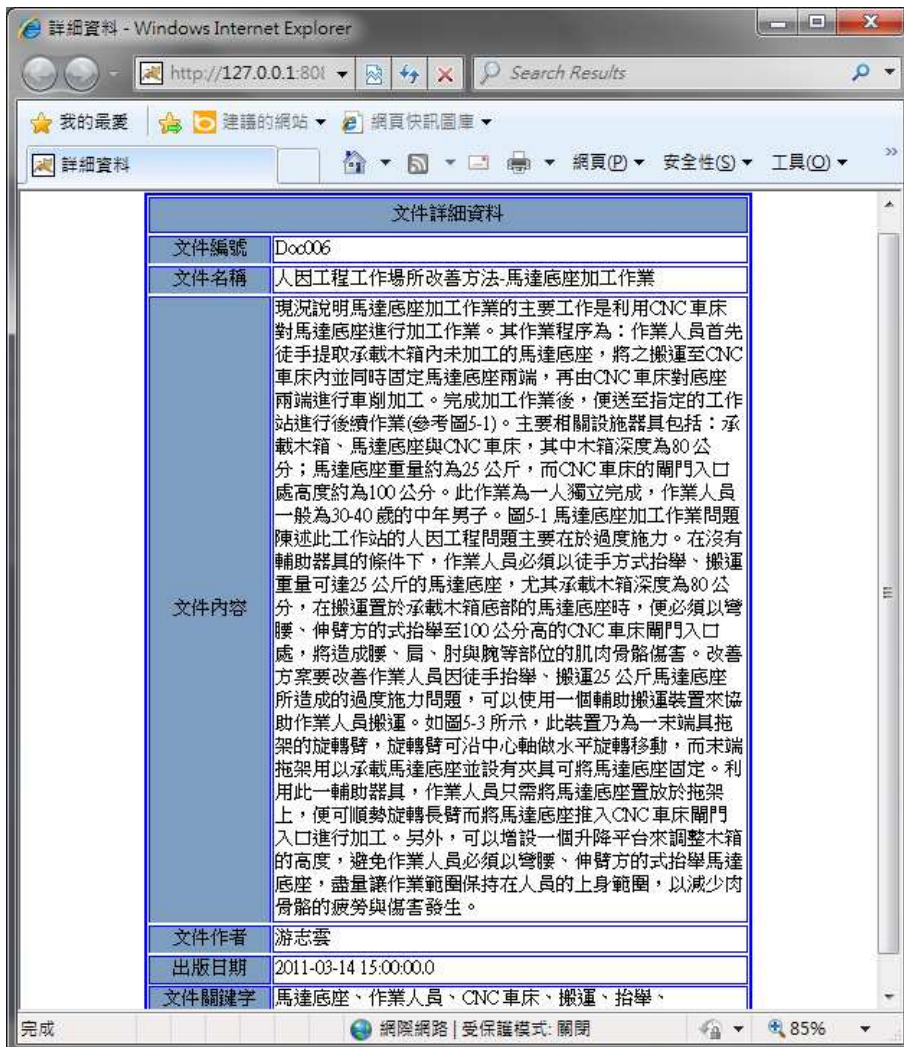


圖 B.11、詢問詞解析與文件篩選(11)



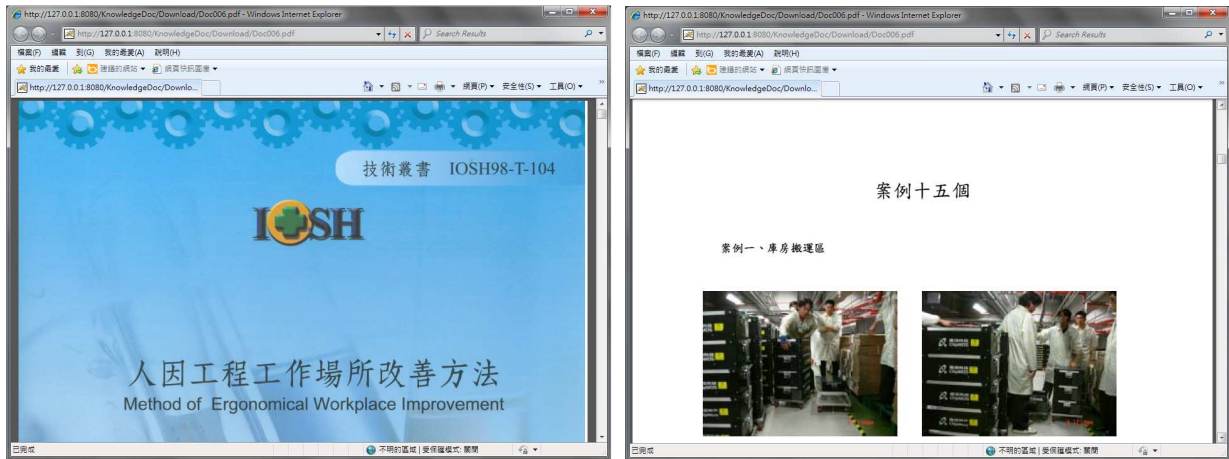


圖 B.12、詢問詞解析與文件篩選(12)

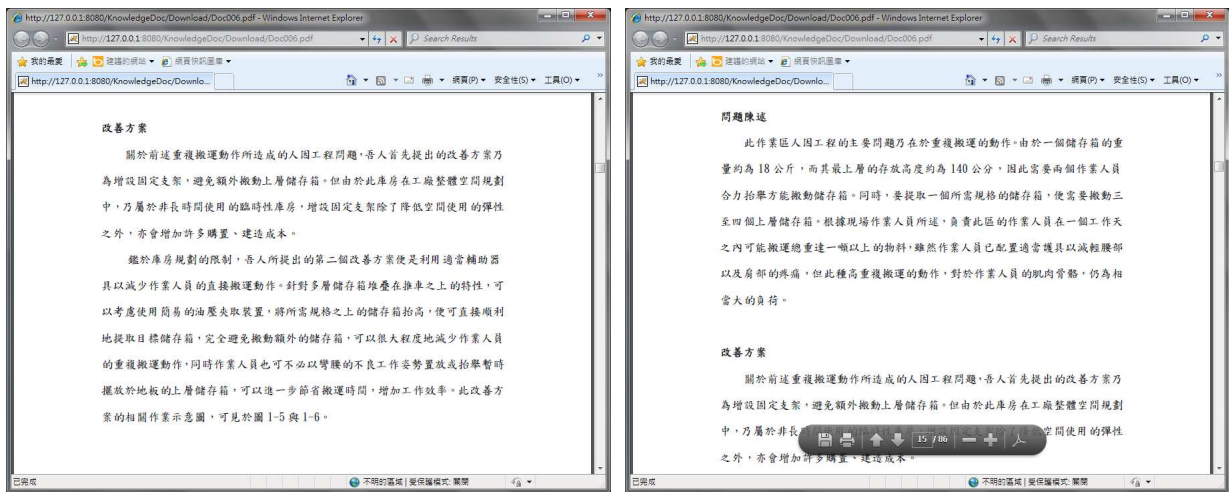


圖 B.13、詢問詞解析與文件篩選(13)

### C. 知識文件結構化摘要推論模組

本系統開發「知識文件結構化摘要推論模組」乃包含「簡略結構化摘要推論」功能及「詳述結構化摘要推論」功能；其中，「簡略結構化摘要推論」功能主要為簡略性描述文件之「動機」與「目的」，使知識搜尋者能於短時間內決定與篩選文件。透過知識文件表達項目解析，以取得具有知識文件之觀念性語句，進而以語句向心性以及結構性進行篩選以擷取最具代表性之觀念性語句。於此功能主要以觀念性語句之向心性以及結構完整性進行計分與擷取。觀念性語句之向心性即檢視該語句是否具有額外之觀念性表達項目，假設該文句同時存在多種觀念性表達項目即表示該語句具有代表性，同時計算語句之結構強度。因考量摘要閱讀之順暢性，於語句之結構強度計算中乃以主詞、受詞與動詞作為判斷依據，透過三種詞性關聯程度以判斷語句之完整性。最後，向心性分數與結構性分數分別乘上權重值再予以加總為總分數，以篩選並擷取成簡略結構化摘要。

