

以圖形結構為基礎之企業智產檢索系統

黃羿綺
國立政治大學
lucy@csim.scu.edu.tw

江家榕
國立政治大學
chiajung.cs97@gmail.com

甯格致
國立政治大學
kcning@nccu.edu.tw

蕭暉議
資策會前瞻所
hhsiao@iii.org.tw

林謙
資策會前瞻所
stevenlin@iii.org.tw

藍啟維
資策會前瞻所
cwlan@iii.org.tw

陳良弼
國立政治大學
alpchen@cs.nccu.edu.tw

劉吉軒
國立政治大學
jlsliu@cs.nccu.edu.tw

摘要

現今企業智產資料數量龐大，種類應用的範圍亦趨複雜，檢索相關智產以達到智產之再利用，往往是一項耗費人力與時間的艱辛任務，資訊維護效果亦不彰。本研究以圖形結構及屬性內容兩層面開發智產庫搜尋技術：圖形結構層面主要處理智產階層結構關係，發展智產結構之間的相似度評量與排序機制，並找出與查詢智產結構最相近的目標智產；屬性內容層面則配合圖形結構快速檢索，以目標智產文件標籤內容進行字串比對，處理智產間描述內容或屬性之相似程度。本研究亦發展智產屬性與內容的相似度評量排序機制，用以找出與查詢智產屬性與內容最相近的數個目標智產，期能提供更完善之智產相似度評量，達成更有效之搜尋比對效果，進而促進智產再利用之具體成果。

關鍵詞：企業智產管理、智產再利用、智產庫應用、資訊檢索、圖形搜尋

Abstract

Enterprise asset management is crucial to achieving stronger competitiveness by faster time-to-market and lower development cost. While it remains a great challenge to well organize inter-relationships among assets which have diversified modalities, our study has addressed several shortcomings of enterprise asset management according to the recent research such as the relatively low usability of asset resources, asset isolation, and lacking of solution as asset portfolio. We are tackling these problems with a hierarchical asset framework, a graph search method and a text matching all together.

Keywords: enterprise asset management, asset reuse, graph search, information retrieval, graph search

1. 智慧資產管理發展現況及問題

在現今資訊時代裡，數量可觀的各式資訊在軟體、硬體及創新思維的推波助瀾下，以各種風貌應用於不同類型的團隊需求中。而各種因應而生之資訊系統開發亦隨著企業組織不斷地累積經驗，如：業務企劃、銷售規劃、資訊系統研發、資訊服務等。這些資訊系統開發，都會憑藉其專案執行過程中習得的知識與經驗，持續地建立相關的智慧資產(簡稱為智產)以利後續再利用，相關類似的智慧資產(簡稱為：智產)例如：資訊服務標準作業流程，領域業務提案樣版，資訊系統函式庫等。然而，日積月累之後的企業智產不僅數量龐大，種類與可應用範圍也漸趨複雜，對於新建立之智慧資產如何連結與既有智產間的關係，往往是一項耗費人力與時間的艱辛任務，且經常有許多謬誤，從而導致智產間關係的建立與維護效能不彰，許多可被再利用的智產因此難以被發掘，衍生出智產孤島效應[1][3][11][13]。因此，面對越趨嚴峻的全球化競爭趨勢，企業必須能夠快速地回應各種客戶需求，而使企業加速提供敏捷、可重複使用並可延伸的解決方案能力，對於既有智產持續性的累積與管理運用，需要建立一套有效的智產管理機制。

然而，企業中智產系統的發展仍受到了許多的限制。首先，智產資訊的儲存資料結構必須要先能夠妥善地規畫定義，以便在前端的資訊檢索呈現良好及友善的搜尋結果，而對後端的資訊管理人員而言，亦能夠有效地更新這些智產資料。再來則是智產與智產間是有階層式或是相關性等關聯存在的，若要避免智產資訊的孤島效應並強化可再利用性的話，則勢必亦需要一妥善的方案來維持智產間的關聯資訊。最後，一個完善的智產檢索系統應當要以能提供完整的智產解決方案為目標，而非僅僅只是將符合檢索條件的智產文件呈現作為搜尋結果[9]。

在本研究中，我們提出一套完整的智慧型智產檢索系統。透過已建立之階層化智產描述架構[2]為基礎，強化智產搜尋比對功能，並以智產有效再利用及智產快速搜尋配對為主要目標。本研究將同時以圖形結構及屬性內容兩個層面，進行比對搜尋方法之設計與發展。在圖形結構層面，主要處理智產之階層結構關係，發展出智產結構之間的相似度評量與排序機制，進而找出與查詢智產結構最相近的數個目標智產。在屬性內容層面，則以文字搜尋技術為基礎，處理智產描述內容及屬性之關聯關係，同樣的，也發展出智產屬性與內容之間的相似度評量與排序機制，進而找出與查詢智產屬性與內容最相近的數個目標智產。兩個層面的智產相似度評量機制也將互相整合，提供更完整、更全面之智產相似度評量，達成智產關聯之有效搜尋比對，進而促進智產再利用之具體成果。歸納而言，本研究提出之重點在於：

