

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

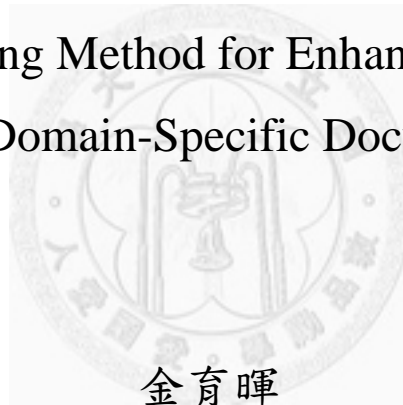
National Taiwan University

Master Thesis

運用特定領域文件集合進行自我學習之知識本體增補

技術

A Self-Learning Method for Enhancing Ontology  
from Domain-Specific Documents



金育暉

Yu-Huei Jin

指導教授：謝尚賢 博士

Advisor: Shang-Hsien Hsieh, Ph.D.

中華民國 101 年 8 月

August, 2012

國立臺灣大學碩士學位論文  
口試委員會審定書

運用特定領域文件集合進行自我學習之知識本體增補技術

A Self-Learning Method for Enhancing Ontology

from Domain-Specific Documents

本論文係金育暉君 ( R99521610 ) 在國立臺灣大學土木工程學系碩士班完成之碩士學位論文，於民國 101 年 07 月 12 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

謝 尚 賢

(指導教授)

王 明 德

郭 榮 欽

吳 翌 禎

謝尚賢

王明德

郭榮欽

吳翌禎

系主任

呂 良 正

呂良正

(簽名)

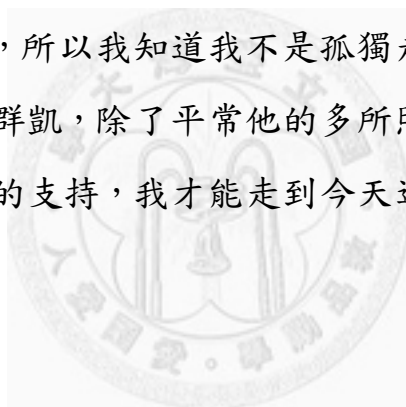
## 誌謝

在 CAE 的兩年奮鬥，終於走到了這一頁，雖然這是論文的開頭，卻是整個碩士生活的結尾。一路走來顛顛簸簸，受到了許多人的幫助與指導，要說聲感謝的人的不計其數。而最為感謝的人則是一路給予我指導的謝尚賢老師，謝老師從我大四修學士論文開始，一路給我研究上的指導並告訴我研究上所需注意的事項。謝老師也相當關心我的生活，讓我在碩士生活中得以安心地做研究，在此深深的感謝他。而在口試時不吝給予我意見的王明德老師、郭榮欽老師、吳翌禎老師也要深深感謝，三位老師都對於我的研究給了很多有用的意見，讓這一篇論文更加完整，並感謝國科會研究計畫（NSC 100-2221-E-002 -224 -MY2）對研究經費上的補助。

同時也要感謝在研究路上一直給予我幫助的乃文學長，他讓我從一個不懂資訊檢索的學生到以資訊檢索作為研究的題目並順利的畢業，這一路有了他的指導讓我少走了許多彎路，也少做了許多無用之功。另外還要特別感謝孟涵學姊，除了對於我的研究提供了許多建議，她還會三不五時對我在研究生活中所需要注意的事項給予提醒，也常常提供了不同的思維給我，讓我對於生活有了新的想像與想法。

再來要感謝 R99 的夥伴們，蔡承諺、jacob、matthew、小古、郭品岑、劉柔君、義彬、南同學、BK、三凱、瀚嶸，我們一起奮鬥了兩年，製造了許多笑點，讓兩年生活不只有研究的苦悶，而是還有許多與你們一起分享的歡樂。而在這兩年認識的 CAE 學長學弟們也都非常感謝，像是奕竹總是被我以奇奇怪怪的問題騷擾、pattern 不斷地嗆飛我、KID 抓我去打羽球、濟華大大不斷幫我製造新的封號……等等，湘如跟銘峰也幫我解決了許多有的沒有的事情。總之在 CAE 的兩年認識了許多有趣的學長學弟學姊學妹，讓這兩年的歡樂無限。

最後要感謝高中的那群，這兩年來跟你們的每次聚會都是一種充電，讓我又有了繼續往前的動力，那群中有人的研究生生活太好命，有人的研究生生活太悲慘，所以我知道我不是孤獨走在這條路上的。而在那群中要特別感謝黃群凱，除了平常他的多所照應，在研究生生活的最後一段路是因為有他的支持，我才能走到今天這個終點，也才有了這一頁的誕生。



## 摘要

知識本體 (Ontology) 是一種用來表達知識的方式，其被廣泛的應用於各種領域中。知識本體利用簡潔的方式表示該領域中所含有的知識概念以及知識概念相互之間的關係，讓知識本體可以被電腦解讀並加以使用。然而目前在建置知識本體的流程上，並未有統一的標準與方法，且在建置知識本體的過程中，需要大量領域專家的投入，並隨著領域中知識的發展，必須不斷對知識本體進行增補。因此如何以自我學習的方式來建議知識本體的增補內容，以減少領域專家人力的參與，是建置知識本體時的重要議題。

本研究的主要目的為縮短增補特定領域的知識本體內容時，所需花費的時間與人力。為此在本研究中選定「建築資訊塑模 (Building Information Modeling)」領域作為研究的特定領域，由分析塑模領域的文獻開始，收集文獻中所整理的知識概念以建置塑模領域的基本知識本體，並建立該領域的文件集合，以驗證使用知識本體的資訊檢索技術之成效表現乃優於使用向量空間模型的資訊檢索技術，再針對資訊檢索的檢索結果進行知識概念擷取，來對知識本體的內容進行補充。本論文將詳述各步驟所進行之研究內容，並先透過實驗來驗證應用知識本體進行資訊檢索確實能提高資訊檢索的成效表現，再驗證經過資訊檢索技術所增補的知識本體亦能提高資訊檢索之成效。

**關鍵字：**知識本體、建築資訊模型、建築資訊塑模、資訊檢索

## ABSTRACT

Ontology, which has been widely used in different domains, concisely represents the knowledge as a set of concepts and the relations of those concepts. However, with the growth of the domain knowledge and its lack of unified standards, building and revising the ontology is not only time-consuming but also requires a large amount of manpower. For making the process more efficient, this research proposed a self-learning method to suggest the enhancement of the ontology on a specific domain.

In this research, the specific domain is focused on building information modeling (BIM). There are three steps of the research progress. First, collect concepts from the related researches as reference to build the base ontology. Then, propose an ontology-based retrieval model to improve the retrieval effectiveness. Finally, propose a methodology to extract the concepts from the ontology-based retrieval results. According to the experiment results, using the enhanced ontology to the ontology-based retrieval could improve the retrieval effectiveness. The enhanced ontology can also help learning and sharing the domain knowledge.

**Keyword:** Ontology, Building Information Model, BIM, Building Information Modeling, Information Retrieval

# 目錄

誌謝.....	iii
摘要.....	v
ABSTRACT.....	vi
目錄.....	vii
圖目錄.....	ix
表目錄.....	xi
第一章 緒論.....	1
1-1 研究背景與動機.....	1
1-2 研究目的及範圍.....	1
1-3 研究方法及步驟.....	2
1-4 論文架構.....	3
第二章 建立特定領域之知識本體.....	5
2-1 知識本體簡介.....	5
2-2 以特定領域之知識來源建置知識本體.....	7
2-2-1 建築資訊塑模簡介.....	8
2-2-2 收集特定領域之知識概念.....	10
2-2-3 建立階層與關聯性.....	12
2-2-4 知識本體展示.....	15
第三章 特定領域之資訊檢索.....	18
3-1 資訊檢索簡介.....	18
3-2 建立特定領域之測試文件集.....	20
3-2-1 選擇與準備文件.....	21
3-2-2 定義資訊需求.....	22

3-2-3 評估文件相關度.....	22
3-3 運用知識本體於特定領域資訊檢索之成效評估.....	26
3-3-1 向量空間模型簡介.....	26
3-3-2 以知識本體為基礎之資訊檢索方法 .....	27
3-3-3 資料檢索成效評估.....	31
第四章 以資訊檢索技術增補知識本體 .....	33
4-1 增補知識本體之流程 .....	33
4-1-1 擷取知識概念.....	35
4-1-2 自動化增補知識本體 .....	41
4-1-3 領域專家協助修訂.....	46
4-1-4 領域知識本體.....	46
4-1-5 系統建置與限制.....	47
4-2 自動化增補知識本體之成效評估.....	48
4-2-1 自動化增補知識本體流程之參數設定 .....	48
4-2-2 自動化增補知識本體展示 .....	55
第五章 結論與建議.....	60
5-1 結論 .....	60
5-2 未來研究方向建議.....	61
參考文獻.....	63



## 圖目錄

圖 1 本研究之進行步驟 .....	3
圖 2 建置知識本體之流程[5] [6] .....	7
圖 3 BIM 基本知識本體 .....	14
圖 4 擴展後的 BIM 基本知識本體 .....	15
圖 5 中文版 BIM 基本知識本體 .....	17
圖 6 資料檢索流程圖[18] .....	19
圖 7 研討會文件掃描檔 .....	21
圖 8 測試文件集範例 .....	22
圖 9 向量空間模型資訊檢索系統 .....	27
圖 10 知識本體資訊檢索系統 .....	28
圖 11 概念分層架構 .....	29
圖 12 權重地圖 .....	30
圖 13 運用資訊檢索技術增補知識本體之流程 .....	34
圖 14 bigram 切分結果 .....	36
圖 15 trigram 切分結果 .....	36
圖 16 自動化增補知識本體之流程 .....	42
圖 17 中文版 BIM 基本知識本體 .....	43
圖 18 $K=3$ 時，不同 $N$ 值的平均求準率 .....	50
圖 19 $K=4$ 時，不同 $N$ 值的平均求準率 .....	51
圖 20 $K=5$ 時，不同 $N$ 值的平均求準率 .....	51
圖 21 $N=3$ 時，不同 $K$ 值的平均求準率 .....	52
圖 22 $N=4$ 時，不同 $K$ 值的平均求準率 .....	53
圖 23 $N=5$ 時，不同 $K$ 值的平均求準率 .....	54

圖 24	$N=3$ 、 $K=4$ 的平均求準率 .....	55
圖 25	第一次增補後的知識本體 .....	56
圖 26	第二次增補後的知識本體 .....	57
圖 27	第三次增補後的知識本體 .....	58
圖 28	第四次增補後的知識本體 .....	59



## 表目錄

表 1 BIM 應用項目使用分類[11] .....	11
表 2 BIM 工具使用分類[12] .....	12
表 3 專家評估結果 .....	23
表 4 Cohen's Kappa measure 計算結果 .....	24
表 5 overlap value 計算結果.....	24
表 6 兩種不同資訊檢索模型的平均求準率 .....	32
表 7 文件關鍵字排名表 .....	39
表 8 知識概念關鍵字排名表 .....	44
表 9 「規劃與設計」的平均求準率 .....	49
表 10 「營運」的平均求準率 .....	49



# 第一章 緒論

## 1-1 研究背景與動機

知識本體 (Ontology) 是一種用來表達知識的方式，其被廣泛的應用於各種領域中，如生醫領域用知識本體表達蛋白質相關知識[1]；地球科學領域的 SWEET (Semantic Web for Earth and Environmental Terminology) 計畫[2]。因為知識本體利用簡潔的方式表示該領域中所含有的知識概念與知識概念相互之間的關係，讓知識本體可以被電腦解讀並加以使用，因此在不同的專業領域上，都會有建置與使用知識本體的需求。

然而目前在建置知識本體的流程上，並未有統一的標準與方法，且在建置知識本體的過程中，需要大量領域專家的投入。並隨著領域中知識的發展，必須不斷對知識本體進行修正。因此如何有效的以自動化的方式來進行知識本體的增補，以減少領域專家人力的參與，是建置知識本體時的重要議題。

## 1-2 研究目的及範圍

本研究的主要目的為縮短建置特定領域的知識本體時，所需花費的時間與人力，因此本研究提出以自動化方式從特定領域文件集中，擷取文件中的知識概念，進行知識本體內容增補之方法，並利用知識本體技術改善資訊檢索技術之成效表現。

本研究中選定「建築資訊塑模 (Building Information Modeling)」領域作為研究的特定領域。而在改善資訊檢索的成效表現上，則是以搜尋結果的準確率為優先考慮的準則，而不以檢索過程所耗費的時間作為評估的標準。

## 1-3 研究方法及步驟

本研究步驟如圖 1，共分為四個步驟，各步驟的內容與採用之方法如下：

### 1. 建置特定領域之基本知識本體

本研究由特定領域相關的文獻資料中，整理並分析該領域中的知識概念，並將這些知識概念建置成該特定領域的基本知識本體。

### 2. 建立特定領域測試文件集

因選用的建築資訊塑模領域，目前並無合適的測試文件集，因此在這個步驟中先建置所需的測試文件集，以進行資訊檢索的成效評估。

### 3. 利用知識本體進行資訊檢索之成效評估

使用第二個步驟中所建置的測試文件集，測試以知識本體進行資訊檢索方法與一般資訊檢索方法之間的成效比較，以驗證使用知識本體技術進行資訊檢索可改善特定領域的資訊檢索成效。

### 4. 利用資訊檢索技術增補知識本體

利用以知識本體進行資訊檢索的檢索結果進行分析，來增補基本知識本體的內容，並將增補後的知識本體再次進行使用知識本體的資訊檢索方法。將其檢索結果與只使用基本知識本體的資訊檢索結果及未使用知識本體的資訊檢索結果進行成效比較，以驗證增補過後之知識本體確能提高資訊檢索的成效表現。

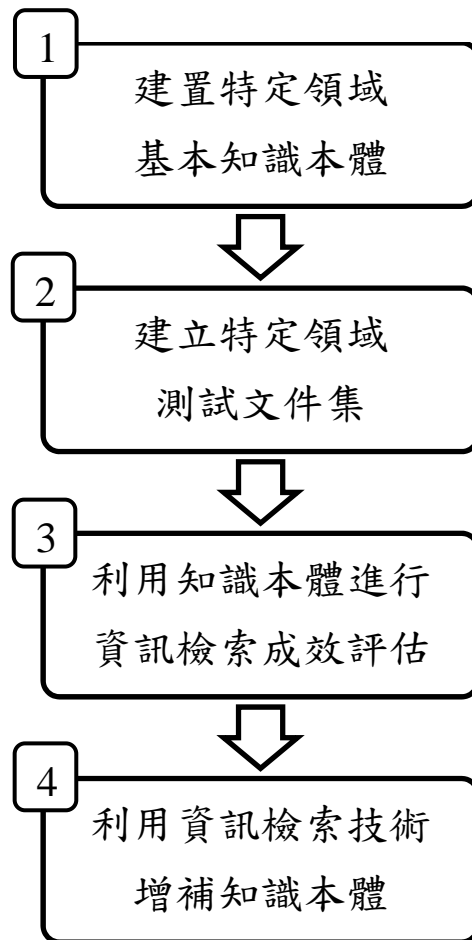


圖 1 本研究之進行步驟

## 1-4 論文架構

本文的內容依照實作流程整理成以下章節：

### 第二章 建立特定領域之知識本體

本章將簡介知識本體與選用的特定領域：建築資訊塑模，並說明目前知識本體於建築資訊塑模的應用以及說明本研究所使用的特定領域之基本知識本體。

### 第三章 特定領域之資訊檢索

本章將說明如何建置特定領域之測試文件集，並利用特定領域文件集合進行以知識本體技術進行的資訊檢索，並與向量空間模型的資訊檢索方式進行成效比較。

#### 第四章 以資訊檢索技術增補知識本體

本章利用第三章以知識本體進行資訊檢索的檢索結果進行分析，擷取出特定領域文件集中的知識概念，以藉此增補知識本體中的知識概念，並利用增補後的知識本體再次進行資訊檢索並與第三章的結果進行成效比較。