「詳述結構化摘要推論」主要功能為描述文件之細部內容，如文件中作業行為、作業環境以及作業時間等。透過詞彙與詞彙之間關聯程度與相依程度計算，同時考量語意順暢性加以判斷語句應歸屬位置，進而取得詞彙鏈並串接語句以形成摘要。於此功能主要進行語彙相依程度計算形成語彙整體先後順序關係，並以連結詞判斷連接文句、標點符號以判斷結束文句即取得候選文句應歸屬位置，以形成制式化結構性摘要。首先以標記方式進行兩詞彙間連結機率之計算，以取得關聯程度後再以詞彙發生於段落之起始與結束之頻率以判定詞彙整體先後順序關係，進而透過順序之關係作為文句連結之依據以預測下一順序之文句，從中以各法則加強判斷語句所屬位置，進而形成詳述結構性摘要。

### **C.1-「簡略結構化摘要推論」之使用說明**

當權限內使用者選擇「簡略結構化摘要推論」功能時，系統則讀取需解析文件並透過「知識文件表達項目解析」功能取得知識文件之觀念性語句。由於簡略摘要多描述文件之動機目的與結果部份，因此在此功能乃取得「知識文件表達項目解析」功能之作業名稱、傷害成因以及改善方法等三個表達項目之觀念性語句，進而計算各語句知文句向心性以及文句結構化程度。針對向心性則判斷該表達項目之觀念性語句是否存在其他表達項目例如：專業動詞、作業環境之作業工具等，以表示該語句同時存在多種表達涵義。針對語句結構化程度部份，則以主詞、動詞與受詞等三種詞性進行判斷語句之結構性，若同時存在三種詞性則為完整結構，若完全不存在則為非結構。最後，根據向心性程度以及結構化程度分別進行加總以取得作業名稱語句總分數、傷害成因語句總分數以及改善方法語句總分數，並透過門檻值進行篩選再以作業名稱、傷害成因、改善方法依序排列形成簡略結構化摘要推論。

舉例而言，當使用者輸入「人因」作為搜尋字串進行查詢再以文件名稱「T-104 人因工程危害預防技術」之知識文件進行解析（如圖 C.1、圖 C.2 所示），系統先取得該文件之內容如「庫房搬運作業：一樓庫房區的主要功能為存放晶元原料與半成品物件，...」（如圖 C.3 所示），透過「知識文件表達項目解析」功能以得知文件各表達項目之觀念性語句，並依照表達項目之語句進行向心性計算以取得各語句性心性分數。當該觀念性語句同時存在其他表達項目語彙，向心性分數則逐筆加分，最後以加總方式取得向心性總分數。如「作業身分」表達項目之觀念性語句「此時便需要兩個作業人員合力將所有上層的儲存箱抬舉並放置在旁側地板上」，同時存在「專業動詞」表達項目之語彙「抬舉」與「彎腰」，因此該語句之向心性總分數為 2，亦表示該語句除「作業身分」表達項

目外另存在「專業動詞」表達涵義。由於「作業身分」表達項目乃以「專業動詞」與「作業工具」等兩筆表達項目進行比對，因此語句之向心性分數若為 2，即表示該語句除本身表達項目外同時存在「專業動詞」或「作業工具」表達涵義較具備文件代表性；反之，若向心性分數為 0 即表示該語句僅存在單筆表達項目，較不具備文件之代表性（如圖 C.4 所示）。除向心性分數計算之外，於此功能同時亦進行文句結構性計算以取得各語句之結構性分數，如「作業身分」表達項目之觀念性語句「此時便需要兩個作業人員合力將所有上層的儲存箱抬舉並放置在旁側地板上，」同時存在主詞「作業人員」、動詞「抬舉」與受詞「儲存箱」因此結構性分數為 3，即表示該語句具備主詞、動詞與受詞以完整表達語句並具有結構性（如圖 C.4 所示）。透過向心性總分與結構性總分分別賦予所對應權重值，並進行加總作為總分數，再根據總分數排序與篩選以篩選得分最高前兩句作為最具代表性語句以形成簡略結構化摘要，如圖 C.5 所示「作業身分」表達項目觀念性語句分別加乘向心性權重值 3 與結構性權重值 1，以取得各語句之總分數。當中，於「作業身分」表達項目最高分數為 12 分，所對應語句為「同時作業人員也可不必以彎腰的不良工作姿勢置放或抬舉...」，其次為「因此需要兩個作業人員合力抬舉方能搬動儲存箱。」此語句之總分為 9 分，因而篩選為簡略摘要並以藍色註記作為文件之作業行為描述、傷害成因則以紅色註記，改善方法則以綠色註記。

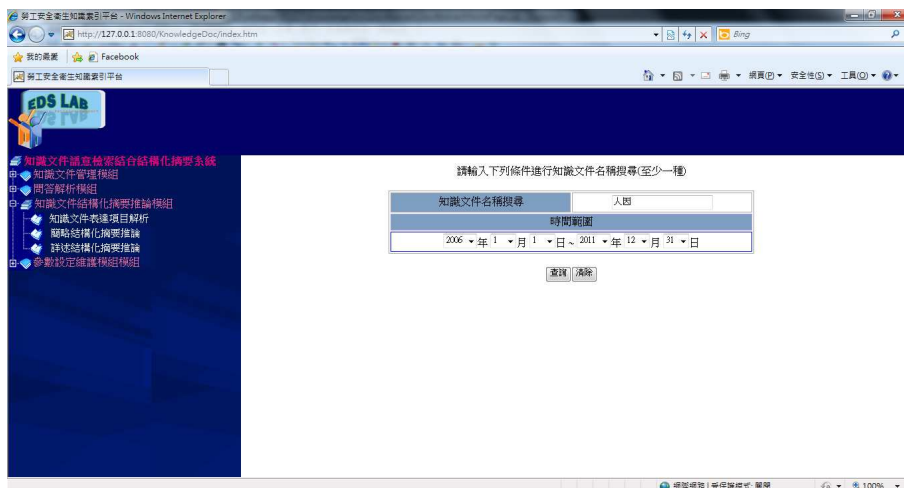


圖 C.1、簡略結構化摘要推論畫面(1)



圖 C.2、簡略結構化摘要推論畫面(2)



圖 C.3、簡略結構化摘要推論畫面(3)



圖 C.4、簡略結構化摘要推論畫面(4)



圖 C.5、簡略結構化摘要推論畫面(5)

## C.2-「詳述結構化摘要推論」之使用說明

透過本研究所開發之「詳述結構化摘要推論」功能，使用者可了解知識文件細部內容，如作業行為、環境與時間等，當權限內使用者選擇「詳述結構化摘要推論」功能時，系統則讀取需解析文件並透「詞頻計算」功能取得文件中出現次數頻繁之詞彙（即稱之頻率辭），並以作業環境、作業行為、傷害成因與改善方法等表達項目之專業語彙進行關聯分析，以取得文件之詞彙鏈與語彙整體先後順序關係。於「詞頻計算」功能中，首先乃以結束標點符號將文件斷句，再進行二字六字詞拆解，從中剔除非關鍵字並將拆解詞彙與原文件進行比對，以取得出現於文件之專業詞彙與出現次數較多之非專業詞彙，同時記錄該詞彙之出現次數（如圖 C.9 所示），進而得知於文件中出現頻繁之詞彙（頻率辭）。當取得頻率辭後，即可與各表達項目之專業語彙進行分析，當中所比對之專業表達項目包含：作業環境、作業行為、傷害成因與改善方法等。當語句存在專業語彙時，即判斷下一語句是否存在頻率辭，若存在頻率辭則表示該專業語彙與該頻率辭具有關聯性，反之則無關聯性。於過程中同時計算頻率詞與下一語句出現之頻率辭之次數，並計算該詞彙作為開頭之加總次數（起之總和）與作為結尾之加總次數（迄之總和），最後則以起迄相除取得詞彙鏈判斷係數（如圖 C.10 左示）。並排序以得知詞彙重要詞彙鏈順序（如圖 C.10 右示）。