- (1) 以圖形結構為檢索要件之解決方案搜尋機制，有效減少傳統需對個別智產進行檢索的繁複過程。
- (2) 依智產描述的多標籤內容比對，有效率地匹配查詢字詞與目標智產，提昇查詢效率與有效性。

2. 智產檢索系統概況及挑戰

以現有市場的資產管理軟體分布來看，與智產庫相關的資產管理軟體大致可以區分為 Knowledge Management (KM) / Content Management System (CMS)、Software Asset Management 及 Information Technology Service Management (ITSM) / Information Technology Infrastructure Library (ITIL) 這三大類，如 Microsoft SharePoint、Oracle WebCenter、IBM Rational Asset Manager、IBM BPM 的 Process Center、IBM Tivoli、BMC Business Service Management 及 HP 的 Asset Manager，以智產庫系統服務的幾個層次及內容關連程度展開其分布如圖 1。

綜觀目前智產相關之檢索系統設計或智產資訊可再利用性議題，大多採用以下數種策略：(1) 智產文件之內容通常以半自動或是全

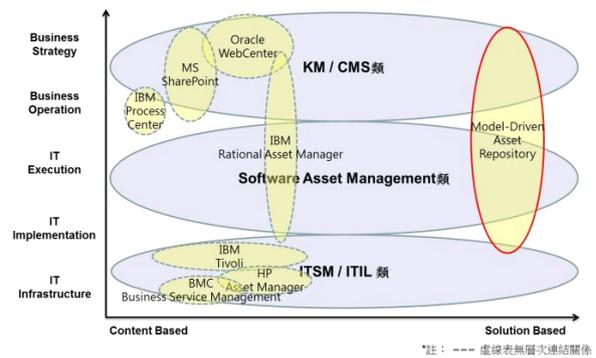


圖 1. 企業智產庫系統服務及內容關係圖[2]

自動化方式處理[3][11][12]；(2) 智產資料在被更新入資料庫時，需擷取相關資訊(文字內容、標記等)及關鍵字，以便建立索引供之後檢索用；(3) 由專家預先設定好之智產分類原則，於檢索時直接篩選後選的目標智產方案資料範圍，來提升搜尋時的效能問題[13]。已知的智產相關應用成果通常都僅著重於單一文件的搜索比對、或是基於後設資訊來搜尋較符合的智產檔案、軟體元件等。然而智產資料庫在應用層面的需求，往往不是僅限於單一智產文件的檢索，而是需要多個智產的組合作為一較完整之智產解決方案；僅考慮單一文件這類的檢索策略往往僅侷限於單一智產文件內的符合結果，進而評估生成符合條件的一項或多項智產文件清單，卻無法提供更切合使用者需求的智產解決方案，也就是經由多個智產文件組合比對後全面性的檢索結果[14][16]。這樣子的檢索範圍落差，除了造成智產檢索上時間的耗費外，也忽略掉智產文件之間的關聯性，比對到的結果往往不符合實務應用需求，整體而言效率上亦是難以達到理想。對企業而言至關重要之智產檢索功能需求，應當要能夠盡可能兼顧到企業的流程化處理、運作及服務條款、基礎硬體設施需求、智產方案的有效組合及再利用。歸納目前智產應用的障礙，大致可列舉為以下三點：

- (1) 缺乏以搜尋解決方案(多項智產組合)為導向的智產檢索功能。
- (2) 不同智產方案中相同元件(Functions)的再利用率偏低。
- (3) 智產孤島效應。

3. 智慧型智產檢索系統

本研究開發以圖形結構為基礎之比對功能模組，為達到符合要求之搜尋精確度，提升智產解決方案之可用性及重用性。我們首先以圖形之搜尋比對技術進行背景分析與探討。圖形被廣泛的運用在表示結構化的事物上，例如：化學化合物的架構、生態方面的食物鏈、網頁的鏈結狀態等。此外，圖形資料的來源也越來越廣，資料量增加的速度十分迅速，例如：網際網路、社群網路、基因資料庫、與政府資料等。隨著圖形資料庫於各層面的應用，圖形搜尋逐漸演變成重要的研究議題。