#### 第五章 結論與建議

對於本研究實作與分析結果做出結論，並針對所面臨之問題提出改善之建議。



## 第二章 建立特定領域之知識本體

### 2-1 知識本體簡介

知識本體為一種整理與儲存知識的方式，其定義為將一知識內容用清楚並標準化的方式「概念化」[3]，而其中「概念化」的含意為將知識內容以不失原意的方式將其內容簡化。因此使用知識本體的方式為當獲得一個知識時，將其知識內容簡化成一知識概念，接著將該知識概念跟其他原有的知識概念進行連結。此項連結可能會包含有此知識概念為另一知識概念的實例，或是此知識概念是另一知識概念的解釋，或更多不同的關聯性。利用這樣的方式，將不同知識概念進行連結。更重要的是可以藉由這樣的連結關係，去推論出兩個未有直接連結的知識概念之間的關係。而將知識本體以 OWL (Web Ontology Language)[4]的格式儲存後，就可以使電腦理解並儲存知識概念。更進一步的可以讓電腦利用知識本體的內容來進行知識概念之間的邏輯推論。

由於每個知識領域對於知識本體的運用方式不同，所以知識本體就需要依照每個知識領域的使用目的進行建置，因此在建置知識本體時的步驟就會有所不同。如圖 2 左為 Noy 與 McGuinness 對於他們所設計的知識本體建置軟體 Protégé，在進行知識本體建置時建議的操作步驟[5]，共分為六個步驟，由決定知識領域範圍開始，先使用已存在的知識本體作為基礎，接著列舉重要的知識概念並定義知識概念彼此之間的關係與屬性。而圖 2 右則為 Uschold 與 Grüninger 對於在進行知識本體建置時的所建議的步驟[6]，共分為四個步驟，首先為確定知識本體的使用目的與應用範圍，再進行知識本體內容的抓取，接著將知識本體編碼儲存後，融入現有知識本體。



分析兩者建置知識本體的步驟後，可以發現兩者有相似之處，在圖 2 中以虛線註明並進行對照，因此可以統整出在建置知識本體的過程中重要步驟，共有三個部分：

- 確認知識本體所涵蓋的領域範圍

確認想要建置的知識本體的領域，並對該領域有一明確定義並限定該領域的範圍。

- 知識概念擷取與知識概念的關聯性建置

將在限定範圍內的知識概念擷取出來並定義不同知識概念之間的關聯性。

- 將知識本體的內容轉換為電腦可閱讀之格式

將知識本體中的知識概念與其相互之間的關係，依造電腦可閱讀的格式進行儲存。

根據以上所統整的知識本體建置流程，第一個與第二個步驟皆需要許多領域專家參與協助，其中又以第二步驟需要領域專家投入許多時間進行，以確認何項知識概念是屬於欲建置的知識本體的範圍內。且根據 Noy 與 McGuinness 的建議，建置知識本體為一不斷循環的流程，因此領域專家須不斷對知識本體的內容進行修訂，以增加該知識本體的內容與完善性。但如此一來在建置知識本體時將耗費大量的人力與時間，所以如何減少領域專家在建置知識本體時所參與的工作量，是本研究的主要目標。

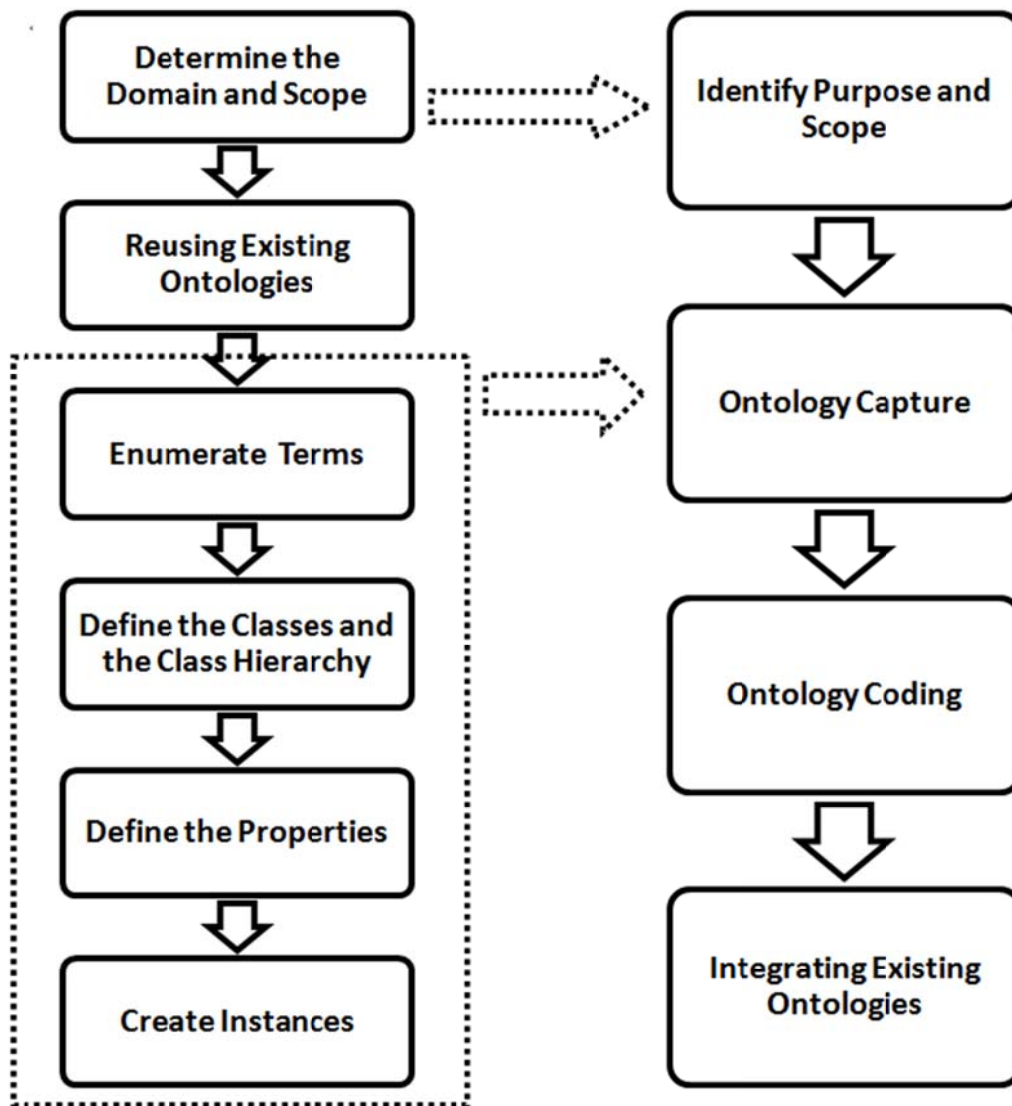


圖 2 建置知識本體之流程[5] [6]

## 2-2 以特定領域之知識來源建置知識本體

在本研究中所選定的特定領域為建築資訊塑模(Building Information Modeling)，建築資訊塑模領域為土木營建工程領域之新興領域，因此以下內容將從介紹建築資訊塑模開始，說明由建築資訊塑模領域中已有的知識來源，建置建築資訊塑模的基本知識本體。

## 2-2-1 建築資訊塑模簡介

建築資訊塑模其主要的表現方式為在三維虛擬空間中的立體模型，但其真正的意義為一整套與生產、溝通、分析相關的 3D 模型技術[7]。因此，建築資訊塑模的使用時間包含了整個建物的生命週期，從規劃設計到現場施工再到建物完成後的營運維護。

而使用建築資訊塑模所帶來的好處，也已經有相關研究[8]，分別有以下幾點：

- 更快與更有效率的流程(Faster and more effective processes)  
建物相關的資訊可以更容易分享、增值與重複使用。
- 更好的設計(Better design)  
在初期設計時，建物可以得到更嚴謹且更快的分析，因此可以在初期設計階段時使用更多不同的設計。
- 包含整個生命週期的成本與環境資料  
(Controlled whole-life costs and environmental data)  
環境對建物的影響更容易預測，對於建物生命週期中所需要的各項花費更加清楚。
- 更好的文件品質(Better production quality)  
在建物的生命週期內所需要的相關文件的可以自動化產出並更加彈性。
- 更好的客戶服務(Better customer service)  
經由 3D 模型的展示，業主可以更輕易的理解建物的設計與施工過程。
- 包含整個生命週期的資料(Lifecycle data)  
所有建物的相關資訊，可以持續使用到營運維護上。

由以上幾點，可以知道一個完整的建築資訊模型(Building Information Model)，其所包含的資料量會相當的龐大。因此，BIM Handbook[9]一書提供 M.A. Mortenson Company 所定義的一個良好的建築資訊模型必須包含六個要點：

- 數位的(Digital)

其儲存資料的方式要以數位資料方式儲存。

- 空間的(Spatial)

其數位模型資料需以虛擬三維(3D)的方式呈現。

- 可量測的(Measurable)

在數位模型資料中的資料是數值化並可計算的，如在模型中有樑的尺寸、數量，並可對各種樑的數值進行計算。

- 完整的(Comprehensive)

在數位資料中不只含有建物模型的數值資料，並包含有建物在規劃設計階段、現場施工階段等，不同階段中與該建物相關的資訊。

- 易取得的(Accessible)

其儲存的資料格式必須是不受限於單一電腦軟體或硬體，其資料檔案可在不同的軟硬體間交換，

- 持久的(Durable)

其資訊必須可以長時間保存並維持一致性於整個建物生命週期。

由此可知建築資訊塑模為運用於建物完整生命週期之技術，且其在三維虛擬空間的表現方式，讓過去在使用平面圖來表達建物設計上，因誤解圖面表示方式而導致物件互相衝突的狀況大量減少，並讓不同階段的參與者，經由建築資訊模型能有更好的溝通方式。

## 2-2-2 收集特定領域之知識概念

由於所需使用的基本知識本體，其領域範圍已限定在建築資訊塑模領域中，而目前在建築資訊塑模領域中，已經有許多文獻對於建築資訊塑模領域的內容進行整理與分析，因此在進行知識概念的收集時，即可根據這些文獻的內容，來整理出建築資訊塑模領域中較為重要的知識概念。因此在此節中將回顧建築資訊塑模的相關文獻，以收集建築資訊塑模領域的知識概念。



由美國國家建築科學研究院(NIBS/National Institute of Building Sciences) 所公布的美國國家 BIM 標準(NBIMS-US/The National BIM Standard-United States)第二版中[10]，在 Term and Definitions 章節裡，有對於建築資訊塑模領域中各式詞彙進行說明，其內容包含有該詞彙的縮寫與定義，如 Component 此詞彙在該書中的定義為，在建築物中的各項設備與該設備在虛擬三維空間中的位置。因此根據此章節的內容，就可收集到在建築資訊塑模中較為重要的知識概念。

而由 buildingSMART 於 2010 年發行的 BIM Project Execution Planning Guide 一書中[11]，其第二章(Identifying BIM Goals and Uses For a Project)的內容為針對建築物不同生命週期所使用的各種應用項目進行說明，總共提出二十五個應用項目，其內容整理後如表 1。其每個應用項目所使用的階段皆有所不同，並在該書的附錄 B 中，有針對每個項目的詳細內容進行描述，因此這些應用項目，可視為目前在建築資訊塑模領域進行應用的各項知識概念。

表 1 BIM 應用項目使用分類[11]

PLAN	DESIGN	CONSTRUCT	OPERATE
Existing Conditions Modeling			
Cost Estimation			
Phase Planning			
Programming			
Site Analysis			
Design Reviews			
Design Authoring			
Energy Analysis			
Structural Analysis			
Lighting Analysis			
Mechanical Analysis			
Other Eng. Analysis			
LEED Evaluation			
Code Validation			
3D Coordination			
Site Utilization Planning			
Construction System Design			
Digital Fabrication			
3D Control and Planning			
Record Model			
Maintenance Scheduling			
Building System Analysis			
Asset Management			
Space Mgmt/Tracking			
Disaster Planning			

	Primary BIM Uses
	Secondary BIM Uses

另一項資料來源則為 Eastman 所編寫的 BIM Handbook[9]，其為建築資訊塑模領域的資訊手冊，該書的內容對建築資訊塑模有詳細的介紹，並針對在建物生命週期中的擔任不同工程角色時，所會面臨到的問題一一列舉與說明。而在該書中的 Glossary 章節有對於建築資訊塑模領域中相關的知識概念進行介紹，並對不同知識概念之間的相關性進行說明。

而由 IBC(Institute for BIM in Canada)組織於 Environmental Scan of BIM Tools and Standards[12]報告中統計目前已存在的 BIM 工具，並對每個 BIM 工具在建築資訊塑模中所應用的領域範圍進行統計，其整理出的分類架構如表 2，此分類架構即是對建築資訊塑模領域中的知識概念的整理。

表 2 BIM 工具使用分類[12]

Tool Category	Areas of Use
Planning and Design	Site Modeling
	Spatial Programming
	Design Authoring
	Design Review
	Engineering Analyses
	Code Evaluation
	Cost Estimation
Construction	Sequential Planning
	Construction Site Utilization
	Temporary Structure Design
	3D Coordination
	Site Analysis and Phase Planning
	Cost Estimation
Operation	Building Record
	Building Performance
	Space and Asset Management
	Disaster Planning
	Maintenance Scheduling
	Building Analysis

### 2-2-3 建立階層與關聯性

在知識本體中，需定義不同知識概念之間的關係，所以由上節的文獻回顧中，本研究選用 BIM Handbook 的 Glossary 章節，來做為基本知識本體的主要知識來源。因該章節的內容除了對於知識概念的說明以外，仍有針對不同知識概念之間的關係進行說明。

以下條例出本研究所選用的知識概念與其彼此之間的關係：

1. Building information modeling (BIM)

基本知識本體的領域範圍為 BIM，因此選用該知識概念作為基本知識本體的最上層結構。

2. BIM system

其內容為整合 BIM 工具並提供平台讓不同工具進行連接，所以此知識概念為 Building Information Modeling 的下層概念。

3. Building Model

其內容為可供 BIM 工具讀取與編輯的數位資料，所以此知識概念為 Building Information Modeling 的下層概念。

4. BIM Process

為使用 BIM 工具處理建物生命週期中所遇到的各種工作事項，所以此知識概念為 Building Information Modeling 的下層概念。

5. BIM application

其內容為工作團隊為某一種特定目的而對 Building Information Modeling 所採取的使用方式，此概念在 BIM system 中有提及並進行描述，所以此概念為 BIM system 之下層知識概念。

6. BIM tool

其內容為可產生並操作 Building Information Model 的軟體。其概念在 BIM system 中提及並進行描述，所以此知識概念為 BIM system 之下層知識概念。

7. Building objects

其定義為組成建築物的基本單位，如柱、梁、版、牆以及建築物的空間概念。所以此知識概念為 Building Model 下層之知識概念。



## 8. Building model repository

其定義為儲存 Building Information Model 的資料庫，該資料庫內容為物件式的儲存方式且具有檢索個別元件的功能。所以此知識概念為 Building Model 之下層知識概念。

## 9. Building Data Model

其定義為儲存 Building Information Model 的檔案格式，所以此知識概念為 Building Model 下層知識概念，用來說明模型的檔案格式。

綜合以上 9 個知識概念以及其相互之間的關係可以得到建築資訊塑模的知識本體，其內容如圖 3。共為三層架構，每一層之間的關係為「包含」，因此根據圖 3 所表示的基本知識本體，BIM 此知識概念「包含」了 BIM System、BIM Model 與 BIM Process 共三個子知識概念。

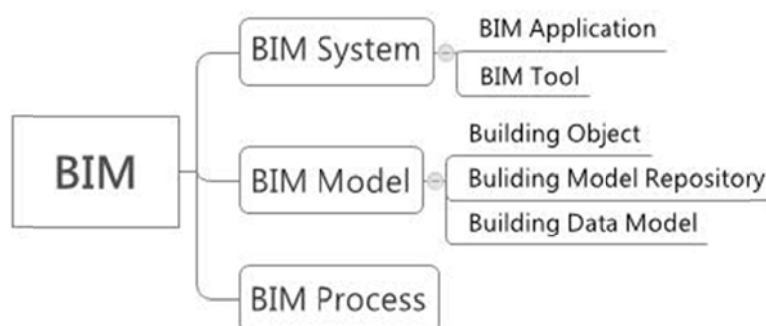


圖 3 BIM 基本知識本體

而在 Process 之下的知識概念，在 BIM Handbook 裡只有部分舉例，而沒有全部列舉。因此本研究參考了 Environmental Scan of BIM Tools and Standards 所提供的知識概念進行知識本體的擴展。

Environmental Scan of BIM Tools and Standards 此篇報告將 BIM Process 分為三個部分：Planning and Design、Construction、Operation，而每一個部分有包含有各種子應用方向。依造該篇報告整理之結果，將各項知識概念添加進目前 BIM Process 知識概念的下層。因此 BIM 領域的基本知識本體就擴展為圖 4。其總共有四層架構，共 27 個知識概念，其知識概念之間的關係為上層的知識概念包含有下層的知識概念。