舉例而言，當使用者輸入「人因」作為搜尋字串進行查詢再以文件名稱「T-104 人因工程危害預防技術」之知識文件進行解析（如圖 C.6、圖 C.7 所示），系統先取得該文件之內容如「庫房搬運作業：一樓庫房區的主要功能為存放晶元原料與半成品物件，...」（如圖 C.8 所示），透過「詳述結構化摘要推論」功能以得知文件中具有專業詞彙以及

於文件中之非專業字詞次數（如圖 C.9 所示）。當文件中某一語句「現況說明一樓庫房區的主要功能為存放晶元原料與半成品物件，...」當中存在專業詞彙「搬運」出現 9 次與非專業詞彙「放於」出現 3 次等。透過篩選方式以取得出現頻率較高之詞彙與專業詞彙，並進行該詞彙與下一詞彙之出現組合次數計算，以取得每一詞彙之開頭次數與結尾次數（如圖 C.10 所示）。如文件中專業詞彙「主要功能」開頭次數為 91 且結尾次數為 5，並取得詞彙鏈判斷係數 18.2，而非專業詞彙「規格」開頭次數為 293 且結尾次數為 187，並取得詞彙鏈判斷係數 1.567 等，透過判斷係數之排序進而得知詞彙鏈依序出現為「主要功能」、「移動」、「原料」等，並以此詞彙鏈進行詳述摘要之呈現，分別以「作業描述」（橘色標記）、「問題描述」（黃色標記）與「改善方法」（綠色標記）表示之。

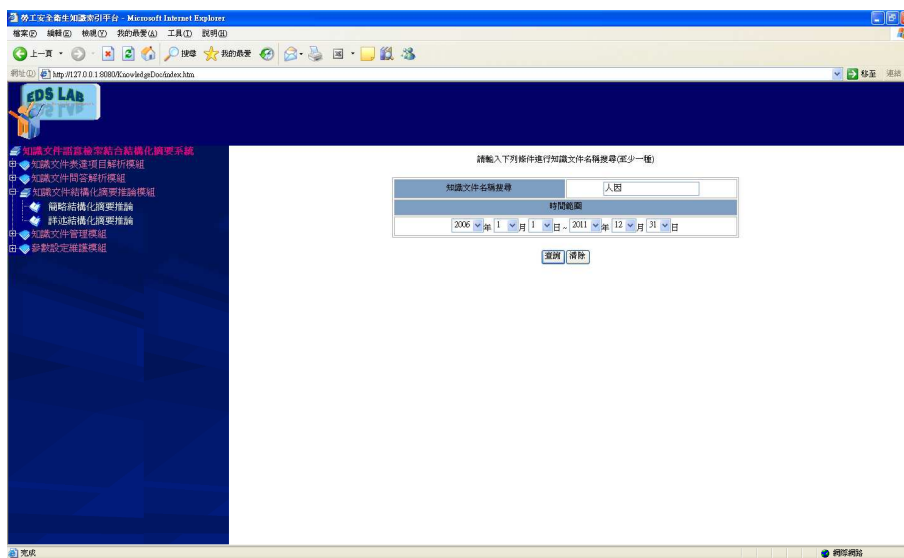


圖 C.6、詳述摘要畫面(1)

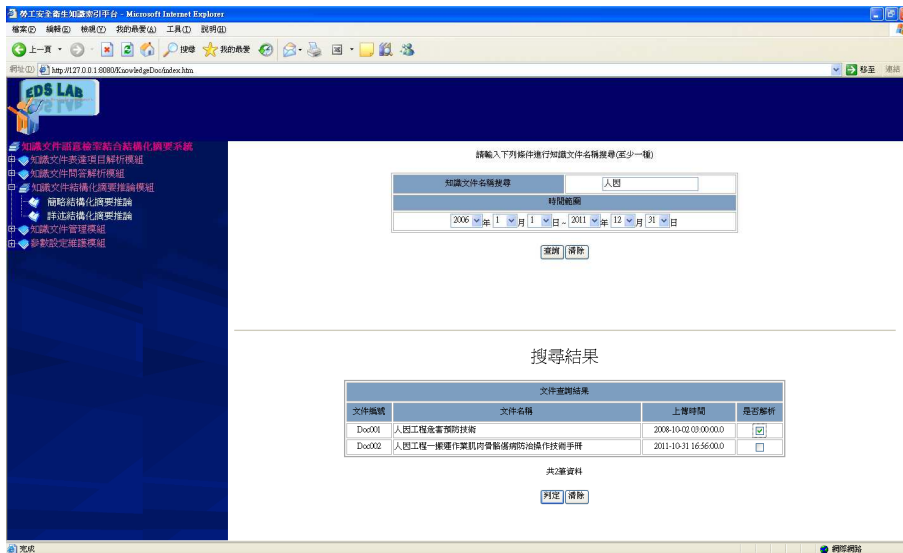


圖 C.7、詳述摘要畫面(2)



圖 C.8、詳述摘要畫面(3)



圖 C.9、詳述摘要畫面(4)

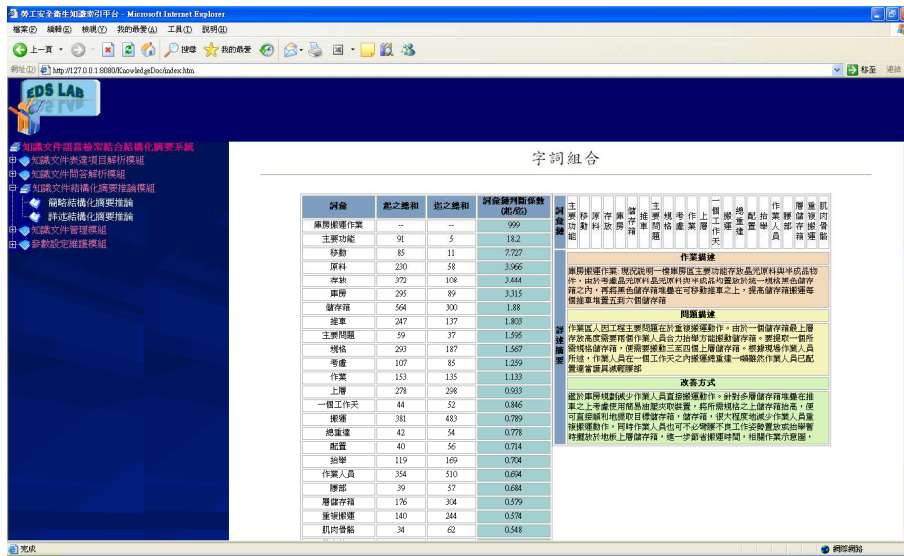


圖 C.10、詳述摘要畫面(5)