雜度，協助去除同義詞及多義詞之模糊性[15]。而以我們使用的階層化智產庫而言，同一階層中的某一點，與此點所連接的邊有兩種關係，第一種關係為上下層關係，如策略層的"節省成本"連到作業層的"設備監控流程模型"、作業層的"設備監控流程模型"連到執行層的"設備資料蒐集服務規格"；第二種為橫向關係，如"設備監控流程模型"連接到"中控管理系統資訊模型"，因為兩個智產的 Layer 是相同的。

圖形搜尋有三種查詢模式，如下所列：

- (1) 完全比對：找出查詢圖為其子圖之智產圖。

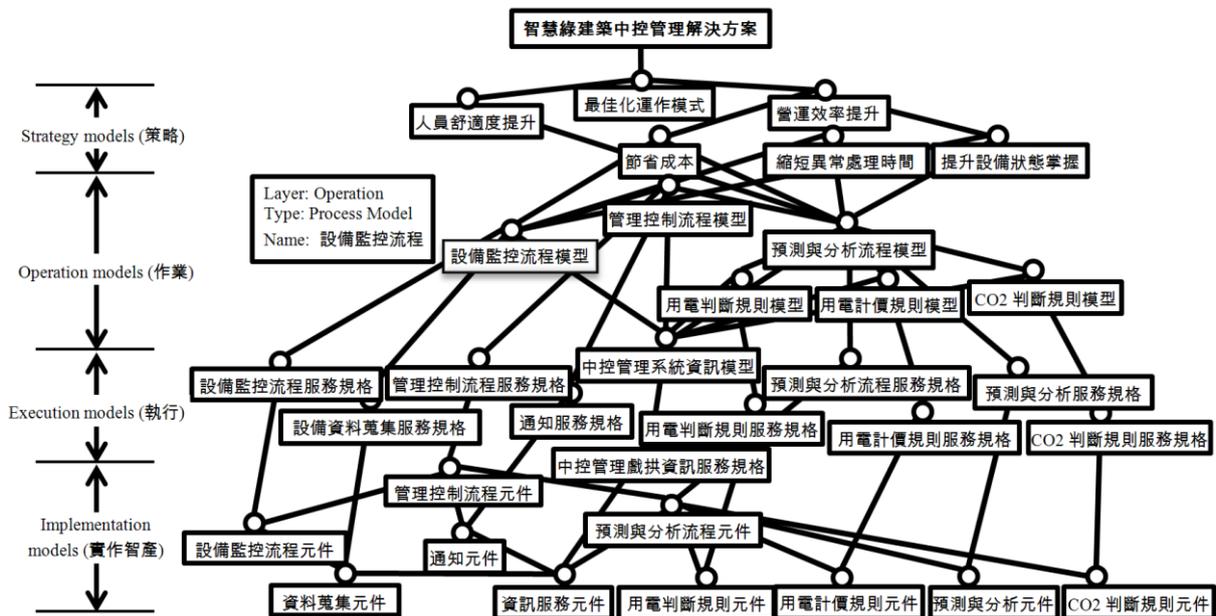


圖 2. 階層化智產庫架構

3.1 圖形結構層面搜尋

圖形資料其實是由圖論所推演而來。圖中包含點(node)、邊(edge)，其中點可以延伸為一個物件(object)的概念。每個點上會帶有各自的屬性(attribute)，具獨特意義的屬性則可視為標籤(label)，而邊則代表兩點之間的關係。圖 2 中以智產庫為例，"設備監控流程模型" 這個點的屬性有 Layer: operation、Type: process model、Name: 設備監控流程模型，而其標籤為 Name: 設備監控流程模型。而特定應用領域中的概念可以被以階層的方式，建構出相互對應關係，而詞彙與概念的對應也可相對建立，並掌握其一對多或多對一關係。像是語意網路與本體知識模型的採用有助於降低資訊檢索問題的複

- (2) 相似比對：找出與查詢圖相似之智產圖。

- (3) 組合比對：透過多張智產圖組合而得查詢圖。

其基本定義為給定一查詢圖，依據查詢需求針對單點與各點之間關聯，搜尋符合條件之資料圖。而此種搜尋所做的同構(isomorphic)比對已經被證明為 NP-complete 問題，故常見的搜尋技術為避免過多的計算，多採用過濾與驗證方法。其中，過濾的基準有許多種而未被過濾掉的查詢圖則要再經過逐點驗證以確保符合使用者需求。在完全比對方法中[5][7]採用擷取特徵方法，並建立索引加速過濾，[4]則是計算最小圖形距離並建立索引。相似比對方法

[19]使用計算兩條相似路徑後比對兩圖相似度，[6][8]計算兩圖距離相似度並做索引。組合比對[18]提出當無法由單一圖找到解時組合多個圖做為解。本篇論文為提供三種搜尋，故採用特徵擷取方式以方便於組合及比對。