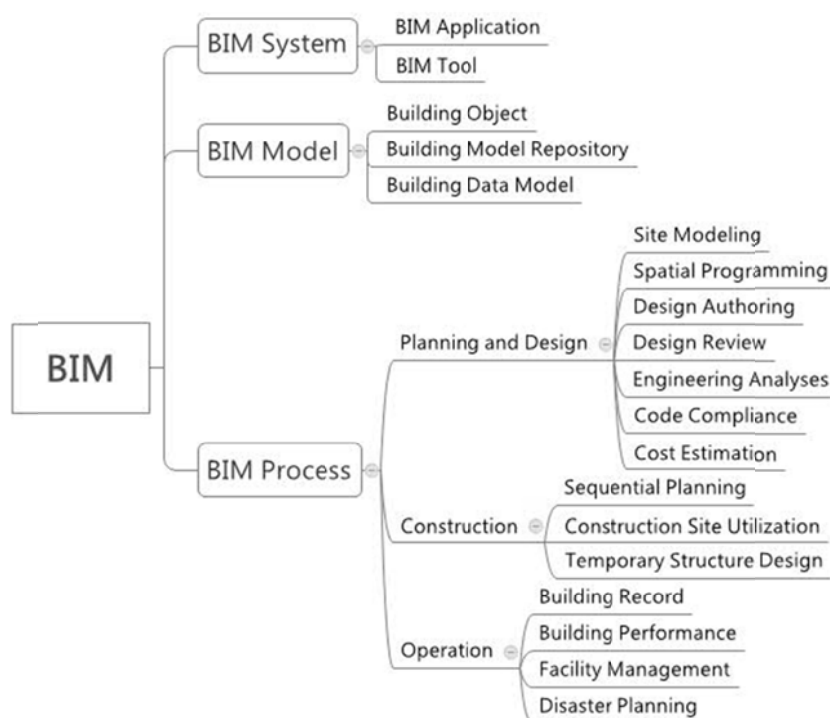


圖 4 擴展後的 BIM 基本知識本體

#### 2-2-4 知識本體展示

經過上兩節的整理之後，本研究已整理出建築資訊塑模領域的基本知識本體，但由於該基本知識本體的知識來源皆為英文的文獻，因此其知識概念的內容皆為英文。為將這些知識概念轉換成中文，需要對知識概念進行翻譯。

本研究首先採用 AEC-STIIRS 工具[13] 來進行翻譯，該工具提供了在營建領域中各式詞彙的中英文版本。其資料來源有兩個部分：CEDICT[14]與牛頓工程詞典[15]。CEDICT 為一般詞典，其中包含了 33000 組詞彙，牛頓工程詞典則為營建工程領域的專業詞典，共有 35000 組詞彙。

因此根據 AEC-STIIRS 工具可翻譯出部分知識概念，以下說明使用該工具在翻譯時的三種狀況：

只有單一解釋，如「tool」此知識概念使用 AEC-STIIRS 工具查詢後，其中文翻譯只有「工具」，而「工具」也符合此知識概念所表達之含意，因此選用「工具」做為此知識概念的中文版本。

含有多種解釋，如「model」此知識概念，在使用 AEC-STIIRS 工具查詢後，其中文翻譯有「模型」、「模範」、「雛型」三種不同的解釋，因此需根據此知識概念在知識本體中所表達的含意，來決定該知識概念的中文翻譯，而在此例選用「模型」做為此知識概念的中文版本。另一個相同的例子為「system」，其中文翻譯有「系統」、「制」、「系」、「綱」、「法」、「式」共六個，本研究選用系統作為此概念的中文翻譯。

未有直接翻譯，如「disaster planning」此知識概念在翻譯工具中並無直接的翻譯，但在 AEC-STIIRS 工具中有一英文詞彙為「disaster prevention plan」，該詞彙的中文翻譯為「防災計畫」，其含意與「disaster planning」相同，因此可以將「disaster planning」翻譯為「防災計畫」。

由於仍有部分知識概念在 AEC-STIIRS 工具中並沒有資料，因此部分知識概念參考前人的文獻進行翻譯。如在「以 BIM 與代理者技術實現營建協同設計審查之研究」[16]中，其作者所定義的中文關鍵字為：代理者、協同審查設計審查、建築資訊模型；英文的關鍵字為：Agent、Collaborative design review、BIM。由此可以知道 Collaborative design review 的中文解釋為協同審查設計審查，而扣除掉 Collaborative 後可以知道 design review 的中文翻譯為設計審查。

因此根據以上的步驟，就可將英文版本的基本知識本體(圖 4)翻譯為中文版本的基本知識本體，其成果如圖 5。

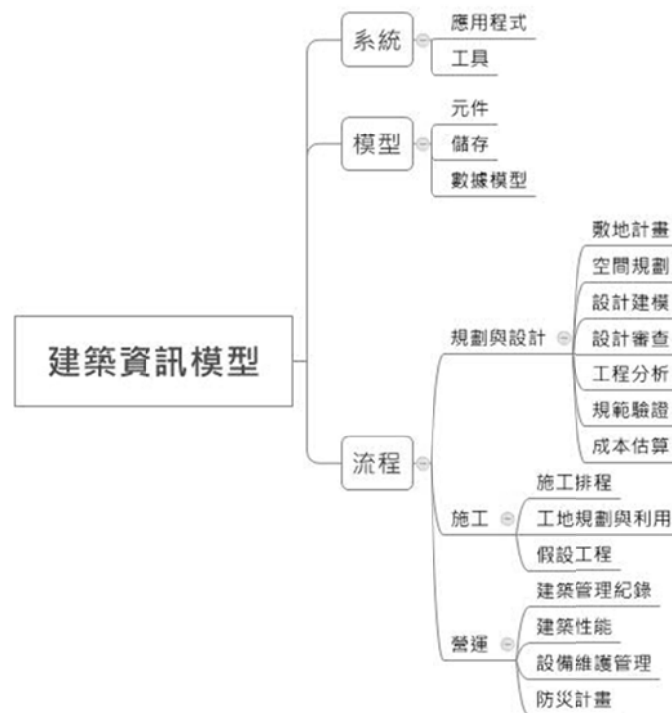


圖 5 中文版 BIM 基本知識本體

## 第三章 特定領域之資訊檢索

### 3-1 資訊檢索簡介

隨著資訊時代的發展，各式資料的電子化已經是進行知識管理的必要程序。但隨著資料數量的增加，若無妥善的整理方式，則儲存的資料裡所蘊含的知識就無法迅速的被找到，如此一來儲存的資料也無法進行有效的應用，因此資訊檢索技術就應運而生。

資訊檢索(Information Retrieval)為在儲存的資料中迅速找到所需要的相關資料的一種技術[17]。其最常見的形式就是網路上的各式搜尋引擎，如 GOOGLE、YAHOO、BING.....等。其運作方式為接收到使用者的資訊需求(Information Needing)後，根據不同的資訊檢索模型(Retrieval Model)進行資訊需求與事先已蒐集好的文件集合(Document Collection)的相關性計算，並依造相關性計算的結果，給予使用者經過排序過後的檢索結果。

參考文獻[18]中所提供的資訊檢索流程，本研究進行整理後將資訊檢索技術分為三個部分並繪製成圖 6，分別為圖 6 上方的使用者介面(User Interface)，圖 6 右下方的文件集合處理(Document Collection)，圖 6 左下方的資訊檢索模型(Retrieval Model)。以下分別說明各部分所進行的內容。

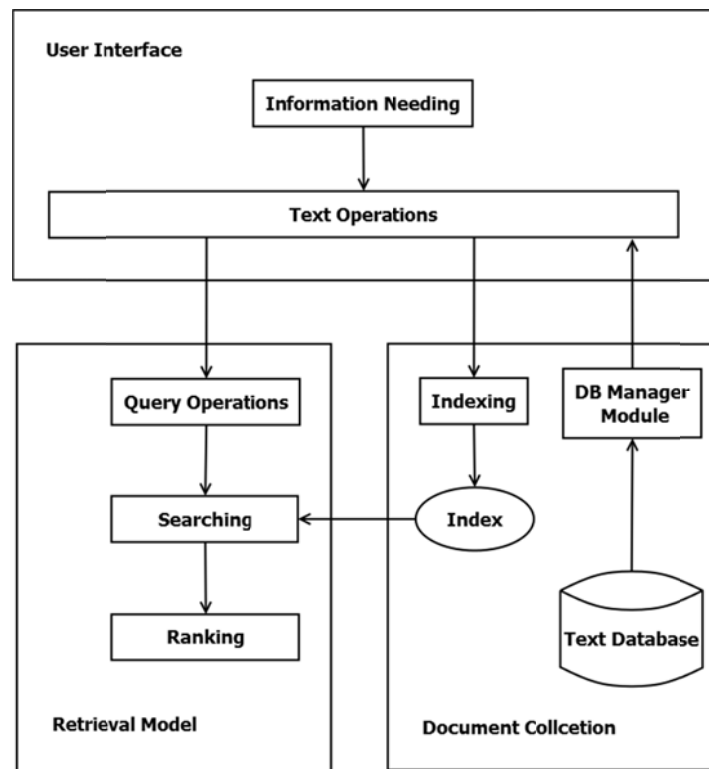


圖 6 資料檢索流程圖[18]

### 一.使用者介面 (User Interface)

處理使用者的資訊需求，延伸或擴展使用者的資訊需求。資訊需求在文字檢索中所代表的意義就是檢索詞彙(Query)，使用者在使用資訊檢索系統時，會利用簡單的詞彙來表達自己的資訊需求。同時資訊檢索系統也會記錄使用者在搜尋結果中實際點選的文件，來修正下一次資料搜索的成效表現。

### 二.文件集合處理 (Document Collection)

分析所儲存的文件集中所有文件包含的資訊。在文字檢索中其分析方式為將文件中的詞彙擷取出來，並將這些詞彙建置成索引，索引為表示該詞彙在那一篇文件中曾出現過的列表，其目的為增加資訊檢索系統在進行檢索時的處理速度，而索引將會搭配不同的資訊檢索模型來建置完整的資訊檢索系統。

### 三. 資訊檢索模型 (Retrieval Model)

將資訊需求與處理好的文件透過資訊檢索模型進行相關性計算，以找出較為符合資訊需求的文件，並依文件與資訊需求的相關程度，對檢索結果進行排序，讓使用者能決定從那一個結果開始看起。

綜合以上回顧，資訊檢索技術能幫助使用者快速而正確的得到其所需要的資訊需求，因此不論在何種專業領域內，都需要使用到資訊檢索技術。而讓使用者選擇何種資訊檢索技術的關鍵在於，該資訊檢索技術是否能檢索出使用者所需要相關資料，也就是說，資訊檢索的成效表現才是使用者最關心的議題。因此本章節將導入上一章的基本知識本體，來進行以知識本體為基礎的資訊檢索系統之建置，以改善資訊檢索系統運用於建築資訊塑模領域的成效表現。

#### 3-2 建立特定領域之測試文件集

資訊檢索的成效由測試文件集加以評估，測試文件集是具有「標準答案」的文件集合，而標準答案的獲得方式為請專家判斷有哪些文件與某一筆資料需求是確實相關的。因此，為評估資訊檢索系統之成效，本研究建置了建築資訊塑模領域的測試文件集。

測試文件集由三個部分組成：文件集合、資訊需求、文件相關度的「標準解答」。文件集合為最原始的文件資料，而為了後續的系統建置的方便，需將實際的文件資料轉化成電子文件。資訊需求則為使用者可能提出的檢索詞彙。文件相關度的標準解答則為文件集合中的每一項文件與某一筆資訊需求的相關程度。以下小節解釋各步驟詳細內容。

### 3-2-1 選擇與準備文件

在準備文件集合時，文件的資料來源有許多途徑。如：現有的書籍、期刊或是各式新聞。而不同的資料來源在文件獲取難度與文件內容的豐富度上各有不同。如期刊的文件內容較為完整但數量較少，而網路文件來源則數量眾多，但文件內容較為雜亂。

考慮到內容資料的正確與完整性，本研究選用了 100 年電子計算機於土木水利工程應用研討會的所有論文共 124 篇，扣除掉內容為英文的文章，實際使用 117 篇文件作為文件集合。圖 7 即為文件集合中的其中一篇文件。



圖 7 研討會文件掃描檔

而為了讓後續的資訊檢索系統無須針對不同的文件檔案格式進行讀取文件的設計。因此本研究採用文獻[13]中，建置 NCREE Collection 時所使用的文件格式，將每一份文件轉換為 XML 格式，而所使用的節點分別為：[DOCNO-文件編號]、[Title-文件標題]、[Author-作者]、[Abstract-文件摘要]、[Subject-關鍵字]、[Content-文件內容]、[Reference-參考文獻]，其中 Title、Author、Abstract、Subject，若有中英文兩種版本，則用 xml:lang 標籤分別註明。經轉換過的文章其內容如圖 8。



```

<DOC>
<DOCNO>124</DOCNO>
<Title lang="ZH">擬時間動力分析法應用於高度非均質地下水滲流問題分析研究</Title>
<Author lang="ZH">王嘉若、顧承宇</Author>
<Abstract lang="ZH">對於地下水滲流分析之非線性問題處理，以往多採牛頓法進行分析，本研究利
<Subject lang="ZH">地下水滲流、擬時間動力分析法、高度非均質、病態問題、非線性</Subject>
<Content lang="ZH">前言在求解地下水滲流問題中對非線性問題之處理，以往學者大多採用數值方法
<Title lang="EN">A study on highly heterogenous groundwater seepage problem usingFTIM-based
<Author lang="EN">Chia-jo Wang, Cheng-Yu Ku</Author>
<Abstract lang="EN">To deal with the nonlinear problems in the groundwater seepage analysis
<Subject lang="EN">Groundwater seepage, FTIM-based dynamical methods, highly heterogenous,i
<Reference>1. Cheng-Yu Ku, Weichung Yeh, Chein-ShanLiu (2010), Solving Non-Linear Algebraic
</DOC>

```

圖 8 測試文件集範例

根據以上所述，本研究建置了一小型文件集合，且因所採用之格式與 NCREE Collection 相符，因此該文件集合能增加 NCREE Collection 的內容，使得土木營建工程領域的測試文件集更加豐富。

### 3-2-2 定義資訊需求

在大型的文件集合中，定義資訊需求的方式通常有兩種方式：1. 由所收集的文件中相關的主題來直接定義資訊需求，2. 將文件集合先行檢索，並從裡面剔除出現次數過少的主題。但由於本研究所選定之領域，並無合適的大型文件集，因此經由上一節收集相關資料後，建置了一小型文件集合，而該文件集合的資訊需求則由上一章所建置的基本知識本體，選取其中的三個知識概念：「規劃與設計」、「施工」、「營運」來作為一開始設定的資訊需求，並在下一節對該資訊需求的合適性進行評估。

### 3-2-3 評估文件相關度

當產生資訊需求後，需評估在文件集合中與該資訊需求相關的文件，來做為該文件集合的標準答案，所以須請專家針對文件與資訊需求的相關性進行評估，而為確保不同專家對於同一個資訊需求所評估出來的結果是相同的，因此需對於專家所評估的結果進行一致性測試，採用的方法分別為 overlap value[19] 及 Cohen's Kappa measure[20]。

為說明兩種方法的計算方式，假設表 3 為針對某一筆資訊需求，兩專家評估每份文件之相關程度的結果。RR 為兩專家皆認為與該資訊需求相關的文件數量，RN 為專家 B 認為相關但專家 A 認為不相關的文件數量。NR 為專家 B 認為不相關的但專家 A 認為相關的文件數量。NN 為專家 A 與專家 B 皆認為不相關的文件數量。

表 3 專家評估結果

	專家 A	相關	不相關
專家 B			
相關		RR	RN
不相關		NR	NN

根據表 3 計算 overlap value 方法可表示為：

$$\text{overlap value} = \frac{RR}{RR + RN + NR} \quad (\text{公式 1})$$

其含義為兩專家評估為相關文件的交集除以兩專家評估為相關文件的聯集，其值越高則代表兩專家所評估的結果一致性越高。

Cohen's Kappa measure 則考慮到兩位專家所評估的一致性結果，可能是偶然為一致性，因此將扣除偶然一致性的部分，其計算方式如下：

$$\text{Kappa} = \frac{P(A) - P(E)}{1 - P(E)} \quad (\text{公式 2})$$

其中 P(A) 為兩專家評估結果確實為一致性的機率，其計算方式為：

$$P(A) = \frac{RR + NN}{\text{SUM}} \quad (\text{公式 3})$$

而 P(E)則是兩專家評估結果偶然為一致性的機率，其計算方式為：

$$P(E) = \frac{RR + NR}{SUM} \times \frac{RR + RN}{SUM} + \frac{NN + NR}{SUM} \times \frac{NN + RN}{SUM} \quad (\text{公式 4})$$