#### D. 「知識文件管理模組」之程式邏輯

「知識文件管理模組」即維護知識文件其中包含「文件新增」、「文件查詢」、「文件刪除」與「文件修改」等四大功能，使用者可於此進行文件之新增、查詢、修改與刪除等功能。透過「文件新增」功能即可匯入文件於系統資料庫內，以解析推論知識文件表達語彙與結構化摘要推論，使系統可達精確之判定效果。藉由「文件查詢」功能乃提供權限內使用者查詢所有文件之內容，以方便使用者瞭解系統內各項文件資料之維護結果。此外，「文件刪除」與「文件修改」功能乃提供權限內使用者進行修改與維護錯誤文件資料，進而保持文件內容之正確性與即時性。

##### D.1- 「文件新增」之使用說明

當權限內使用者執行文件新增功能時，系統乃提供文件新增之系統介面予權限內使用者，即可透過「文件新增」功能將文件資料匯入並維護於系統資料庫中。權限內使用者可於該介面輸入欲新增之文件名稱、文件內容、文件上傳時間與文件作者等基本資料，並上傳文件。待權限內使用者輸入/點選介面之所有選項並按下「上傳」鍵後（如圖 D.1 所示），系統即完成文件之新增作業，並輸出文件新增資訊如：文件名稱、文件作者與上傳時間等訊息提供予使用者，並可選「繼續閱讀」以檢視完整內容（如圖 D.2 與圖 D.3 所示）。

舉例而言，當使用者點選文件上傳功能並根據文件名稱「點燈測試作業」、文件作者「游志雲」、上傳時間「2012年6月17日」與文件內容「點燈測試作業是作業員將直



徑極細小的電燈管放置桌面的兩電極上檢驗燈管是否可點亮之測試。...」等輸入相關資訊，同時瀏覽欲上傳之文件並點選「上傳」按鈕（如圖 D.1 所示）。待上傳完畢後，即呈現資料新增成功訊息，並輸出文件之基本資料如文件名稱「點燈測試作業」、文件作者「游志雲」、上傳時間「2012 年 6 月 17 日」等，並於資料表下方呈現文件內容，使用者可點選「繼續閱讀」以完整檢視文件內容（如圖 D.3 所示）。

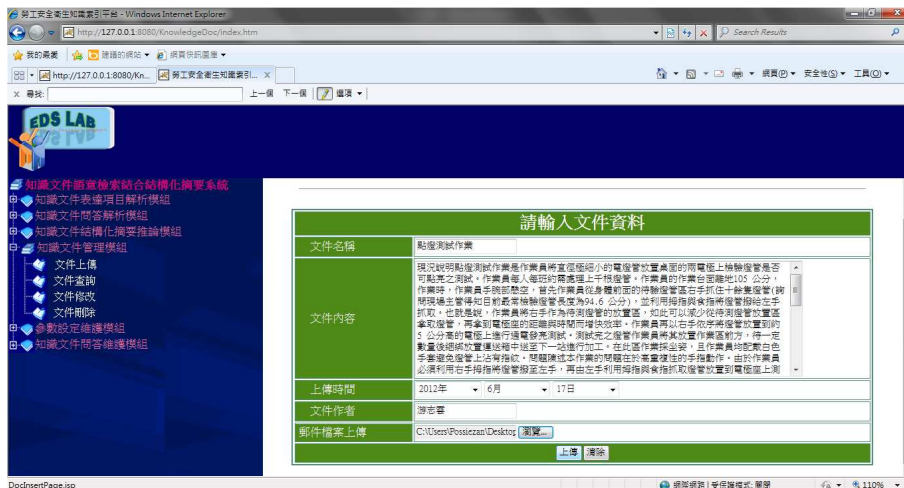


圖 D.1、知識文件新增(1)



圖 D.2、知識文件新增(2)



圖 D.3、知識文件新增(3)

## D.2-「文件查詢」之使用說明

為方便權限內使用者查詢所需之文件資料，本系統乃開發「文件查詢」功能，以提供權限內使用者查詢已上傳之文件內容。當權限內使用者選擇文件查詢功能時，於下拉式選單選取查詢條件並輸入查詢字串與時間範圍搜尋（如圖 D.4 與圖 D.5 所示）。待權限內使用者輸入查詢字串、選取查詢條件與時間範圍等條件，並按下「查詢」鍵後，系統將符合該查詢條件之文件資料顯示於系統頁面下方。此外，權限內使用者亦可於此介面中點選「詳細資料」之連結功能，系統以彈跳視窗方式顯示文件詳細資料（如圖 D.6 與圖 D.7 所示），亦可點選「已分析」檢視文件表達項目解析判定、簡略摘要判定與詳述摘要等判定結果（如圖 D.8 與圖 D.9 所示），並完成查詢文件之步驟。

當權限內使用者執行「文件查詢」功能時，系統乃提供條件查詢欄位供使用者輸入查詢條件「人因」並以「文件名稱」為查詢字串，同時限制文件時間範圍於「2007 年 1 月 1 日」至「2012 年 12 月 31 日」之間進行篩選，即可取得「人因工程工作場所改善方法-庫房搬運作業」、「人因工程工作場所改善方法-車椅組裝室作業」與「人因工程工作場所改善方法-氣泡料攪拌作業」等六筆文件，此外，若權限內使用者點選「詳細資料」連結功能，系統以彈跳視窗方式顯示該文件之詳細資料如文件「人因工程工作場所改善方法-車椅組裝室作業」詳細內容為「車椅組裝室現況說明車椅組裝...」（如圖 D.6 與圖 D.7 所示）。當該文件已完成推論之判定，權限內使用者即可點選「已分析」連結功能，並以彈跳視窗方式顯示該文件之文件表達項目分析結果、簡略結構化摘要結果與詳述結構化摘要推論結果（如圖 D.8 與圖 D.9 所示）。



圖 D.4、知識文件查詢(1)



圖 D.5、知識文件查詢(2)

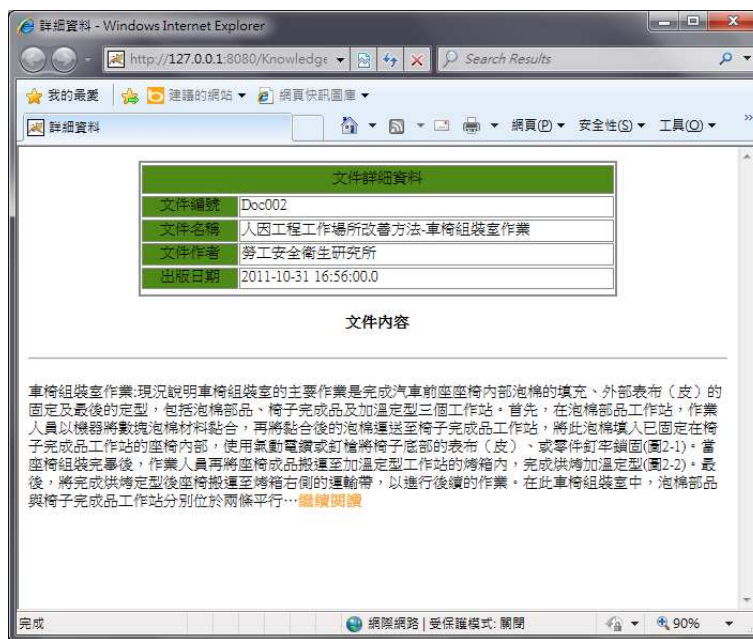


圖 D.6、知識文件查詢(3)



圖 D.7、知識文件查詢(4)