3.2 屬性文字內容層面搜尋

近二十年來網路應用與使用者參與的蓬勃發展，很大程度上立基於搜尋引擎(search engine)的強大功能。搜尋引擎協助使用者在廣泛的資料來源及大量的資料中，快速的找到可能符合其資訊需求的資料，其中最關鍵的網頁或文件搜尋技術，乃源自學術研究中的資訊檢索領域(information retrieval)。資訊檢索的問題定義為給予一個含有大量文件的集合及一個使用者資訊需求的表達，找出一個由相關文件組成的子集合。因此，以數學的角度而言，資訊檢索的問題可以化約為從一個大集合中找出一個子集合，其成員必須符合相關性的條件[13]。

資訊檢索系統內部的核心運算為檢索詞與文件內容之間的字串搜尋比對，再依據搜尋比對結果計算各文件與檢索詞的相關程度，進而產生篩選及排序的相關文件子集合。目前資訊檢索系統的實際應用，通常都是在大量文件資料的情形下，希望能透過搜尋而提供有用資訊，在最短的時間內，充分發揮滿足使用者資訊需求的效益。因此，資訊檢索系統的運算速度成為系統是否能實際應用的重要關鍵之一。對

於搜尋問題的傳統方法為字串比對(string matching)，也就是將每一份文件視為一個很長的字串，再用字串比對演算法，將檢索字串與文件字串，從頭到尾的循序比對，而搜尋到使用者的檢索詞出現在哪些文件中。然而這種作法是屬於線性演算法，搜尋時間與文件長度及文件量成正比，而且必須是在線上即時的情境下操作。因此，傳統的字串比對法，在面對大量的比對基礎下，需要也因應提升硬體處理的效率，才以便能在合理的時間內，提供大量文件的檢索功能[17]。

然而儘管有先天上的限制，而以目前資訊檢索技術在運算速度考量上，大都以索引建立(indexing)的技術來因應，則可以透過資料庫快速比對提取目標資訊。然而實務上而言，索引的建立不光是索引自身的資料量有可能隨著文本量而大量擴張以外，索引的更新及維持也是需要耗費大量運算時間空間的任務。

3.3 智慧型整合搜尋

考慮以上兩項搜尋比對的原則，並應用於階層化智產架構資料中，本研究提出一個以圖形搜尋為基礎的智慧型智產檢索系統，除了考量階層化智產庫中，兩兩智產之間的關聯性以外，也包括了智產文件中描述相關細節之屬性文字內容比對，而在每一次進行檢索時，先行以智產圖形結構快速比對，篩選出可能的智產方案集合，再依據智產內容中的細節文字比對結果，來再次篩選最接近檢索條件之智產方案集合，最後輸出符合的智產結果。

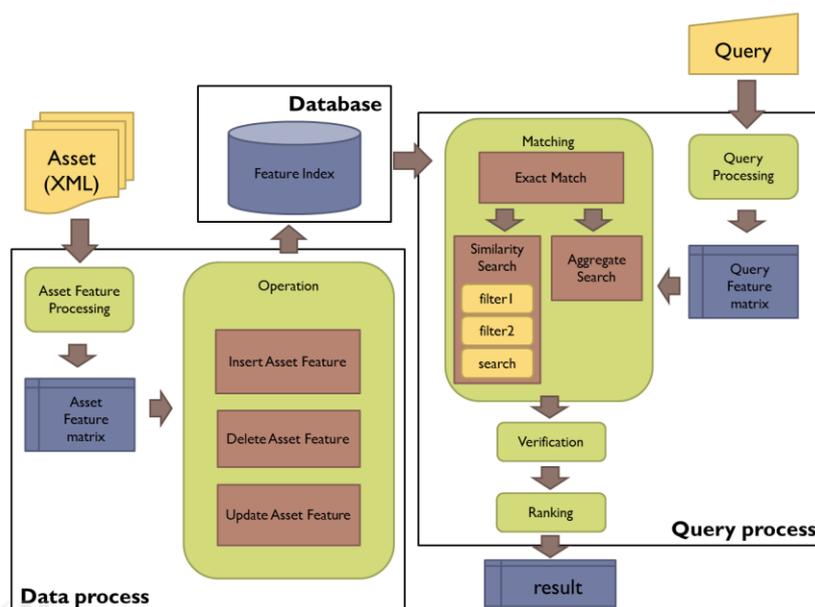


圖 3. 系統架構圖

我們希望能妥善利用階層化智產庫的優勢，配合圖形檢索根據已建立好的特徵庫進行快速搜尋比對，最後才利用傳統文字比對的方式來確認最後輸出的最符合智產方案結果。此設計即能達成不但考量智產之間關連性、避免智產孤島效應、增加智產的再被利用率、更於效能上有別於以往繁複冗贅缺乏效率的檢索策略，遂能建構出一個具有實用價值的智慧型智產檢索系統。