其中 SUM 所代表的意義為搜尋結果的總文件數量，即為 RR+RN+NR+NN。因此依照 Cohen's Kappa measure 的計算方式，即可獲得扣除偶然一致性的一致性評估結果。

本研究對文件集合進行三個資訊需求的一致性評估。每個資訊需求皆邀請了三位專家進行評估，而為了讓每一位專家的評估標準相同，因此要求專家在進行評估時，其評估標準為判斷該文件內容是否與該檢索詞彙所表達之生命週期階段的概念相關，如：「營運/Operation」此資訊需求，若該文件內容與「營運/Operation」相關且與「建築資訊模型/BIM」相關，則該文件與此資訊需求為相關。

將三位專家所評估的結果進行 overlap value、Cohen's Kappa measure 的計算，但由於這兩種計算方式是比較兩位專家之間的評估結果，因此每個檢索結果會有三筆數值，其結果如表 4 與表 5。

表 4 Cohen's Kappa measure 計算結果

資訊需求	I & II	II & III	I & III
規劃與設計	$1.009 \times 10^{-16}$	0.615	-0.222
施工	0.368	0.300	0.300
營運	0.565	0.327	0.722

表 5 overlap value 計算結果

資訊需求	I & II	II & III	I & III
規劃與設計	0.353	0.667	0.421
施工	0.625	0.500	0.563
營運	0.556	0.417	0.727

參考文獻[19]統計 TREC topic 202-250 的 overlap value 的數值，多落在 0.42 至 0.49 之間；文獻[21]則指出 Kappa 值在 0.67 以上即代表兩位專家所評估的結果較為一致。但由於本研究所建置的文件集合數量較少，因此對於此兩項數值所選定的門檻值調整為 Overlap 高於 0.5，Kappa 高於 0.6。在此門檻值之下，則有兩個資訊需求在兩位專家的評估下符合要求，分別是「規劃與設計」及「營運」，因此選取此兩筆資訊需求作為檢驗資訊檢索系統成效的指標。



### 3-3 運用知識本體於特定領域資訊檢索之成效評估

#### 3-3-1 向量空間模型簡介

本研究的資訊檢索系統採用向量空間模型(Vector space model)進行建置，向量空間模型為在資訊檢索技術中常見的檢索模型。其為將文件與資訊需求都轉換為向量的表示方式，而在向量空間中，當兩向量的夾角越小時，代表兩向量越為相似，因此利用轉換而成的向量可以求出文件與資訊需求之間的相關程度[22]。

轉換文件向量的方式為向量的每一個維度代表在文件中每一個詞彙的權重，其中詞彙的權重計算是以 tf-idf 方法[23]來進行計算，其計算方法如下：

$$tf = \frac{\text{目標詞彙在文件中出現的次數}}{\text{文件的總詞彙數量}} \quad (\text{公式 5})$$

$$idf = \log\left(\frac{\text{文件集中的文件總數}}{\text{有出現過目標詞彙的文件數量}}\right) \quad (\text{公式 6})$$

tf 為 term frequency 之縮寫，其中文為詞頻，其含義為當目標詞彙在某一文件出現越多次則 tf 值越高。idf 則為 inverse document frequency 的縮寫，其中文為逆向文件頻率，當目標詞彙在許多文件皆出現時 idf 值越小，兩者相乘後則為在文件向量中某一個維度的詞彙權重。計算完每一個詞彙的權重後，這些不同維度的權重就會組成該文件的向量，可表示成 $\vec{D}$ 。而當使用者提出資訊需求時，也對該資訊需求用同一方式建置該資訊需求之向量，可表示成 $\vec{Q}$ ，接著在利用向量內積來求其相關性，其式如下：

$$\cos \theta = \frac{\vec{D} \cdot \vec{Q}}{|\vec{D}| |\vec{Q}|} \quad (\text{公式 7})$$

向量空間模型的相關性計算為當兩向量所夾的角度越小，則該資訊需求與該文件的相關性就越高，因此可利用向量內積來進行計算，當向量內積的結果越大的時候，則相關性越高。因此根據向量空間模型可以將文件集中的文件依照與資訊需求的相關性進行排序，也就是使用者所需要的搜尋結果。

### 3-3-2 以知識本體為基礎之資訊檢索方法

本研究運用向量空間模型進行資訊檢索系統建置，其檢索流程如圖 9，將文件集合(Document Collection)以向量空間模型(Vector Space Model)進行分析後，就可對資訊需求(Information Needing)進行向量模型的檢索，以得到該資訊需求的檢索結果(Retrieval Result)。

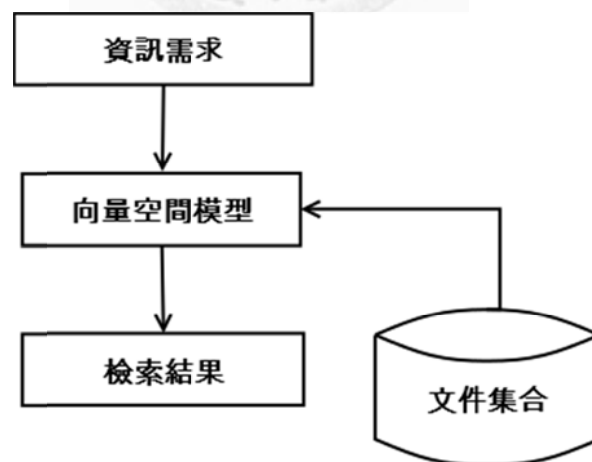


圖 9 向量空間模型資訊檢索系統

根據圖 9 向量空間模型的資訊檢索系統，當使用者輸入資訊需求後，向量空間模型的資訊檢索系統只會對使用者所輸入的資訊需求進行分析，而以知識本體為基礎的資訊檢索方法，其主要的內容是在資訊需求的分析上進行擴充。因為當擁有知識本體後，資訊檢索系統就可以針對使用者所輸入的資訊需求依照知識本體的內容去進行擴充，其流程如圖 10。

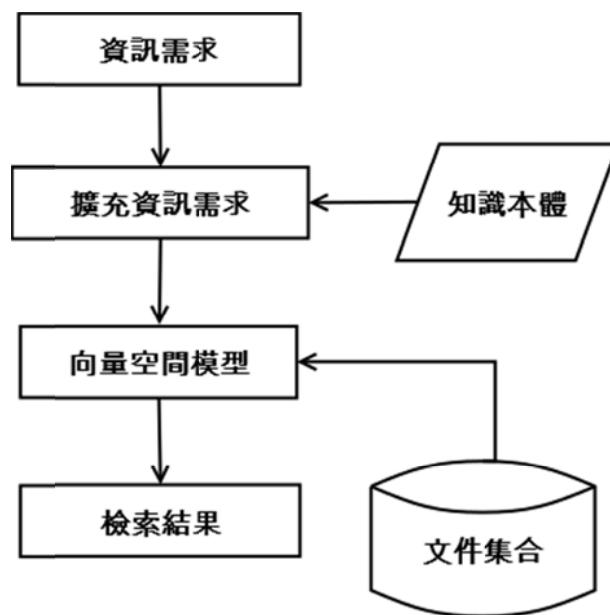


圖 10 知識本體資訊檢索系統

因此，在運用知識本體於資訊檢索上，本研究將知識本體視為一知識來源，藉由該知識本體對使用者的資訊需求進行擴充。例如當使用者輸入「BIM」、「營運」這兩個資訊需求時，知識本體資訊檢索系統就會認為該使用者是想要查找在 BIM 領域中與營運相關的文件，因此知識本體資訊檢索系統就會開始查找知識本體中與營運相關的內容進行資訊需求的擴充。其擴充的方式為將在知識本體中屬於營運下面層級的知識概念皆補充到資訊需求中，進而讓資訊需求能包含較多資訊。

而由於在知識本體中不同的層級的知識概念與營運的相關性不盡相同，因此根據知識本體的結構，每一個補充的知識概念皆有不同的權重，其權重分配方式為從使用者所輸入的資訊需求開始，往下層的概念開始分配權重。

為說明權重分配方式，假設目前的知識本體架構如圖 11，且此時使用者輸入的資訊需求為節點 E，則其權重分配方式為由節點 E 開始往下找尋沒有子節點的節點，依照此例，無子節點的節點為 P、Q、R、S、T、U。因子節點越多則代表該節點所蘊含的知識概念較多，則該節點的重要程度也較高，因此將這些沒有子節點的節點，其權重數值初始化為 1。接著將此節點的權重賦予此節點的父節點，如節點 M 為 S、T、U 的父節點，因此其權重數值為 3，依照相同計算方式則節點 K 的權重為 2，節點 L 的權重為 1，並再將 K、L、M 的權重賦予三節點的父節點，因此節點 E 的權重為 6，而其他所有節點的權重為 0。其結果如圖 12。

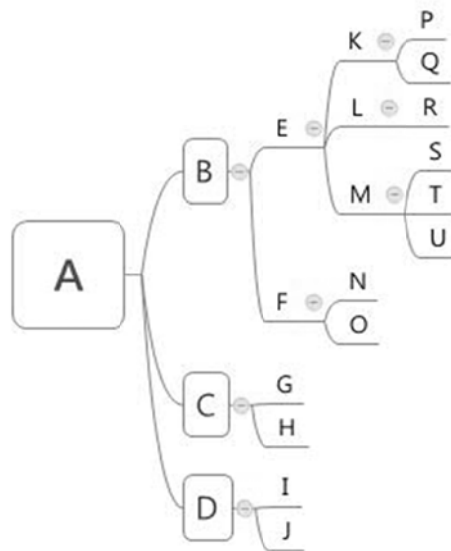


圖 11 概念分層架構



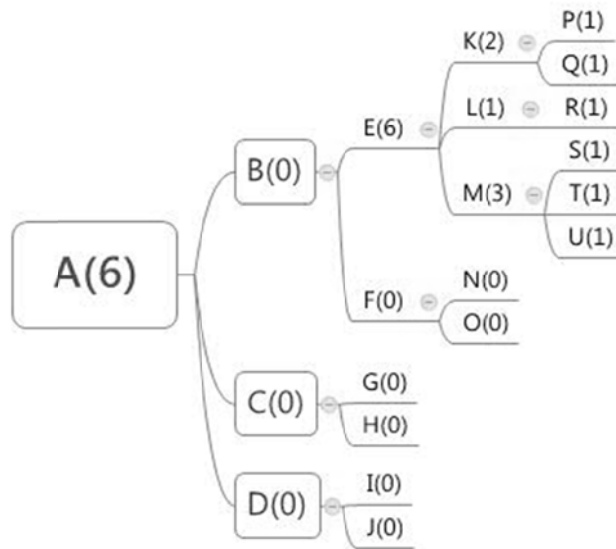


圖 12 權重地圖

得到節點 E 的權重地圖後，則需考慮使用者所輸入的另一資訊需求：BIM，由於 BIM 與節點 E 皆為使用者輸入的資訊需求，所以兩者的重要性一樣重要。因此將節點 E 的權重賦予給 BIM，也就是 BIM 的權重與節點 E 相等。而因為 BIM 為整個知識本體架構中最上層的知識概念，所以在圖 11 的例子中節點 A 即為 BIM，因此將節點 E 的權重賦予給節點 A，也就是節點 A 的權重為 6。而為了使每一筆資訊需求的重要性都相等，因此將權重地圖進行正規化，讓權重地圖上的所有權重加總起來為 1，以保持每一筆資訊需求的重要性相等。

根據以上步驟，使用者所輸入的資訊需求，就可經由知識本體的內容進行補充，而增補完的資訊需求，就可以轉換為向量空間模型中的資訊需求向量，並依照向量空間模型的相關性評估方式，得出該資訊需求的資訊檢索結果。

### 3-3-3 資料檢索成效評估

為評估向量空間模型檢索與以知識本體為基礎之資訊檢索方法於資訊檢索系統之成效，因此本研究針對兩種檢索系統所返回之檢索結果進行評估。所採取的評估標準分別是求全率與求準率，其計算方式如下：

#### 1. 求全率(Recall)

$$\text{Recall} = \frac{\text{檢索結果中相關文件數量}}{\text{文件集合內所有相關文件的總數}} \quad (\text{公式 8})$$

#### 2. 求準率(Precision)

$$\text{Precision} = \frac{\text{檢索結果中相關文件數量}}{\text{檢索結果的文件總數}} \quad (\text{公式 9})$$

在評估資訊檢索系統的成效上，這兩項是最主要的指標，求全率可以得知資訊檢索系統是否能將所有相關的文章檢出，而求準率可以知道資訊檢索系統的精確程度。因此為了不偏重於其中一項指標，本研究採用了平均求準率(AP/Average Precision)來做為評估資訊檢索系統的成效表現的指標，其計算方法如下：

$$\text{AP} = \frac{\sum_{i=1}^N (P(i) \times \text{rel}(i))}{R} \quad (\text{公式 10})$$

其中 R 為與該資訊需求相關的文件數量；P(i) 為計算檢索結果中到第 i 名文件的求準率；rel(i) 為第 i 名文件是否與資訊需求相關，若相關則為 1，不相關為 0；N 為搜尋結果之總數。

而根據資訊檢索領域的重要國際會議 TREC(Text Retrieval Conference)所提供的資訊檢索成效測試工具[24]，平均求準率的計算方式可簡化為當求全率的值為 0、0.1、0.2 至 1.0 時的求準率的平均，其方法為在求全率與求準率的曲線圖上以內差的方式，求取這十一個點的求準率數值。此種計算方式同時考慮了求全率與求準率的表現，因此本研究根據此指標來評估資訊檢索系統成效。

根據以上評估方式，對兩種檢索方法使用上一節通過一致性分析的資訊需求，並根據由專家所評估出來的標準答案，可得到使用向量檢索模型與以知識本體為基礎的檢索技術的資訊檢索系統成效，其平均求準率的數值如表 6。

表 6 兩種不同資訊檢索模型的平均求準率

資訊需求 \ 檢索模型	Vector space model	Ontology-based model
規劃與設計	0.528	0.554
營運	0.475	0.570

以表 6 的第一行為例，其所代表的意義為使用「規劃與設計」做為資訊需求在兩個不同的資訊檢索模型中所得的平均求準率，如第一行第一列的 0.528 即是使用向量空間模型進行資訊檢索後所得到的平均求準率。若其值越高，則代表使用該資訊檢索模型的資訊檢索成效表現越好。

因此根據表 6 的第一行的數值表現來看，使用知識本體進行資訊檢索的平均求準率高於使用向量空間模型的平均求準率，而在第二行的數值表現，使用知識本體進行資訊檢索的平均求準率也同樣高於使用向量空間模型的平均求準率。所以在此兩項資訊需求的檢索結果中，以知識本體進行資訊檢索確有改善資訊檢索系統之成效。

## 第四章 以資訊檢索技術增補知識本體

### 4-1 增補知識本體之流程

在第二章中，本研究建置了建築資訊塑模領域的基本知識本體，並在第三章中利用該基本知識本體建置了以知識本體為基礎的資訊檢索系統，同時對於該資訊檢索系統的成效與向量空間模型的資訊檢索系統的成效進行比較，其結論為以平均求準率(average precision)作為指標進行評估後，得出以知識本體為基礎的資訊檢索的成效較佳。本章將延續上兩章的成果，對以知識本體為基礎的資訊檢索系統的檢索結果進行分析，來對基本知識本體的內容進行增補。

本研究第三章所建置的資訊檢索系統可以得到每筆資訊需求的檢索結果。在資訊檢索中檢索結果是與資訊需求相關的文件的排名順序，若文件在檢索結果中的排名越高就代表該文件與該筆資訊需求越相關。而因為該資訊檢索系統是使用知識本體來對資訊需求的内容進行補充，因此在檢索結果中排名較高的文章，就是與知識本體中的知識概念相關性較高的文件。