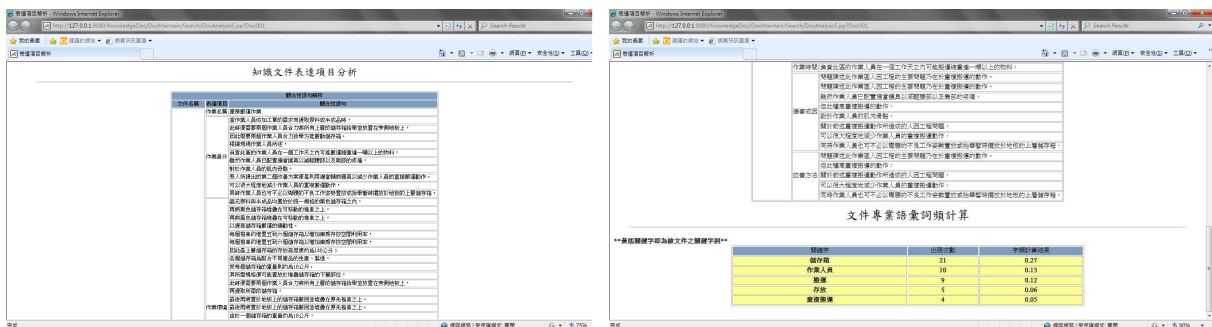


圖 D.8、知識文件查詢(5)

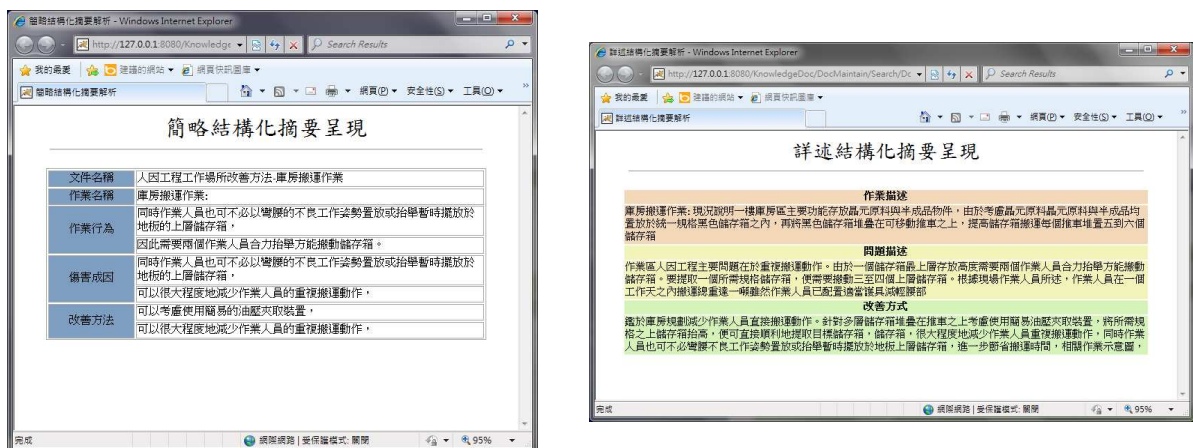


圖 D.9、知識文件查詢(6)

### D.3-「文件修改」之使用說明

文件修改功能乃提供權限內之使用者修改或錯誤之文件資料，進而保持文件內容之正確性與即時性。當權限內使用者選擇文件修改功能時，於下拉式選單選取查詢條件輸

入查詢字串與選擇時間範圍搜尋（如圖 D.10）。待權限內使用者輸入查詢字串、選取查詢條件、選擇時間範圍等條件，並按下「查詢」鍵後，系統將符合該查詢條件之文件資料顯示於系統頁面下方（如圖 D.10）。此外，權限內使用者亦可於此介面中點選「資料修改」之連結功能，系統以彈跳方式顯示修改文件基本資料視窗，使用者於修改視窗更改錯誤資料（如圖 D.11 所示），並於詳細資料視窗下方按下「確定修改」鍵後，系統自動將該筆文件資料於系統資料庫中修改，同時執行訊息「文件修改成功」顯示於系統介面上（如圖 D.12 所示），即表示資料修改完成。

舉例而言，當權限內使用者執行「文件修改」功能時，系統乃提供條件查詢欄位供使用者輸入。當使用者以「人因」並以「文件名稱」等條件篩選文件，以取得相關文件如「人因工程工作場所改善方法-車椅組裝室作業」與「人因工程工作場所改善方法-氣泡料攪拌作業」等，並點選「人因工程工作場所改善方法-氣泡料攪拌作業」文件之「資料修改」系統以彈跳視窗方式顯示修改文件基本資料介面（如圖 D.11 所示），使用者即可修改文件作者更新為「游志雲」按下「確定修改」鍵後，系統即呈現資料更新成功訊息同時重新呈現資料更新後畫面（如圖 D.12 所示）。



圖 D.10、知識文件修改(1)



圖 D.11、知識文件修改(2)

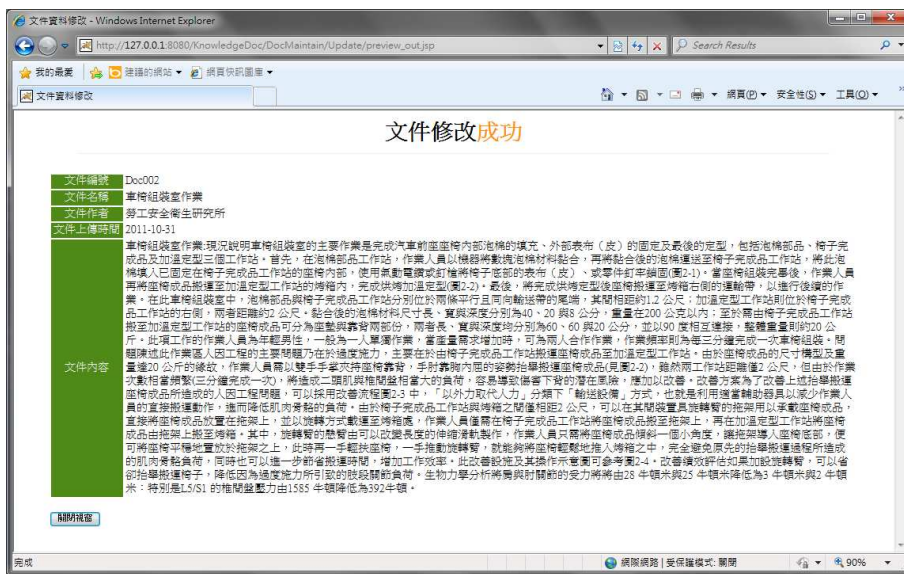


圖 D.12、知識文件修改(3)

#### D.4-「文件刪除」之使用說明

文件刪除功能乃提供權限內之使用者刪除錯誤之文件，進而保持文件資料之正確性與即時性。當權限內使用者選擇文件資料刪除功能時，於下拉式選單選取查詢條件輸入查詢字串與選擇時間範圍搜尋等條件。待權限內使用者輸入查詢字串、選取查詢條件與

時間範圍等篩選條件，並按下「查詢」鍵後，系統將符合該查詢條件之文件顯示於系統頁面下方（如圖 D.13 所示）。此外，權限內使用者亦可於此介面中點選「詳細資料」之連結功能，系統以彈跳視窗方式顯示欲刪除文件之詳細資料（如圖 D.14 所示），且於詳細資料視窗中按下「刪除」鍵後，系統自動將該筆文件於系統資料庫中刪除，同時執行訊息「資料已成功刪除」顯示於系統介面上（如圖 D.15 所示），即完成刪除文件之步驟。

當權限內使用者執行「文件刪除」功能時，系統乃提供條件查詢欄位供使用者輸入。當使用者根據篩選條件「文件名稱」出現「作」之搜尋字串等條件進行篩選，以搜尋相關文件「庫房搬運作業」、「車床組裝室作業」與「人因工程工作場所改善方法-氣泡料攪拌作業」等，權限內使用者即可點選「庫房搬運作業」文件之「詳細資料」連結功能，系統以彈跳視窗方式顯示欲刪除文件詳細資料（如圖 D.14 所示），使用者即可按下刪除欄位之「刪除」按鈕以刪除文件，同時執行訊息「資料已成功刪除」顯示於系統介面上（如圖 D.15 所示）。