4. 系統實作與方法

目前智產庫已經發展出階層式智產描述架構，針對此架構將結合圖形結構及屬性內容兩個層面，進行比對搜尋方法之設計與發展，如圖 3。利用擷取智產庫中各智產圖的結構特徵圖，比對查詢圖的結構特徵圖，找出等於或近似於查詢圖之智產候選圖。若是無法由單一智產圖中尋找到相似智產圖時，可藉由多個智產圖的組合求得候選答案。最後再經過屬性驗證，針對符合使用者需求的智產候選圖排序以回傳最終解。

4.1 智產庫特徵圖擷取

針對智產庫中每個智產圖作特徵圖擷取。由於智產庫有階層的架構，可將智產圖之特徵圖分三類擷取：(1)策略層的同層特徵擷取 (2)僅上下層之間的特徵擷取、以及 (3)上下層+下同層特徵擷取。假設智產庫中有多個智產圖如圖 4 所示，針對每個智產圖可以擷取出的特徵圖如表 1 所示。

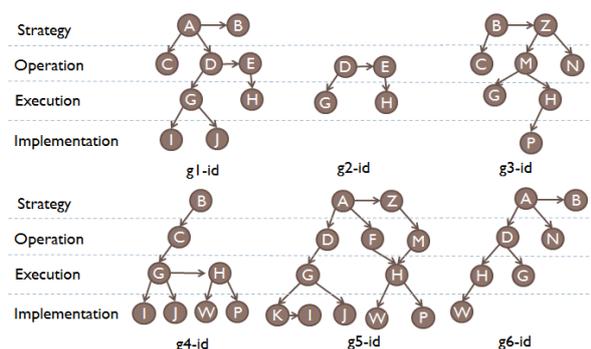


圖 4. 多個智產圖呈現於不同階層狀況

針對不同的智產圖，我們可以擷取出不同的特徵圖。其中，若智產圖並非完整四層結構時，將增設虛擬點：”@”，以利符合智產圖特徵擷取之條件。將這些特徵圖整理記錄後形成特徵表，除基本特徵資訊外，因智產之間具有階層關係，同名之智產位於不同階層視為不同智產，故特別儲存特徵圖之第一個點位於哪個

階層，以利初步過濾不合條件之特徵圖。

4.2 查詢圖特徵擷取

對於查詢圖來說，查詢圖是一個由多個智產組成的階層關係，因此在進入系統後亦會先做特徵圖的擷取，再利用這些查詢特徵圖做為查詢的目標。

4.3 圖形比對

查詢結果需回傳包含查詢圖之智產圖、相似於查詢圖之智產圖或是由多張智產圖組合成包含查詢圖的解。

表 1. 智產圖結構特徵表

Feature IDs	Feature Types	First Node's Layer	Features	Graphs
1	1	Strategy	A, B	g1-id
2	2	Strategy	A, C	g1-id
3	2	Strategy	A, D	g1-id
4	3	Strategy	A, D, E	g1-id
5	2	Operation	D, G	g1-id
6	2	Operation	E, H	g1-id
7	2	Execution	G, I	g1-id
8	2	Execution	G, J	g1-id
9	3	Strategy	@,D,E	g2-id
10	2	Operation	D, G	g2-id
11	2	Operation	E, H	g2-id
...

4.3.1 完全比對

智產圖中有與查詢圖完全相同之圖形。假設查詢圖所擷取的特徵有 n 個，因特徵擷取方式採重疊取法，故比對至智產庫中有包含此 n 個特徵之智產圖即為此智產圖有包含完整的查詢圖。根據表 2 可以得到 g1-id 為包含查詢圖之智產候選圖。

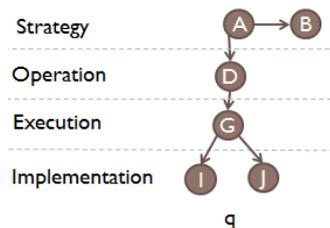


圖 5. 查詢圖

表 2. 智產圖是否存在查詢圖之特徵圖

Feature Types	First Node's Layer	Features	g1-id	g2-id	g3-id	g4-id	g5-id	g6-id
1	Strategy	A, B	1	0	0	0	0	1
2	Strategy	A, D	1	0	0	0	1	1
2	Operation	D, G	1	1	0	0	1	1
2	Execution	G, I	1	0	0	1	0	0
2	Execution	G, J	1	0	0	1	1	0