為了對知識本體的内容進行增補，本研究針對在檢索結果中排名較高的文件進行分析。因為在同一篇文件中會包含許多知識概念，除了與資訊需求相關的知識概念外，該文件仍會含有其他的知識概念，而這些知識概念與資訊需求中的知識概念同時出現在同一篇文件中，則代表這兩個知識概念之間相關性程度較高，因此應將這些知識概念增補進知識本體中，所以藉由分析檢索結果中排名較高的文件中的知識概念，就可以對知識本體的内容進行增補。

根據以上概念將利用資訊檢索技術對知識本體進行增補的流程整理成圖 13，其主體流程延續第三章以知識本體為基礎的資訊檢索流程(圖 10)，其增加的部分為：「擷取知識概念」、「知識概念是否存在於知識本體中?」、「增補知識本體」、「領域專家協助修訂」、「領域知識本體」五個部分，以下各小節分別說明各部分之內容。

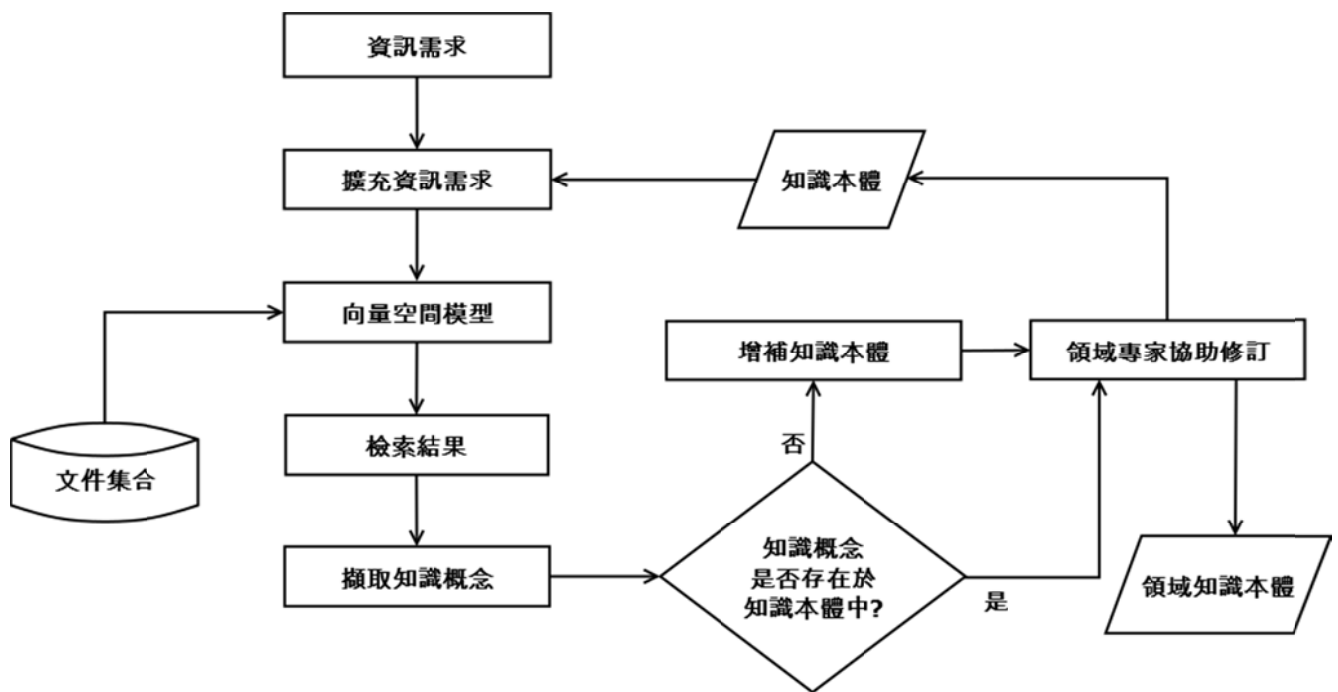


圖 13 運用資訊檢索技術增補知識本體之流程

#### 4-1-1 擷取知識概念

得到某一筆資訊需求的檢索結果後，對該檢索結果中排名較高的文件進行知識概念的擷取。而在中文的文件中，代表該篇文件所包含的知識概念則為該文件中的所出現的詞彙，因此為了擷取出每篇文件的知識概念，就須對文件中的語句進行分析。所謂的語句的分析就是如何將語句拆分成詞彙，在資訊檢索中此項技術稱之為「斷詞」。

斷詞技術為讓文件內容裡的語句依照正確的語意，以詞彙為單位來進行儲存。但是如何將語句以正確的方式進行斷詞，在不同語言中就有著不同的方式，因為每種語言在詞彙組合成語句的規則上並不相同，例如，英文的詞彙是使用空白做為間隔，而中文的詞彙則是以連續無間隔的方式來組成語句。

斷詞技術在中文上尤其困難，因為中文的詞彙在語句中並沒有明顯的間隔，所以中文斷詞方法多是參考中文辭典，先行在語句中擷取出在辭典中出現的詞彙，再分析語句中所剩下的內容，來得到斷詞結果。此種方式的優點為在文件內容與該辭典所包含的領域相同時，斷詞結果會較為準確，但若文件內容是屬於新發展或較為專業的領域時，則會因為辭典中沒有與該領域的相關資料，而容易發生斷詞錯誤。

由於本研究所選定的特定領域為建築資訊塑模領域，其為新發展的知識領域，因此本研究採用 n-gram 斷詞方法[25]對文章進行分析，其優點為無需先建置辭典，而是直接針對文章的內容進行分析，來擷取出文件中所出現的詞彙。因此 n-gram 方法在擷取新發展的知識領域的詞彙時，就不會因新發展的知識領域的詞彙尚未被編進辭典中，而無法對文件的內容進行正確的斷詞。

n-gram 方法為假設每個語句都存在有一個最小的表示單位，在中文裡最小的表示單位為「字」，每個最小表示單位可以跟另一個最小的表示單位合成為一個概念，也就是中文的「詞」。n-gram 方法並不限定多少個最小表示單位來進行組合成一個概念。若只用一個最小表示單位稱之為 unigram，以兩個單位的長度進行連接稱 bigram，以三個單位的長度進行連結稱為 trigram。

在實際進行斷詞時需要多少個最小表示單位組合成一個概念，則是由語言特性與文件內容來決定的。以下以「工程變更設計」此語句為例，表示 bigram 與 trigram 的切分結果。

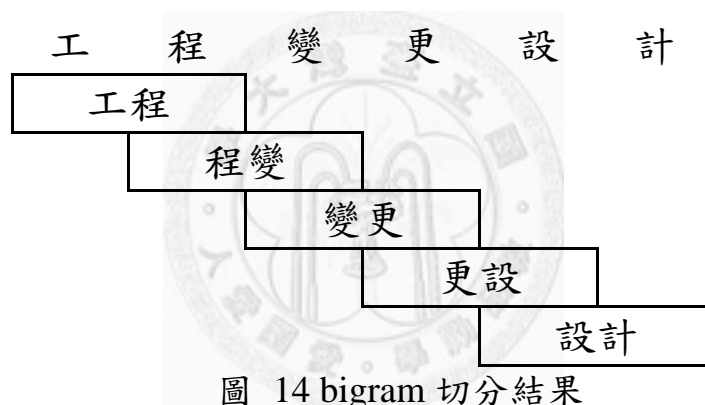


圖 14 bigram 切分結果

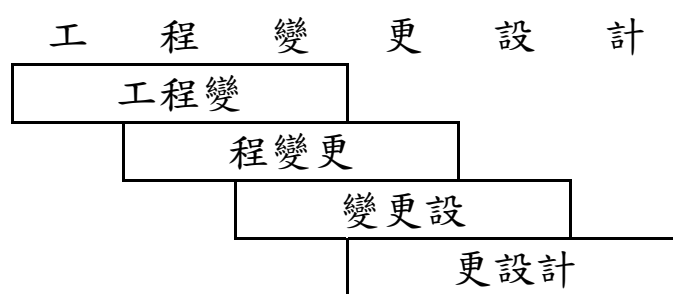


圖 15 trigram 切分結果

對語句進行以 n-gram 方法的不同長度切分後，為得到這些詞彙的重要程度，因此對這些詞彙進行詞頻計算，詞頻為計算每一個詞彙在文件中所出現的頻率，其計算方式為該詞彙在文件中所出現的次數除以整篇文件中的總詞彙數量：

$$\text{詞頻} = \frac{\text{目標詞彙在文件中出現的次數}}{\text{文件的總詞彙數量}} \quad (\text{公式 11})$$

其中文件的總詞彙數量的計算方法，則是根據 n-gram 方法的不同切分長度進行計算。每個詞彙在計算詞頻時的分母為該詞彙的切分長度的總字彙數量。例如依照「工程變更設計」此語句，「工程」此詞彙在計算詞頻的時候的分母為 5，因為該語句在兩個單位的長度下被切分為 5 個詞彙；「工程變」此詞彙的分母則為 4，因為在三個單位的長度下，該語句被切分為 4 個詞彙。因此切分長度較長的詞彙，其分母就會較小，詞頻的數值就會較高，來凸顯長詞彙的重要性。

而由於在中文文件中，仍然會有許多專業詞彙以英文方式表示，因此在進行實作時會將這些英文詞彙獨立抽出，再對剩下的中文詞彙進行 n-gram 方法分析，而英文詞彙的權重也是依照詞頻計算公式進行計算，所以在本研究中同時也會分析在中文文件的英文詞彙的重要程度。

利用 n-gram 方法與詞頻計算方法，就可以對文件集中所有文件的內容進行分析，來得到在每篇文件中詞頻表現較高的詞彙。其結果如



表 7，該表為在文件集中某五篇文件進行分析後，在該篇文件中詞頻前十名的詞彙。



表 7 文件關鍵字排名表

排名	文件一	文件二	文件三	文件四	文件五
1	研究室	檢查	BIM	BIM	RFID
2	節能	檢查維護	Revit	專案經理人	讀取
3	誘導	消防安全設備	產出	執行流程	構件
4	用電	消防	施工圖	專案	BIM
5	誘導模式	檢查維護人員	圖面	建國工程	標籤
6	生態系	消防設備	日照	團隊協作	鋼構
7	虛擬生態系	BIM	日照輻射量	團隊	Tag
8	虛擬	設備	疑義	員工	驗收
9	虛擬對象	維護	日照輻射量分析	流程	鋼結構
10	回饋	管理權人	建築	營造公司	讀取器

從



表 7 中可以看出，每篇文章的前十名關鍵字就代表了這篇文章中所蘊含的重要的知識概念，所以計算出每篇文件中關鍵字的排名順序後，就可以對每個資訊需求的檢索結果進行關鍵字排序，再依據關鍵字排名表中的關鍵字來對知識本體的內容進行增補。



#### 4-1-2 自動化增補知識本體

此小節包含知識本體增補流程的兩個部分：「知識概念是否存在於知識本體」與「增補知識本體」。此兩部分為針對每一筆資訊需求的檢索結果，選取在檢索結果中排名較高的文件的關鍵字後，對這些關鍵字進行排序，接著判斷排名較高的關鍵字是否已在原本的知識本體中，若關鍵字不存在於目前的知識本體中，則增加該關鍵字於知識本體中。

由於增補知識本體的流程為等待終止條件的循環，因此為了能自動化進行知識本體的增補，本研究建置了自動增補知識本體的系統。其系統的運行流程如圖 16，分為「在知識本體中同階層的知識概念」、「找尋知識概念的候選增補知識概念」、「檢查候選增補知識概念是否有重複」、「候選增補知識概念添加進知識本體」、「檢查候選增補知識概念在父知識概念的排名順序」以下分別說明各部分之內容。

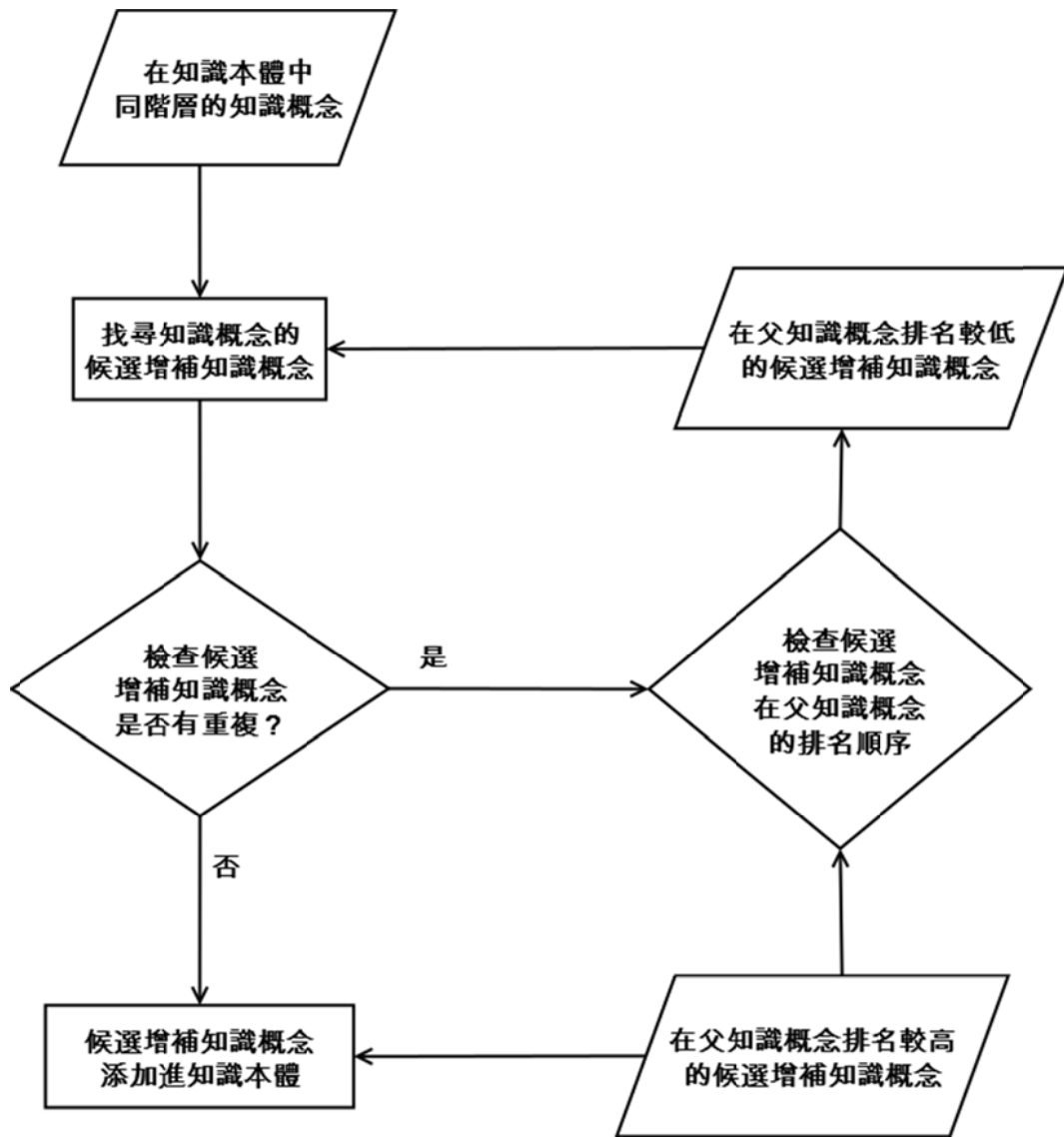


圖 16 自動化增補知識本體之流程

- 在知識本體中同階層的知識概念

自動化增補知識本體之流程的初始步驟。在本研究中由知識本體中最下層的知識概念為起始，進行增補知識本體之流程，結束該層的增補流程後，會再選取在知識本體中上一層的知識概念進行增補，直到整個知識本體中的知識概念都已經進行增補為止。

以圖 17 為例，在知識本體架構中，最下層的概念為第四層的敷地計畫、施工排程、防災計畫……等，因此增補知識本體的流程將從此階層開始往上進行增補流程，在第四層的知識概念增補結束後，依序為第三層的應用程式、元件、規劃與設計、施工、營運……等，第二層的系統、模型、流程，再到最上層的建築資訊模型。