圖 D.13、知識文件刪除(1)



圖 D.14、知識文件刪除(2)

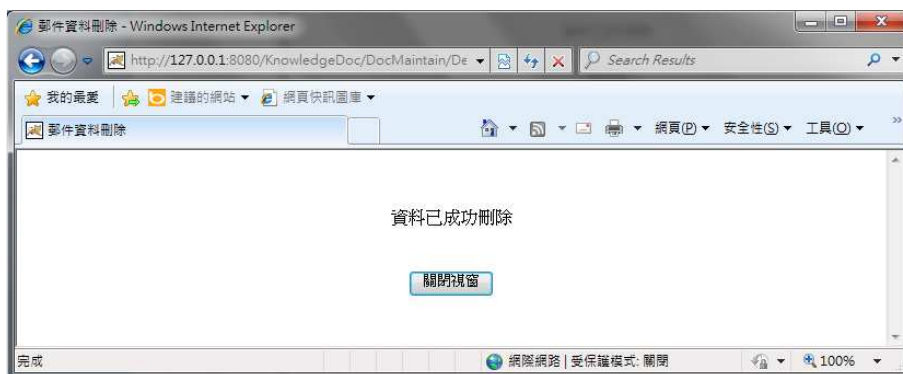


圖 D.15、知識文件刪除(3)

## E. 參數設定維護模組

為使權限內使用者方便維護各系統相關資料，本系統乃開發「參數設定模組」。本程式功能乃提供權限內使用者於線上修改各系統參數資料。系統參數設定模組包含「系統門檻值設定」與「系統權重值設定」兩功能；其中，「系統門檻值設定」功能乃提供權限內使用者進行修改與維護各模組之門檻如文件關鍵字篩選門檻、詢問詞與回覆詞配對組合篩選門檻等設定，以加強系統準確性；「系統權重值設定」乃提供權限內使用者進行修改與維護簡略結構化摘要推論、詳述結構化摘要推論之權重質設定，進而保持系



統參數之正確性。

### **E.1- 「系統門檻值設定」使用說明**

參數設定功能乃提供權限內之使用者修改系統參數，進而提升文件表達項目解析、問答解析與結構化摘要推論等模組之正確性，本系統乃開發「系統門檻值設定」與「系統權重值設定」兩功能。當權限內使用者選擇「系統門檻值設定」功能時，系統即呈現表達項目解析模組之文件關鍵字篩選筆數門檻參數、結構化摘要模組之簡略結構化摘要輸出筆數門檻參數、知識文件詞彙鏈詞頻篩選門檻與詢問詞回覆詞配對組合門檻參數等，並提供各參數之門檻值說明以解釋門檻值定義與用處（如圖 E.1 所示）。待權限內使用者點選設定鈕後，系統即跳出門檻值設定頁面並顯示修改資料與欄位，當使用者更改數值完畢即可按下「確定修改」（如圖 E.2 所示），系統自動將系統參數數值於系統資料庫中修改，同時執行訊息「資料修改成功」顯示於系統介面（如圖 E.3 所示）。

舉例而言，當權限內使用者執行「系統門檻值設定」功能時，系統即呈現表達項目解析模組之文件關鍵字篩選筆數門檻參數名稱「Kwcnt」、參數說明「知識文件關鍵字筆數設定...」以及門檻值為「5」；結構化摘要模組之簡略結構化摘要輸出筆數門檻參數名稱「DocEZtotal」、參數說明「簡略結構化摘要各表達項目呈現筆數設定，即門檻值設置越高...」以及門檻值為「2」等各參數門檻值名稱、說明與門檻值等。當使用者點選「文件關鍵字篩選筆數門檻值」之「設定」連結，系統即呈現「門檻值設定頁面」並顯示該筆門檻值之參數名稱「Kwcnt」與門檻值為「5」，使用者將門檻值為「5」修改「3」（如圖 E.2 所示），並按下「確定修改」鍵後，系統自動於系統資料庫中修改，同時執行訊息「資料修改成功」顯示於系統介面上（如圖 E.3 所示）。



圖 E.1、系統門檻值檢視畫面

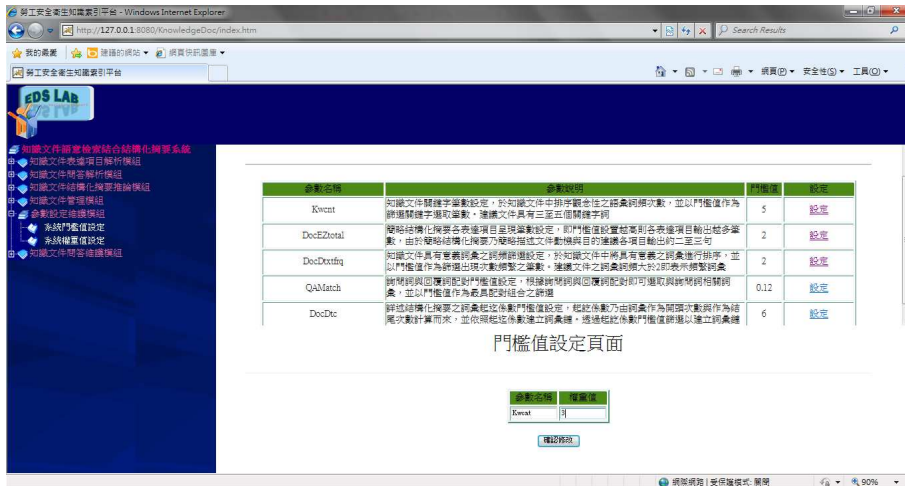


圖 E.2、系統門檻值設定畫面

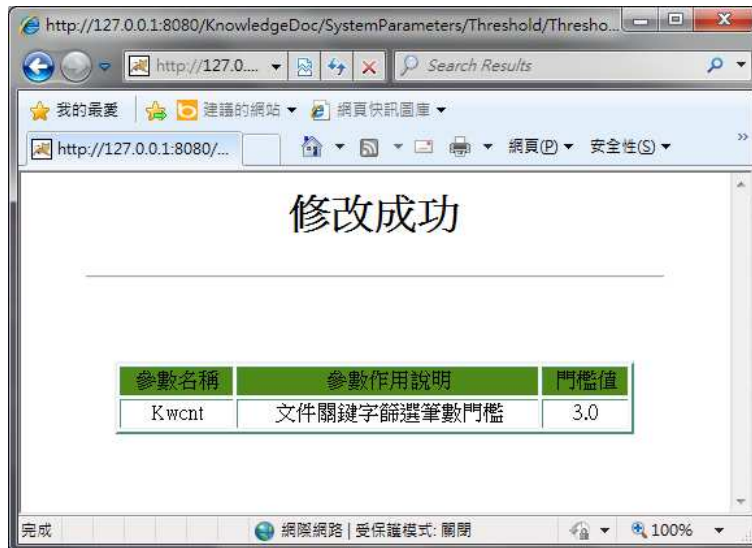


圖 E.3、系統門檻值修改成功畫面

## E.2- 「系統權重設定」使用說明

當權限內使用者選擇「系統權重值設定」功能時，系統即呈現結構化摘要模組之簡略結構化摘要向心性權重值與簡略結構化摘要結構性權重值，並提供各權重值說明以解釋權重值定義與用處（如圖 E.4 所示）。待權限內使用者點選設定鈕後，系統即跳出權重值設定頁面並顯示修改資料與欄位，當使用者更改數值完畢即可按下「確定修改」（如圖 E.5 所示），系統即顯示「資料修改成功」顯示於系統介面亦表示系統資料庫修改完成（如圖 E.6 所示）。