4.3.2 相似比對

若查詢圖擷取出的 n 個特徵中，比對至智產庫中智產圖 m 個特徵 ($m < n$)，則使用相似度比對，尋找出與查詢圖相似的智產圖。在做相似比對之前，會先針對智產圖中是否存在查詢圖所擷取出之特徵圖做第一次過濾，以減少需做比對之智產候選圖。使用者決定其所需回傳結果之智產圖與查詢圖至少相似程度為 $x\%$ ，而相似特徵圖計算中最高相似度為 y ，可得知若智產圖中未包含查詢圖之特徵圖數量大於等於 $z(y^z < 0.x)$ 時不符合使用者需求，即可從候選圖中刪除此張智產圖。以圖 4.、圖 5. 及其生成之表 <X5> 為例，若使用者需相似度為 75% 以上之智產圖，且相似特徵圖最高分數為 0.9，即是智產圖缺少 3 個以上特徵圖就無需做比對，候選圖中只剩 g1-id、g5-id 以及 g6-id。

表 3. 智產圖是否包含查詢圖之特徵結構

Feature Types	First Node's Layer	q	g1-id	g5-id	g6-id
1	Strategy	1	1	1	1
	Strategy	1	2	3	2
2	Operation	1	2	3	2
	Execution	2	2	4	1
3	Strategy	0	1	0	0
	Operation	0	0	0	0
	Execution	0	0	1	0

第二次過濾則是經由查詢圖之特徵結構比對，智產圖中各個特徵圖結構必須大於等於查詢圖對應之各個特徵結構。以表 3. 為例，其中 g6-id <2, Execution> 的特徵結構數量為 1 而查詢圖 <2, Execution> 的特徵結構數量為 2，因此 g6-id 的特徵結構數量小於查詢圖特徵結構數量，候選圖中只剩 g1-id、g5-id。

相似圖的搜尋是透過相似特徵圖做判斷，而相似特徵圖有兩種情形：(1) 智產圖之特徵圖所擁有的點集合包含查詢圖之特徵圖的點集合，且此兩特徵圖共同點之排列順序及階層關係皆相同、(2) 智產圖之特徵圖與查詢圖之特徵圖的點之數量、階層相同，但點之屬性並非完全相同，而是有高相似度。

第一種相似特徵圖可以藉由放置 wildcard * 於查詢特徵圖各字元間做搜尋而得，取與查詢特徵圖長度越相近者計。以查詢圖之 GI 特徵圖為例，圖 4. 之 g5-id 中的 GKI 符合相似特徵圖之情況(1)。

第二種相似特徵圖則是藉由選取在同張

圖上點數量相同、階層相同之特徵圖後，與查詢圖之特徵圖做兩兩比對，取相似程度較高者計。此兩種相似特徵圖皆會計算與查詢特徵圖之間的相似分數，而此分數需再乘以一懲罰值 $p(0 < p < 1)$ ，以利於跟完全比對作區隔，故相似特徵圖之分數會介於 0 至 p 之間。且逐一計算智產圖所缺少之特徵圖時會同時做智產圖與查詢圖之間的相似分數計算，若目前分數已經小於使用者所規定之相似程度 $x\%$ ，則此張智產圖不可能符合使用者需求，故無需再計算。而相似特徵圖之情況(2)則以查詢圖之 AB 特徵圖與圖 4. 之 g5-id 中的 AZ 為例，此相似特徵圖之分數由 B 與 Z 相似度決定。

4.3.3 圖形組合比對

若完全無法找到相同或相似的圖形，或是回傳解數量不足時，可利用兩張以上的智產圖做組合，以產生包含查詢圖的智產組合圖作為答案。所需組合之智產圖數量越少越佳，即是指兩張智產圖所組合而成的答案會比由三張以上智產圖所形成的答案較佳。組合方法是先確定查詢圖之特徵圖是否皆存在於智產庫中，若無則無法完整組合，若有則將各張智產圖依其擁有查詢圖之特徵圖數量做排序，再依序針對各張智產圖缺少之部分搜尋能補足之智產圖，直到完整組合查詢圖為止。

4.4 屬性內容比對

經由智產檢索輸入的圖形結構與階層化智產庫所建立出來的圖形結構特徵資料庫比對後，我們已可有效地初步篩選出可能的智產解決方案，然而智產文件內部涵蓋的諸多標籤中，仍有用以描述該智產詳細內容、或是附加說明的文字部分，這部分的文字亦須被納入比對的考量，如此才能夠更符合檢索的條件。

這部分我們針對目標字串比對的原則如下：

- (1) 首先找出那些已經透過圖形結構篩選後，提取相對應存在於智產庫中的智產文件內容及其關聯。
- (2) 依據檢索目標圖形關聯，建立一連串兩兩比對的智產文件清單(包括智產的索引以及即將用來比對的欄位內容或是標籤內容)。
- (3) 以 Levenshtein Distance 演算法[10]，以 fully-connected pairwise 逐個計算出評估後的相似度數值，亦即兩兩字串比對