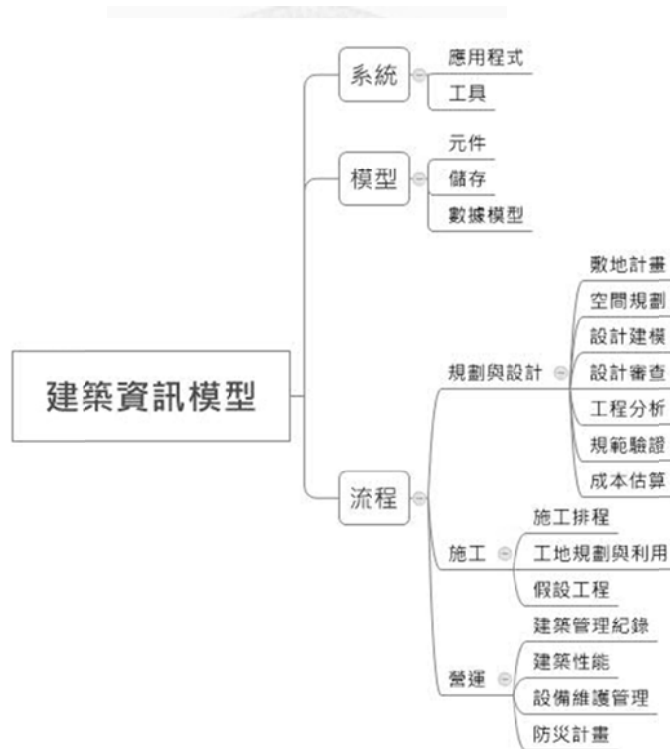


圖 17 中文版 BIM 基本知識本體

- 找尋知識概念的候選增補知識概念

此步驟為找尋所有在知識本體中為同一階層的知識概念的候選增補知識概念。其方式為在以該知識概念作為資訊需求的關鍵字排名表中，找尋候選增補知識概念。選擇候選增補知識概念的邏輯為由排名最高的關鍵字開始選取，若該關鍵字已出現在知識本體中，則選取下一位關鍵字，若前 K 名的關鍵字皆已選取，則該知識概念無候選增補知識概念。

以表 8 為例，以「敷地計畫」作為資訊需求進行檢索後，可得到該資訊需求的關鍵字排名表，而在該排名表中第一名為「BIM」，但因此知識概念已在知識本體中，因此選取第二名的「植栽」作為候選增補知識概念。

表 8 知識概念關鍵字排名表

知識概念 排名	敷地計畫	工程分析	工地規劃與利用	規畫與設計	施工
1	BIM	BIM	BIM	BIM	BIM
2	植栽	進度	進度	團隊	介面溝通
3	銷售	專業分包商	專業分包商	領域團隊	專業分包商
4	建築	甘特圖	甘特圖	專業分包商	施工介面
5	購屋者	模型	模型	跨領域	WBBIMS
6	景觀	WBBIMS	WBBIMS	WBBIMS	更新
7	視覺運算	更新	更新	更新	更新資
8	彩現	更新資	更新資	3D	施工介面溝通
9	現時間	工程進度	工程進度	階段成本占設計總	構件
10	Revit	更新資訊	更新資訊	段成本占設計總成	介面管理

- 檢查候選增補知識概念是否有重複 & 候選增補知識概念添加進知識本體

這一步驟為檢驗在同一層的知識概念的候選增補知識概念是否有重複，若候選增補知識概念沒有重複則直接將該知識概念添加進知識本體中。若有重複的候選增補知識概念則需檢驗該候選增補知識概念在該知識概念的父知識概念的關鍵字排名表的排名順序。

以表 8 為例，「工程分析」與「工地規劃與利用」其關鍵字排名相同，而排名第一名的「BIM」已在知識本體中，因此需選用第二順位的「進度」，但「進度」是另一知識概念的第一名的關鍵字，因此無法選用該知識概念作為候選增補知識概念，所以選取第三名的「專業分包商」作為候選增補知識概念，但因「專業分包商」同時為「工程分析」與「工地規劃與利用」的候選增補知識概念，因此需進行下一步驟的檢驗，來確定「專業分包商」為哪一個知識概念的候選增補知識概念。

- 檢查候選增補知識概念在父知識概念的排名順序

若某候選增補知識概念與該知識概念的相關性較高，則在該知識概念的父知識概念的關鍵字排名表中，該候選增補知識概念的排名也會較高。因此若兩個知識概念的候選增補知識概念相同時，則需比較該候選增補知識概念在兩個知識概念的父知識概念的關鍵字排名表的排名順序。將候選增補知識概念增補進在父知識概念的關鍵字排名表排名較高的知識概念之下，而排名較低的知識概念則須再次找尋合適的候選增補知識概念。

根據上一步驟所使用的例子，「工程分析」與「工地規劃與利用」就須判斷兩知識概念的候選增補知識概念在父知識概念的排名順序，而此兩知識概念的父知識概念分別為「規劃與設計」與「施工」，而父知識概念的關鍵字排名表則如表 8，在「規劃與設計」的排名表中「專業分包商」為第四名，在「施工」的排名表中「專業分包商」為第三名，因此「專業分包商」為「工地規劃與利用」的候選增補知識概念，而「工程分析」則需再次尋找合適的候選增補知識概念。



根據圖 16 所表示的自動化增補知識本體之流程，可以說明在上一節中的圖 13 的流程裡「知識概念是否存在於知識本體中？」的部分，利用自動化方式進行增補的流程，其方式為從知識本體最下層的知識概念開始，進行候選增補知識概念的檢驗，並逐步往上層知識概念進行增補知識概念的檢驗，直到所有知識概念都檢驗結束為止。

#### 4-1-3 領域專家協助修訂

此步驟為請領域專家對知識本體的內容進行評估與修訂，總共分為兩個部份：

第一個部份為請領域專家對於每一次增補進知識本體中的知識概念進行修訂，以確定該知識概念的表達方式與在知識本體架構中的位置是否合適，如將相近的知識概念進行整合、將知識概念的縮寫與簡稱進行整合，並對知識概念在知識本體中的位置進行調整。

第二個部分則是請領域專家評估該知識本體的完善性，是否已足夠代表該知識領域，以確定是否繼續進行增補。若需繼續進行增補，則再次利用增補後的知識本體進行增補流程。若無需進行增補，則產出領域知識本體。

#### 4-1-4 領域知識本體

在完整的增補流程中，將由領域專家對增補的知識本體的內容進行評估，來確定是否繼續增補知識本體，若不再進行增補後，則獲得最後的領域知識本體。然而在本研究中，將以自動化方式來取代專家評估，其原因為可快速獲得增補後的知識本體，並可立即分析知識本體對於資訊檢索的影響。且若使用自動化方式能取得較佳的成效後，則請專家協助修訂亦能獲得較佳的成效，因此為快速驗證此增補流程之成效，本研究以自動化方式進行增補知識本體，而自動化的終止條

件為當排名較高的關鍵字皆已出現在知識本體後，則停止整體增補流程，產出領域知識本體。

#### 4-1-5 系統建置與限制

根據以上增補知識本體的流程，本研究建置知識本體增補系統，該系統在使用以知識本體為基礎的資訊檢索系統後，會選取在每筆資訊需求的檢索結果前 N 篇的文件的關鍵字排名表，將前 N 篇文件的關鍵字進行排名，得到該資訊需求的關鍵字排名表。接著檢視排名表的前 K 個關鍵字是否已出現在知識本體中，若沒有該關鍵字則增加該關鍵字到知識本體當中，並再次進行資訊檢索。若關鍵字排名表的前 K 個關鍵字皆已出現在知識本體，則結束增補知識概念之流程，產出領域知識本體。

同時為了快速驗證該知識本體增補流程之可行性，該系統在知識本體的架構上，將知識本體簡化為樹狀結構，且所有知識概念之間的關係皆簡化為一種關係：「包含」，因此在樹狀結構的上層概念包含了所有的下層概念，每一個知識概念只會有父節點，不會有多重的連接關係。而在文件集合的選擇上也以中文文件來做為知識本體增補系統建置的資料來源。

這幾項簡化之條件皆可以使增補過來的知識本體快速產出，以驗證此知識本體增補流程之成效。由於完整的知識本體增補流程未限制各步驟的詳細實作內容，因此針對知識本體的架構與資料來源的處理方式可再進一步地進行不同的實作方式，來讓增補過後的知識本體的內容更加完善。

## 4-2 自動化增補知識本體之成效評估

### 4-2-1 自動化增補知識本體流程之參數設定

在進行自動化增補知識本體的流程中，有兩項參數需進行設定，分別是選取前 N 篇文章的 N 值與前 K 名關鍵字的 K 值，為了得到何項參數能表現出較佳的資訊檢索成效，因此對不同的參數設定進行實驗。使用的資訊需求分別是「規劃與設計」、「營運」，其結果如表 9 與表 10。表中的數字分別代表在使用不同參數設定下所增補出來的知識本體，在進行資訊檢索之後的平均求準率(average precision)。

以表 9 第一行為例，其所代表的意義為在搜尋「規劃與設計」的結果中，選取前三篇(N=3)文件的關鍵字進行關鍵字排序，並只對在關鍵字排名表的前三名(K=3)關鍵字進行是否增補進知識本體的檢驗。而每一列的數值為使用經過增補後的知識本體進行該資訊需求的檢索後的平均求準率。如第一行第一列的 0.518 為在 N=3、K=3 的參數下，使用只進行一次增補後所產生的知識本體，然後對「規劃與設計」進行資訊檢索後的平均求準率。若其值越高，則代表該資訊需求下的資訊檢索成效表現越好。第一行第六列的 X 則代表在該參數設定下，其增補知識本體的流程已經終止，因此無該資訊需求的平均求準率的數值。

表 9 「規劃與設計」的平均求準率

迭代 次數	N=3			N=4			N=5		
	K=3	K=4	K=5	K=3	K=4	K=5	K=3	K=4	K=5
1	0.518	0.518	0.518	0.518	0.518	0.518	0.545	0.545	0.566
2	0.530	0.533	0.533	0.533	0.533	0.524	0.546	0.534	0.534
3	0.530	0.595	0.559	0.533	0.540	0.535	0.543	0.540	0.459
4	0.533	0.595	0.595	0.533	0.576	0.549	0.543	X	0.464
5	0.533	X	0.595	X	X	0.535	0.533	X	0.464
6	X	X	0.595	X	X	0.535	X	X	0.487
7	X	X	0.595	X	X	0.535	X	X	0.487
8	X	X	X	X	X	0.535	X	X	0.487
9	X	X	X	X	X	0.535	X	X	X

表 10 「營運」的平均求準率

迭代 次數	N=3			N=4			N=5		
	K=3	K=4	K=5	K=3	K=4	K=5	K=3	K=4	K=5
1	0.570	0.570	0.566	0.570	0.570	0.570	0.566	0.566	0.566
2	0.570	0.570	0.563	0.570	0.570	0.570	0.566	0.566	0.579
3	0.570	0.570	0.563	0.570	0.570	0.570	0.566	0.566	0.579
4	0.570	0.570	0.563	0.570	0.570	0.570	0.566	X	0.579
5	0.570	X	0.563	X	X	0.570	0.566	X	0.579
6	X	X	0.563	X	X	0.570	X	X	0.579
7	X	X	0.563	X	X	0.570	X	X	0.579
8	X	X	X	X	X	0.570	X	X	0.579
9	X	X	X	X	X	0.570	X	X	X

由表 9 與表 10 可以了解不同參數對資訊檢索成效的影響，其中表 10 的數值變動幅度並不大，使用進行一次增補後的知識本體與進行多次增補後的知識本體，其平均求準率並無變化。因此為比較不同參數對於資訊檢索系統的成效所造成的影響，將依據表 9 的數值表現，也就是「規劃與設計」的資訊需求的平均求準率進行進一步探討。

為確定不同 N 值對於資訊檢索的成效影響，首先固定 K 值進行繪圖，其結果如圖 18、圖 19、圖 20。在圖中的 Base 數值為未使用知識本體進行資訊檢索的平均求準率，Ontology 為只使用基本知識本體進行資訊檢索的平均求準率，這兩者的數值與成效比較已於第三章探討過。而在圖中每一個資料點則為使用不同 N 值後，使用不同次數增補後的知識本體進行資訊檢索的平均求準率。

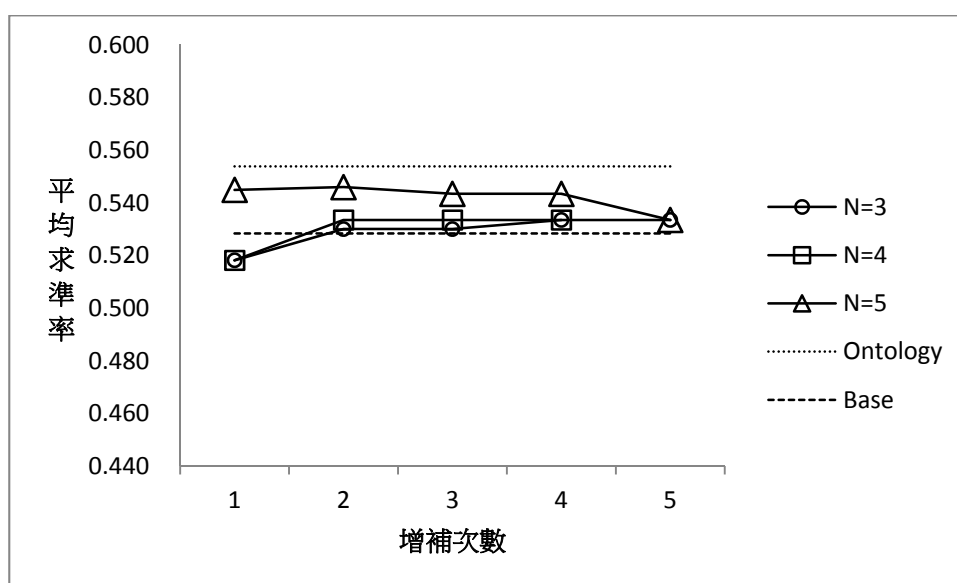


圖 18 K=3 時，不同 N 值的平均求準率

圖 18 為固定 K=3 時，不同 N 值的平均求準率，從圖上來看可發現三種不同的 N 值的平均求準率皆高於未使用知識本體的平均求準率，但也都低於只使用基本知識本體的平均求準率。但使用 N=3 與 N=4 的平均求準率會隨著增補的次數不斷的增加，N=5 則會逐漸下滑。

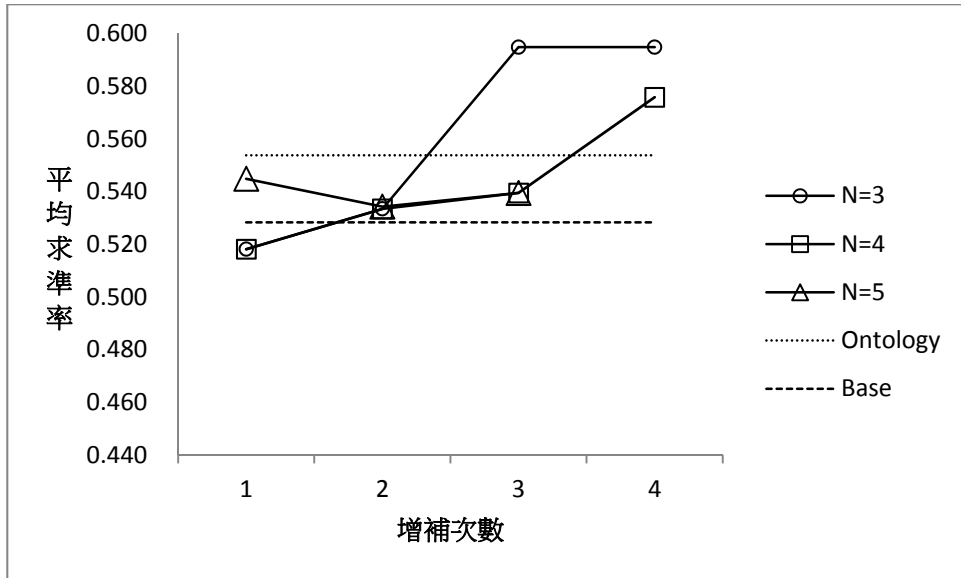


圖 19 K=4 時，不同 N 值的平均求準率

圖 19 為固定 K=4 時，不同 N 值的平均求準率，從圖上來看可發現三種 N 值的平均求準率皆高於未使用知識本體的平均求準率，而其中只有 N=5 經過三次增補後仍然低於只使用基本知識本體的平均求準率。在 N=3 與 N=4 的平均求準率則是隨著增補的次數的增加，平均求準率開始上升，且 N=3 的平均求準率會高於 N=4。