舉例而言，當權限內使用者執行「系統權重值設定」功能時，系統系統即呈現結構化摘要模組之簡略結構化摘要向心性權重值參數名稱「DocEZhart」、參數說明「簡略結構化摘要之向心性分數權重值。於簡略結...」以及門檻值為「3」；簡略結構化摘要結構性權重值參數名稱「DocEZstr」、參數說明「簡略結構化摘要之結構性分數權重值。於簡略結構化摘要...」以及門檻值為「1」等相關資訊（如圖 E.4 所示），當使用者點選「簡略結構化摘要向心性權重值」之「設定」連結，並將門檻值為「3」修改「4」（如圖 E.5 所示），待修改完畢並按下「確定修改」鍵後，系統自動於系統資料庫中修改，同時執行訊息「資料修改成功」顯示於系統介面上（如圖 E.6 所示）。



圖 E.4、系統權重值檢視畫面



圖 E.5、系統權重值設定畫面

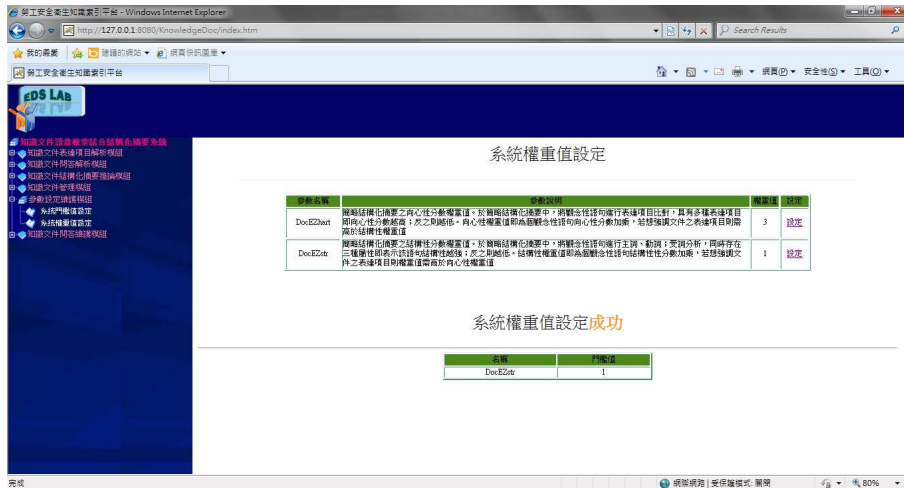


圖 E.6、系統權重值設定成功畫面

## F. 知識文件問答維護模組

「知識文件問答維護模組」即維護知識文件詢問詞與回覆詞其中包含「詢問詞與回覆詞新增」、「詢問詞與回覆詞查詢」、「詢問詞與回覆詞刪除」與「詢問詞與回覆詞修改」等四大功能，使用者可於此進行詢問詞與回覆詞配對組合新增、查詢、修改與刪除等功能。透過「詢問詞與回覆詞新增」功能即可匯入詞彙於系統資料庫內，以進行問答解析模組推論，使系統精確判定詞彙語意達到問答效果。藉由「詢問詞與回覆詞查詢」功能乃提供權限內使用者查詢所有詢問詞與回覆詞資訊，以方便使用者瞭解系統內詢問詞與回覆詞資料之維護結果。此外，「詢問詞與回覆詞刪除」與「詢問詞與回覆詞修改」功能乃提供權限內使用者進行修改與維護錯誤資料，進而保持問答解析之詢問詞與回覆詞正確性與即時性。

### F.1-詢問詞與回覆詞新增說明

當權限內使用者執行詢問詞與回覆詞新增功能時，系統乃提供詢問詞與回覆詞新增介面予權限內使用者，即可透過「詢問詞與回覆詞新增」功能將詢問詞與回覆詞資料匯入並維護於系統資料庫中。首先，權限內使用者可於該介面輸入詢問詞，並根據詢問詞搜尋知回覆句以及回覆詞並按下確定鍵送出（如圖 F.1 所示），待系統解析配對組合完畢即呈現配對組合結果（如圖 F.2 與圖 F.3 所示），權限內使用者即可按下「新增配對結果」鍵完成詢問詞與回覆詞新增作業，系統亦輸出詢問詞與回覆詞新增資訊如：詢問詞、相關回覆詞與各配對係數等資訊（如圖 F.4 所示）。

舉例而言，當使用者輸入詢問詞「肌肉」並根據肌肉搜尋取得回覆句「一般抬舉重物對於腰椎會形成數倍甚至數十倍於物品重量之受力，因...」以及該回覆句主要回覆詞「人工抬舉」、「腰椎」、「搬運」，待輸入五筆回覆句以及十五個回覆詞等相關資訊，並點選「確定」按鈕（如圖 F.1 所示）。系統即呈現配對分析結果與詢問詞「肌肉」具有相關回覆詞包含「搬運」、「人工抬舉」、「腰椎」等詞彙（如圖 F.2 與圖 F.3 所示），使用者即可點選「新增配對結果」將上述配對組合以完成新增作業，系統新增成功後乃呈現新增成功訊息，並輸出詢問詞與回覆詞新增資料如詢問詞「肌肉」、回覆詞「搬運」以及配對係數為「1」等資訊（如圖 F.4 所示）。



圖 F.1、詢問詞與回覆詞輸入畫面



圖 F.2、詢問詞與回覆詞配對結果畫面(1)



圖 F.3、詢問詞與回覆詞配對結果畫面(2)



圖 F.4、詢問詞與回覆詞新增成功畫面

## F.2-詢問詞與回覆詞查詢說明

為方便權限內使用者查詢系統之詢問詞與回覆詞，本系統乃開發「詢問詞與回覆詞查詢」功能，以提供權限內使用者查詢已新增之詢問詞與回覆詞。當權限內使用者選擇詢問詞與回覆詞查詢功能時，於下拉式選單選取查詢條件並輸入查詢字串進行查詢（如圖 F.5 所示）。待權限內使用者輸入查詢字串、選取查詢條件與時間範圍等條件，並按下「查詢」鍵後，系統將符合該查詢條件之文件資料顯示於系統頁面下方（如圖 F.6 所示）。

當權限內使用者執行「詢問詞與回覆詞查詢」功能時，系統乃提供條件查詢欄位於使用者輸入查詢條件「肌肉」並以「詢問詞」為查詢字串，同時限制文件時間範圍於「2007年1月1日」至「2012年12月31日」之間進行篩選（如圖 F.5 所示），即可取得詢問

詞「肌肉」以及相關回覆詞「搬運」、「抬舉」等六筆詢問詞與回覆詞組合，以及各組合係數「1.0」、「0.27」、「0.27」等（如圖 F.6 所示）。



圖 F.5、詢問詞與回覆詞查詢畫面



圖 F.6、詢問詞與回覆詞查詢結果畫面

### F.3-詢問詞與回覆詞修改說明

詢問詞與回覆詞修改功能乃提供權限內之使用者修改或錯誤之詢問詞與回覆詞資料，進而保持詢問詞與回覆詞詞彙之正確性與即時性。當權限內使用者選擇詢問詞與回覆詞修改功能時，於下拉式選單選取查詢條件輸入查詢字串與選擇時間範圍搜尋（如圖 F.7）。待權限內使用者輸入查詢字串、選取查詢條件、選擇時間範圍等條件，並按下「查詢」鍵後，系統將符合該查詢條件之文件資料顯示於系統頁面下方（如圖 F.8）。此外，權限內使用者亦可於此介面中點選「資料修改」之連結功能，系統以彈跳方式顯示修改



詢問詞與回覆詞料視窗，使用者於修改視窗更改錯誤資料（如圖 F.9 所示），並按下「資料確定修改」鍵後系統自動將該筆資料於系統資料庫中修改，同時執行訊息「修改成功」顯示於系統介面上（如圖 F.10 所示），即表示資料修改完成。