之差異量(string distance)，而對於 Levenshtein Distance 演算法中的權重參數(replace、insert、delete)，則給予均一的基本權重。

- (4) 相似度數值從完全不同至完全一致，數值將介於 0.0~1.0，我們將每一筆(pair)的相似值加總並進行均化(normalization)。
- (5) 回傳該個智產方案組合與檢索目標的字串比對相似度值。

4.5 驗證與排序

將查詢圖上的所有點逐個對應於經過比對後留下之智產候選圖，透過屬性比對若皆有對應處，代表此智產候選圖為近似查詢圖之智產圖。最後再依照智產圖與查詢圖之間的相似度做排序，選出最近似於查詢圖之前 K 張智產圖。

5. 系統實務應用結果

此查詢系統包含資料庫建置、系統介面以及結果呈現，將儲存為 XML 檔的智產圖資訊讀取後，擷取特徵圖存入特徵資料表，而解決方案資料表則是儲存智產圖的名稱與 ID，如圖 6 所示。

特徵圖資料表

欄位	型態	Null	註解
<u>Id</u>	int(10)	否	流水編號
FeatureType	int(10)	否	特徵圖類型 Strategy、 Operation、 Execution、 Implementation
FirstNodeLayer	varchar(255)	否	智產階層(1、2、3、4)
FirstNodeLayer Type	int(10)	否	
Features	varchar(255)	否	特徵圖
FeaturesId	varchar(255)	否	特徵圖 ID
GraphId	varchar(255)	否	解決方案 ID

智產解決方案資料表

欄位	型態	Null	註解
<u>Id</u>	int(10)	否	流水編號
SolutionId	varchar(255)	否	解決方案 ID
SolutionName	varchar(255)	否	解決方案名稱

圖 6. 資料庫之特徵資料表、解決方案資料表

管理智產圖的模式有新增、刪除、修改。如圖 7，將 XML 智產圖所擷取的多個特徵圖新增至特徵資料表中。

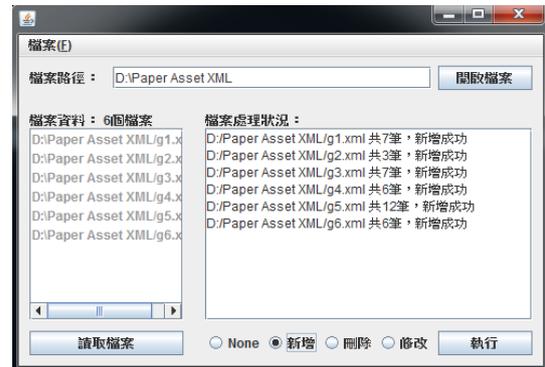


圖 7. 特徵資料庫之新增、刪除、修改

查詢智產圖的模式有四種：(1)預設 (2)完全比對 (3)相似比對 (4)組合查詢，其中，預設同時包含三種查詢模式。圖 8 則依序呈現完全比對狀況下單純輸出符合之智產圖名稱；相似比對狀況下則除了顯示智產圖名稱及其與查詢圖之相似分數外，還會顯示何處為相似特徵；最後，組合比對則會顯示由哪幾個智產圖做結合可得查詢圖，且各智產圖擁有的哪幾個查詢特徵圖。

6. 綜合效益評估

由於企業智產數量龐大，其種類與應用範圍也相當複雜，如何快速尋找有效解決方案是個困難點。過去方法中，關鍵字搜尋雖然可以找到許多符合各種查詢條件之智產，但解決方案是由多個不同的智產所組成，單純的關鍵字搜尋無法確定智產之間的關聯性，若想將找到的智產組合得到解決方案，則需耗費人力成本進行逐一比對或是利用既有智產組合方法拼湊，但既有智產組合方法隨著智產的數量增加，其所需比對的數量也增加，以效率上而言理想的智產解決方案結果的產生則相對不容易達成。

本智慧型智產檢索系統利用工具自動化萃取出智產庫中每個智產圖的結構特徵值，比對查詢圖的結構特徵值、相似路徑及其智產之屬性，找出等於或近似於查詢圖之智產圖，或是藉由多個智產圖的組合求得答案。以下就本研究提出之智慧型智產檢索系統的三個層面來說明優勢為何。