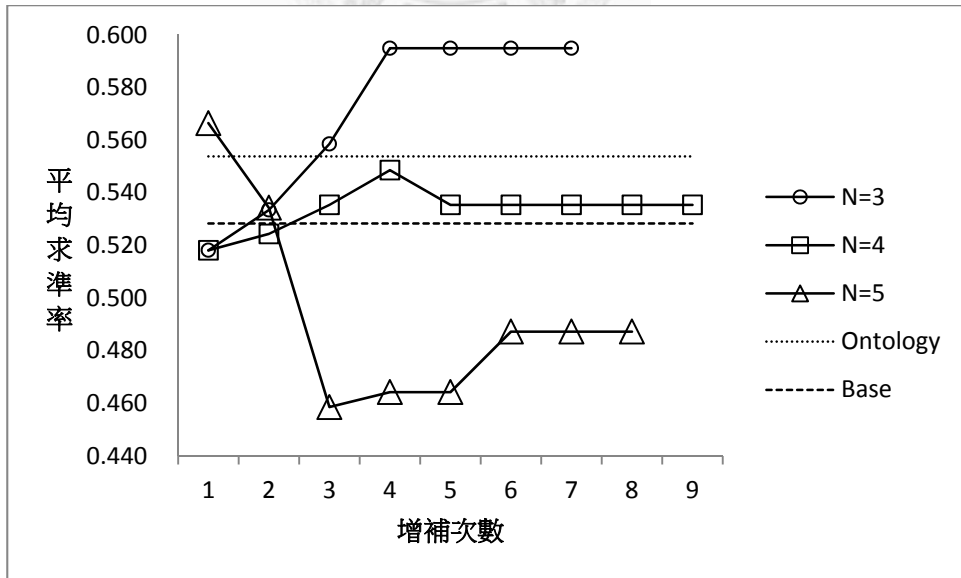


圖 20 K=5 時，不同 N 值的平均求準率

圖 20 為固定  $K=5$  時，不同  $N$  值的平均求準率。從圖上可發現三種  $N$  值的平均求準率有很大的差異，其中  $N=3$  會隨著增補次數的增加，平均求準率有明顯的上升且會大於只使用基本知識本體的平均求準率。 $N=4$  則是在一開始略微上升後，下降成平緩的表現，並每一次增補的結果都低於只使用基本知識本體的平均求準率。 $N=5$  則是隨著迭代次數，平均求準率不斷的下降，並低於未使用知識本體的平均求準率。

由圖 18、圖 19、圖 20 的比較可以發現，在三種不同  $K$  值的參數下， $N=3$  可以有比較穩定且較高的表現， $N=5$  的成效表現會有比較大的差異，且成效不佳。因此在選用  $N$  值的選擇上，選用  $N=3$  作為最後進行增補知識本體的參數設定。

而為了解  $K$  值對於資訊檢索系統的成效影響，以下固定  $N$  值進行繪圖，其結果如圖 21、圖 22、圖 23。

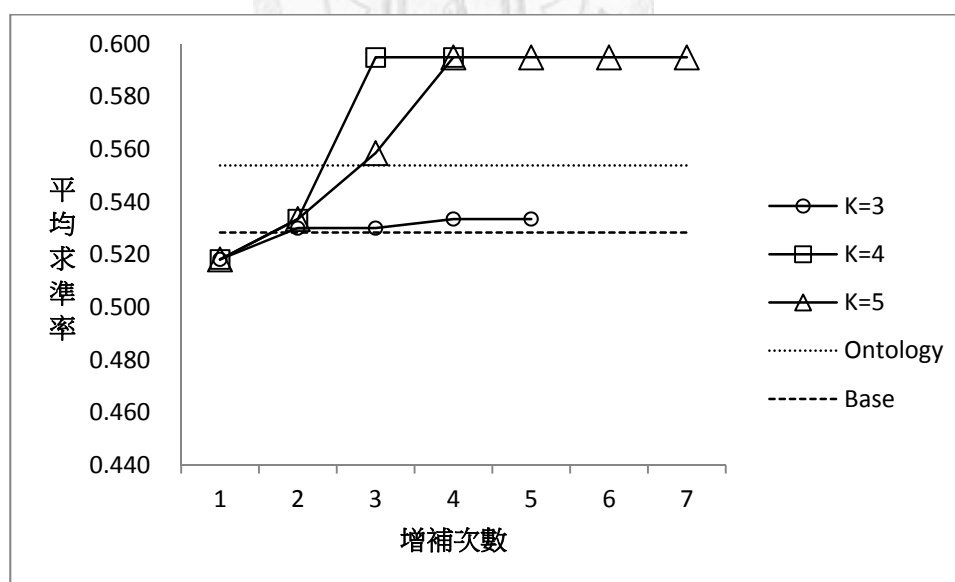


圖 21  $N=3$  時，不同  $K$  值的平均求準率

圖 21 為固定  $N=3$  時，不同  $K$  值的平均求準率。從圖面上來看三種  $K$  值的平均求準率都會隨著增補的次數的增加而提高，而  $K=4$  及  $5$  的數值會在迭代後超越只使用基本知識本體的平均求準率，且  $K=4$  上升的速度較快。

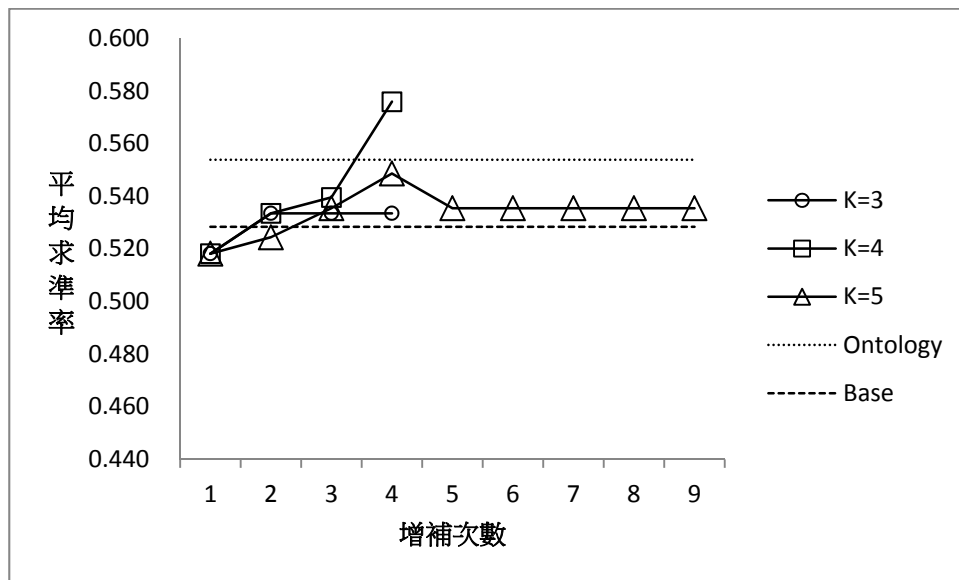


圖 22  $N=4$  時，不同  $K$  值的平均求準率

圖 22 為固定  $N=4$  時，不同  $K$  值的平均求準率。從圖面上來看，可明顯發現只有在  $K=4$  的時候，平均求準率會超過只使用基本知識本體的平均求準率，在  $K=3$  及  $5$  雖然會隨著增補的次數提高平均求準率，但其值皆不會超過只使用基本知識本體的平均求準率。



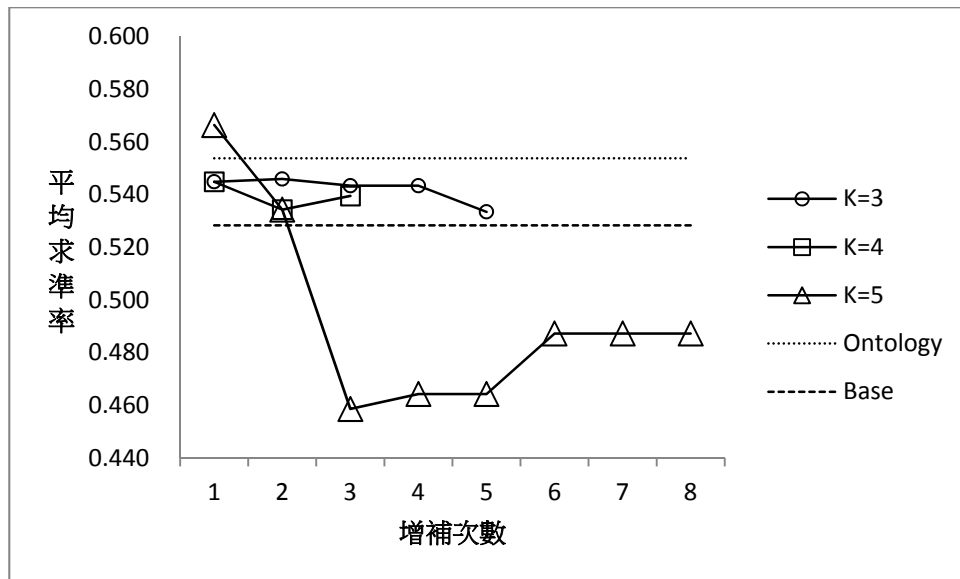


圖 23 N=5 時，不同 K 值的平均求準率

圖 23 為固定 N=5 時，不同 K 值的平均求準率。從圖面上來看可以發現不管使用何項 K 值，平均求準率的表現皆低於只使用基本知識本體的平均求準率，且會隨著增補次數的增加平均求準率開始下降。而 K=5 則會隨著迭代次數大幅下降，且比未使用知識本體的平均求準率還低。

總和圖 21、圖 22、圖 23，可以發現使用 K=4 時，在不同 N 值的表現下，其資訊檢索的成效表現較為穩定，並能高於只使用基本知識本體的平均求準率。因此本研究最後選用 N=3, K=4 做為最後進行自動化增補知識本體的參數值，使用該參數進行增補的知識本體其進行資訊檢索的平均求準率的表現繪製如圖 24。

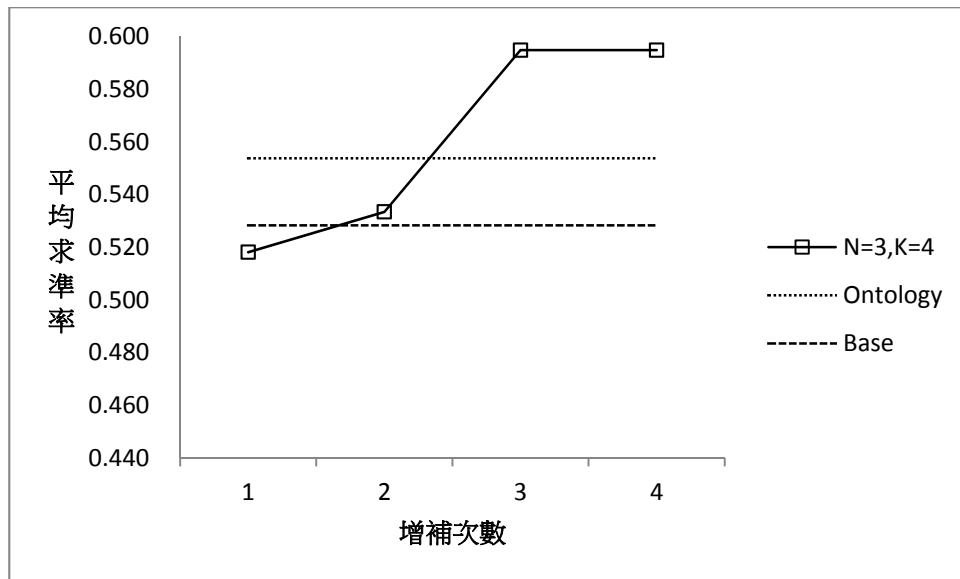


圖 24 N=3、K=4 的平均求準率

由圖 24 可看出在 N=3、K=4 的參數設定下，使用增補過後的知識本體的資訊檢索其平均求準率會隨著增補的次數慢慢提升，且超過只使用基本知識本體的平均求準率。

#### 4-2-2 自動化增補知識本體展示

經過上一節的實驗後，在自動增補知識本體的系統中，選用 N=3，K=4 為參數，進行自動增補知識本體的建置。依照圖 13 增補知識本體的流程，總共經過了四次增補的流程。以下展示每次增補後的知識本體與增加的知識概念。

● 第一次增補成果

與基本知識本體的架構相比，增補過後的知識本體多增加了一層，變為五層架構，增加了 17 個知識概念，其增補後的知識本體架構如圖 25。在此次迭代中增加的知識概念以粗體字標示，分別為：甘特圖、WBBIMS、變更、消防安全設備、救災、領域團隊、銷售、植栽、施工介面、檢查、施工規範、團隊、介面溝通、專業分包商、進度、檢查維護、疏散。

其中 WBBIMS 為該篇文件所做的研究中，建置的 Web-based BIM 資訊管理平台的縮寫，因此該詞彙確實與系統這個概念相關。

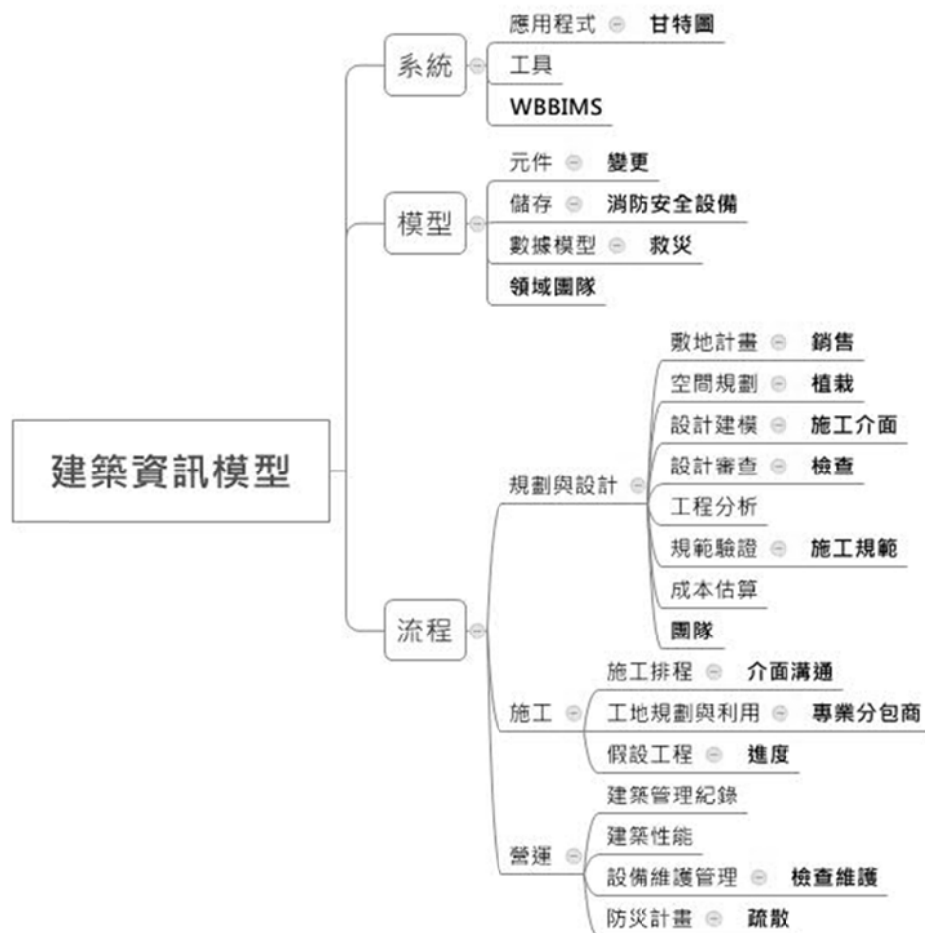


圖 25 第一次增補後的知識本體

● 第二次增補成果

依照增補知識本體的流程，以第一次增補的知識本體為基礎，再次進行資訊檢索，並根據檢索結果對知識本體的內容進行增補。其成果為第二次增補的知識本體的架構多增加了一層，變為六層架構，增加 13 個知識概念，其知識本體架構如圖 26。在此增補過程中增加的知識概念以粗體標示，分別為：工程進度、版次、類行、救災人員、跨領域、維護人員、景觀、建築、施工介面溝通、衝突檢、更新、檢查維護人員、避難