舉例而言，當權限內使用者執行「詢問詞與回覆詞修改」功能時，系統乃提供條件查詢欄位供使用者輸入。當使用者以「肌肉」並以「詢問詞」等條件篩選詢問詞與回覆詞，以取得相關詞彙「般運」、「抬舉」等（如圖 F.8）。，並點選「肌肉」與「般運」配對組合之「資料修改」，系統以彈跳視窗方式顯示修改文件基本資料介面（如圖 F.9 所示），使用者即可修改回覆詞「般運」更新為「搬運」並按下「確定修改」鍵後，系統即呈現資料更新成功訊息同時重新呈現資料更新後畫面（如圖 F.10 所示）。



圖 F.7、詢問詞與回覆詞查詢畫面



圖 F.8、詢問詞與回覆詞查詢結果畫面



圖 F.9、詢問詞與回覆詞修改畫面

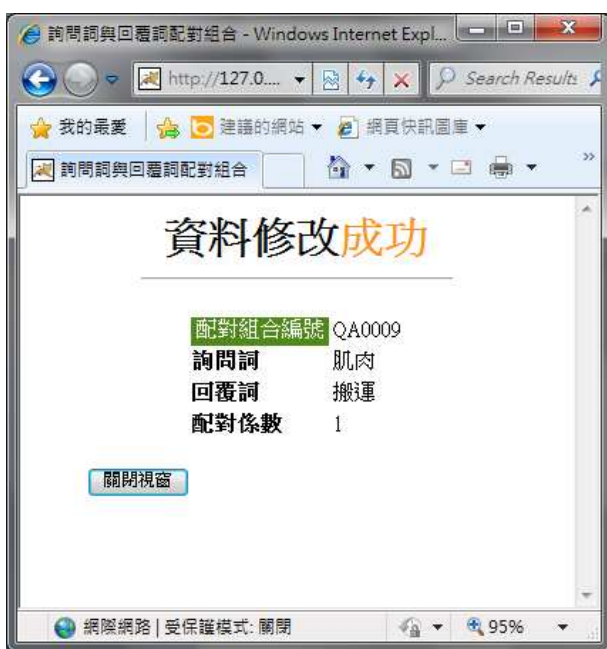


圖 F.10、詢問詞與回覆詞修改成功畫面

#### F.4-詢問詞與回覆詞刪除說明

詢問詞與回覆詞刪除功能乃提供權限內之使用者刪除錯誤之詢問詞與回覆詞詞彙，進而保持資料之正確性與即時性。當權限內使用者選擇詢問詞與回覆詞刪除功能時，於下拉式選單選取查詢條件輸入查詢字串與選擇時間範圍搜尋等條件。待權限內使用者輸入查詢字串、選取查詢條件與時間範圍等篩選條件，並按下「查詢」鍵後，系統將符合該查詢條件之詢問詞與回覆詞詞彙顯示於系統頁面下方（如圖 F.11 所示）。權限內使用者亦可於此介面中點選「資料刪除」之連結功能（如圖 F.12 所示），系統以彈跳視窗方式顯示欲刪除詢問詞與回覆詞之資料進行確認（如圖 F.13 所示），且於資料確認視窗中按下「刪除」鍵後，系統自動將該筆詢問詞與回覆詞組合於系統資料庫中刪除，

同時執行訊息「資料已成功刪除」顯示於系統介面上（如圖 F.14 所示），即完成刪除文件之步驟。

當權限內使用者執行「詢問詞與回覆詞刪除」功能時，系統乃提供條件查詢欄位供使用者輸入。當使用者根據篩選條件「詢問詞」出現「肌肉」之搜尋字串等條件進行篩選，以取得詢問詞「肌肉」與相關回覆詞「搬運」、「抬舉」等組合結果，權限內使用者即可點選「肌肉」與「搬運」組合之「資料刪除」連結功能，系統以彈跳視窗方式顯示資料確認視窗以及欲刪除之資料（如圖 F.13 所示），使用者即可按下刪除欄位之「刪除」按鈕以刪除文件，同時執行訊息「資料已成功刪除」顯示於系統介面上（如圖 F.14 所示）。



圖 F.11、詢問詞與回覆詞查詢畫面



圖 F.12、詢問詞與回覆詞查詢結果畫面

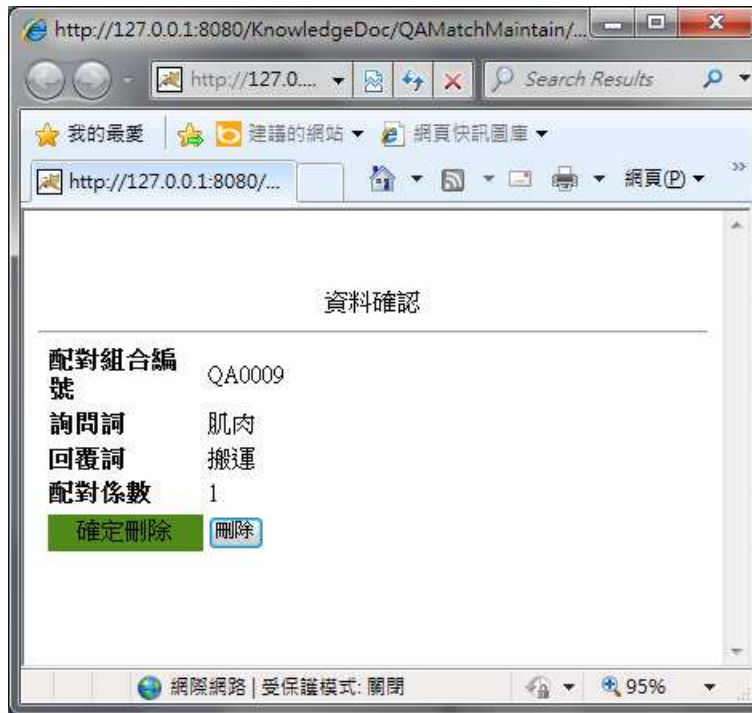


圖 F.13、詢問詞與回覆詞資料確認畫面

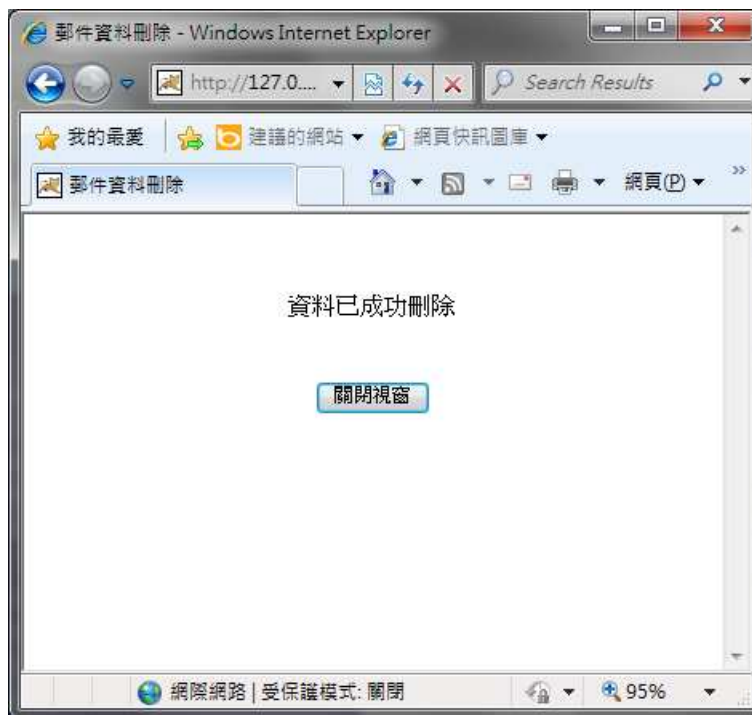


圖 F.14、詢問詞與回覆詞刪除成功畫面