其中「類行」這個知識概念，經過檢查文件中的內容後，確定為「人類行為」斷詞錯誤的結果。另一個「衝突檢」則為「衝突檢查」斷詞錯誤的結果。

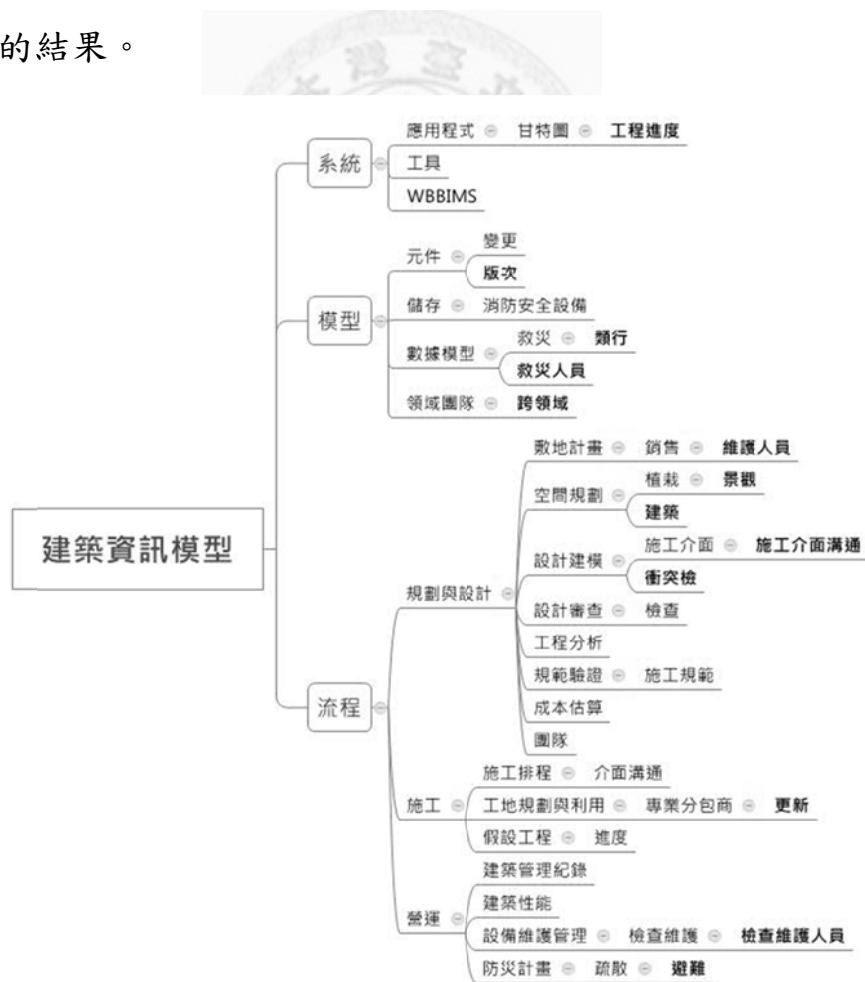


圖 26 第二次增補後的知識本體

● 第三次增補成果

在此次增補過程中知識本體的架構多增加了一層，變為七層架構，增加了 5 個知識概念，其知識本體架構如圖 27。在此次增補過程中增加的知識概念以粗體標示，分別為：變更設計、災害疏散、NILM、統包、管線。

其中 NILM 為非侵入式負載監測(Nonintrusive Load Monitoring)的縮寫，其用途為監測各式電器開關，於該篇文件的研究中，其用途為用於監測何處的電器已無法使用，來確認目前的建築物受災情形。

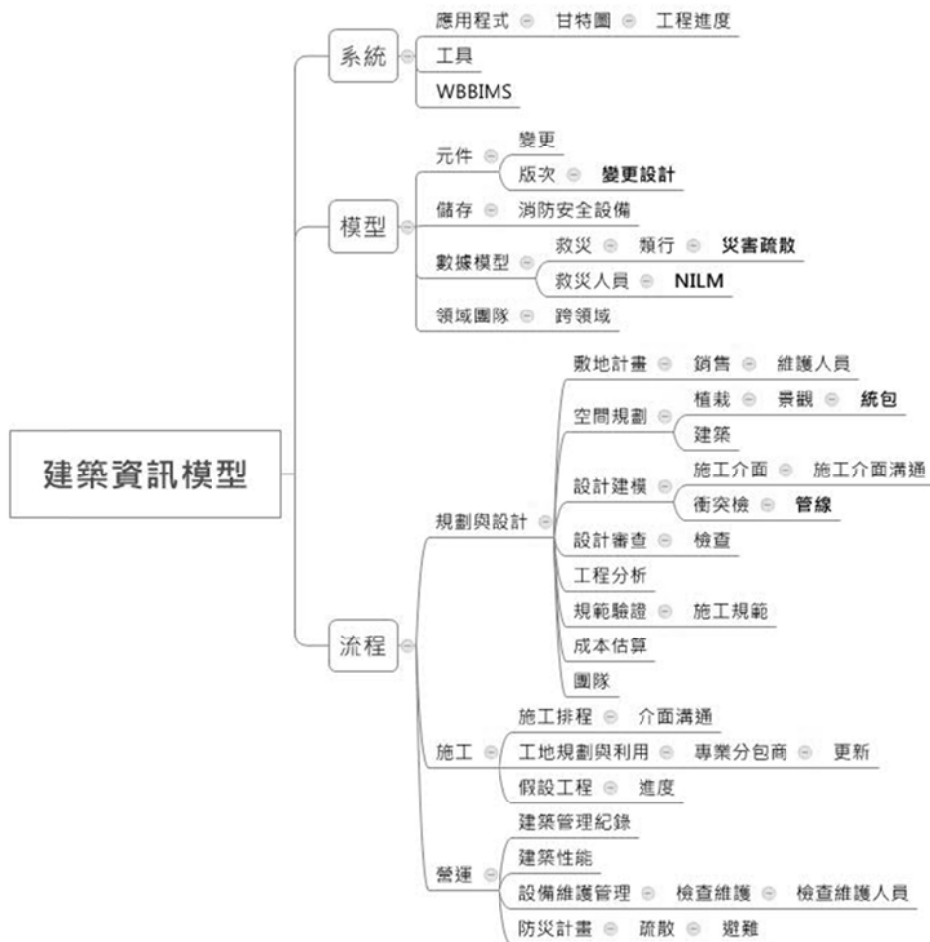


圖 27 第三次增補後的知識本體

## ● 第四次增補成果

在此次增補過程中知識本體的架構多增加了一層，變為八層架構，但只增加 1 個知識概念：界面問題。其知識本體架構如圖 28，並將增加的知識概念以粗體標示。在此次增補後，整個增補知識本體的流程結束，產生最後的領域知識本體。

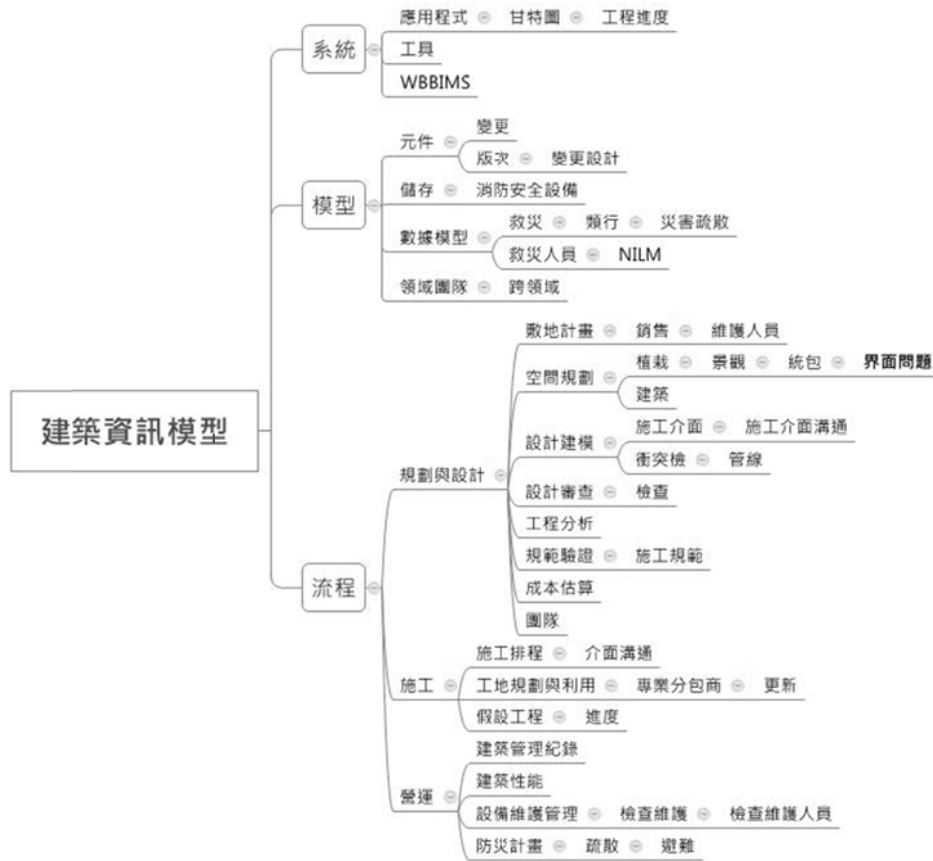


圖 28 第四次增補後的知識本體

經過四次增補之後，所產生的領域知識本體為八層架構、62 個知識概念。而由上一節的實驗中，可以知道使用領域知識本體進行資訊檢索的成效會比只使用基本知識本體的資訊檢索的成效好。因此利用資訊檢索技術確實可以對知識本體的內容進行增補，並達到增進資訊檢索系統之成效的目的。

## 第五章 結論與建議

### 5-1 結論

隨著資訊時代的到來，資訊的數量每日都在快速的增長，因此若無妥善管理資訊的方式，則資訊中所含有的知識就無法迅速的被使用者利用，因此不論在何種專業領域內，都可以運用知識本體技術對該領域內的知識進行整理。但建置知識本體的過程需要許多領域專家的參與，且需隨著專業領域的發展不斷的對知識本體的內容進行修正，因此發展知識本體的過程將耗費許多時間，所以本研究為縮短建置特定領域的知識本體時，所需花費的時間與人力。提出利用資訊檢索技術進行增補知識本體內容的方法。在研究中以建築資訊塑模領域為例，從與建築資訊塑模相關的文獻資料中，整理並分析建築資訊塑模領域的知識概念，並將這些知識概念建置成建築資訊塑模的基本知識本體後，使用以知識本體進行資訊檢索的檢索方法，對建築資訊塑模領域的文件集合進行檢索，由檢索結果中進行知識概念擷取，來進一步發展知識本體的內容。

而為了證明以知識本體進行資訊檢索的檢索方法，確能增進資訊檢索技術之成效，需有測試文件集來進行資訊檢索方法的成效測試，但由於目前在建築資訊塑模中無合適的測試文件集可以使用，因此本研究建置了建築資訊塑模領域的小型測試文件集，並根據增補知識本體的流程，對 N 及 K 兩參數的不同數值對於結果的影響進行分析。其驗證的結果為在使用 N=3 及 K=4 的參數下，以知識本體進行資訊檢索的檢索方法的成效確實優於未使用知識本體進行資訊檢索的檢索方法的成效。

承上所述，本研究之具體貢獻如下：

- 提出使用知識本體進行資訊檢索的檢索方法，此方法能提高資訊檢索技術的成效表現，並於不同知識領域皆可以使用此一檢索方法。
- 提出以資訊檢索技術進行增補知識本體的流程，此一方法由分析檢索結果來擷取知識概念，並將擷取出的知識概念增補進知識本體中。此方式可大幅減少在建置知識本體時領域專家所需參與的時間。

## 5-2 未來研究方向建議

資訊檢索與知識本體的理論基礎已經發展得相當成熟，而在應用方面也已經有許多資源與工具可以使用。但在建築資訊塑模領域，目前還缺乏因應此領域的文件內容所進行發展的知識本體與資訊檢索技術。而由本研究提出以知識本體進行資訊檢索的檢索方法及增補知識本體的流程，尚有下列發展空間：

- 在建築資訊塑模領域中尚無合適的測試文件集，因此本研究建置了小型測試文件集，未來應繼續補充該測試文件集的內容，以增加在建築資訊塑模領域中不同子領域的文件。
- 在進行知識本體內容的增補時，有 N 及 K 兩個參數需進行設定，所以在本研究中對於不同的參數設定進行比較後，得到使用 N=3 及 K=4 可得到較佳的表現。因此未來可對於兩參數在不同文件集合中，影響增補知識本體的結果進行比較，以進一步了解兩參數與增補知識本體的相互關係。



- 本研究為進行實作來得到此方法的成效表現，因此只選用中文文件來做為文件集合，但此知識本體增補流程在不同語言上皆可進行實作，因此未來可針對不同語言的文件進行分析，來更加豐富知識本體的內容。
- 因應建築資訊塑模領域為一新發展的知識領域，因此使用 n-gram 方法進行知識概念的擷取，而此一方式在本研究中仍會有擷取錯誤的狀況發生，因此未來可進行建築資訊塑模領域的辭典建置，來減少在進行知識概念擷取上的錯誤發生。
- 在目前的知識本體架構中，不同知識概念之間關係皆為「包含」的關係，因此未來可藉由分析文件內容中，找出不同知識概念在語句間進行連接的詞彙，來修正目前的關係連結。

總和以上的部分，在建築資訊塑模領域的知識本體建置技術與資訊檢索技術尚有繼續發展之空間，且目前所進行之研究皆以文字為基礎，而在建築資訊塑模領域中，有許多的資料是以三維虛擬方式表示，因此在未來的研究上可再進一步往此方向進行發展。

## 參考文獻

- [1] A.S. Sidhu, T.S. Dillon, E. Chang, and B.S. Sidhu, "Protein ontology: vocabulary for protein data," in *Proceedings of 3rd IEEE ICITA'05*, Sydney, pp. 465-469 vol.1, 2005.
- [2] R.G. Raskin and M.J. Pan, "Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)," *Computers and Geosciences*, vol. 31, pp. 1119-1125, Nov 2005.
- [3] T.R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," *Human-Computer Studies*, vol. 43, pp. 907-928, Nov-Dec 1995.
- [4] *OWL Web Ontology Language*. Available: <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>
- [5] N.F. Noy and D.L. McGuinness, "Ontology development 101: a guide to creating your first ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
- [6] M. Uschold and M. Gruninger, "Ontologies: principles, methods and applications," *Knowledge Engineering Review*, vol. 11, pp. 93-136, Jun 1996.
- [7] D.A. Campbell, "Building information modeling: the Web3D application for AEC," in *Proceedings of the twelfth international conference on 3D web technology*, Perugia, Italy, pp. 173-176, 2007.
- [8] CRC Construction Innovation, "Adopting BIM for facilities management : solutions for managing the sydney opera house," Cooperative Research Center for Construction Innovation, Brisbane, Australia, 2007.

- [9] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM handbook: a guide to building information modeling for owners, managers, designers, engineers and contractors*, Wiley, 2008.
- [10] National Institute of Building Sciences, "National BIM standard - United States," 2012.
- [11] Computer Integrated Construction Research Program, "BIM project execution planning guide," The Pennsylvania State University, July 2010.
- [12] Institute for BIM in Canada, "Environmental scan of BIM tools and standards," 2011.
- [13] 林顯堂，「以知識本體為基礎之特定領域文件段落擷取方法及應用」，國立臺灣大學土木工程學系博士論文，2009。
- [14] CEDICT. *Chinese English Dictionary*. Available: <http://www.mdbg.net/chindict/chindict.php>
- [15] 劉君祖，「牛頓工程辭典」，牛頓出版股份有限公司，台北市，1989。
- [16] 何松柏，「以BIM與代理者技術實現營建協同設計審查之研究」，國立臺灣大學土木工程學系碩士論文，2010。
- [17] C.D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*, Cambridge University Press, 2008.
- [18] R.A. Baeza-Yates and B.A. Ribeiro-Neto, *Modern information retrieval*, Addison Wesley, 1999.
- [19] E.M. Voorhees, "Variations in relevance judgments and the measurement of retrieval effectiveness," in *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, Melbourne, Australia, pp. 315-323, 1998.

- [20] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, pp. 37-46, 1960.
- [21] J. Carletta, "Assessing agreement on classification tasks: The kappa statistic," *Computational Linguistics*, vol. 22, pp. 249-254, Jun 1996.
- [22] G. Salton, A. Wong, and C.S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, pp. 613-620, 1975.
- [23] G. Salton and C. Buckley, "Term-Weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, pp. 513-523, 1988.
- [24] W.R. Hersh and R.T. Bhupatiraju, "TREC genomics track overview," in *Proceedings of the twelfth Text Retrieval Conference*, pp. 14-23, 2003.
- [25] P.F. Brown, P.V. deSouza, R.L. Mercer, V.J.D. Pietra, and J.C. Lai, "Class-based n-gram models of natural language," *Computational Linguistics*, vol. 18, pp. 467-479, 1992.