圖 8. 查詢結果輸出之完全比對、相似比對以及組合比對

(1) 人工手動比對與圖形搜尋比對

若使用人工手動比對，需要循序查找智產庫中所有的智產圖，且對圖中每個智產做屬性檢驗、智產間做關聯檢驗，若所需比對之智產關聯及屬性越多，其所以耗費時間也會更長。而使用智慧型智產庫檢索系統，在離線狀態(offline)下會針對各張智產圖擷取特徵，在線上狀態(online)當使用者需求下達時，也會自動對查詢圖做特徵擷取，且因特徵以包含智產之間的關聯，故只需對特徵資料庫做簡單的 SQL 查詢搜尋，將所需資料提出後運用演算法自動運算，快速地得到答案。

(2) 個別模式圖形搜尋比對與綜合圖形搜尋比對

- (a) 完全比對：在真實的情況下，智產圖與查詢圖完全相同的情形可能較少，且這樣的查詢僅適用於熟悉智產庫的專業人員運用，否則容易造成無任何回傳結果，以至於無法達到智產有效再利用。
- (b) 相似比對：相較於完全比對的搜尋條件較為寬鬆，可增加有回傳結果的機率，適合使用者不確定查詢時使用，但也只能提供多個單一解決方案。
- (c) 組合比對：針對肯定的智產關聯，想要尋找其他缺失部份的使用者，能夠利用多種解決方案組合以提供多種完善的解決方案，但並未提供相似的搜尋。

(3) 經由圖形結構篩選後之字串比對

傳統字串比對的作法是屬於線性演算法，也就是搜尋時間與文件長度及文件量成正比，因此在面對大量的比對基礎下，除了需要不斷維持索引的正確性外，也需要因應漸增的文件資料量而不斷提升硬體處理的效率，才能在合理的時間內提供大量文件的檢索功能。本研究使用圖形結構預先篩選出那些已經透過圖形結構篩選後，提取相對應存在於智產庫中的智產文件內容及其關聯再依據檢索目標圖形關聯，建立一連串兩兩比對的智產文件清單(包括智產的索引以及即將用來比對的欄位內容或是標籤內容)，以 fully-connected pairwise 依序計算出評估後的相似度數值，可大幅度減少字串檢索上所需要耗費的比對時間以及索引存放空間。

本研究之綜合效益評估可以表格方式呈現如表 4。在此綜合效益評估中，本研究評估比較了過去既有智產檢索系統以及傳統人工方法，針對當前智產檢索及應用相關文獻所提到的諸項重要問題(如智產再利用度、智產孤島效應問題以及提供以智產解決方案為導向的檢索結果)，及方法自身的時間、空間效率(例如檢索的精確度、耗費的時間等)，同時也考量了架構本身的擴充性是否能隨著資料擴張或是智產領域的特殊性而能夠進行調適。藉由此綜合性的評估，我們可以看出其中人工方法不光在智產檢索領域已知問題、方法自身的效率、以及架構擴充性上，效益都是最差的(或是無法達成)。此評估也說明了人工方法並不適合用來處理如智產庫的大量資料；而其他智產檢索系統則在這三大項目的評估中，僅字串檢索的部分尚能仰賴大量且有效的索引來達成檢索的功能，其餘的評估項目亦是效益不彰。而本研

究提出基於圖形結構之智慧型智產檢索系統，由於是針對解決智產檢索的問題而發展的，在此三類效益評估的表現都是較好的，唯因為引用階層式智產庫架構的關係，在“智產資料的更新便利程度”的項目中，需花費額外運算時間

及儲存空間來更新圖形結構的特徵索引資料。然而階層式智產庫的關聯架構，也給予了檢索系統較佳的需求適應性，可藉由分層關聯權重或是關聯回饋的方式來達成應用於不同智產領域時的檢索需求。

表 4. 綜合效益評估

	基於圖形結構之智慧型 智產檢索系統 (本研究提出)	其他智產檢索系統	人工方法 (無智產檢索系統)
智產再利用度	高	低	低
智產孤島效應	不易發生	容易發生	容易發生
以解決方案(智產組合) 為基礎的檢索結果	有，以回傳完整的解決方案 (智產組合)為優先	無，皆以智產內容 為檢索標的	有，耗時費力
智產資料的更新便利程度	可	可	差
檢索精確度	高	低	極低
字串檢索效能	高，索引需求量很低	普通，需仰賴大量 已建立索引	差，人工判斷
整體檢索效率	高	低	極低
檢索需求適應性	高，可藉由權重機制來調整 階層偏重的傾向	無	低，人工界定
關聯回饋	可，經由階層畫智產庫之關 聯分析可得檢索 交集目標域	無	無
查詢問句擴展	可，經由階層化製產庫分析 可得檢索關聯圖	無	無

7. 結論及未來方向

本研究探討當前產學界中智產檢索系統尚不足之處，順應全球智產資訊快速增長的生態以及智產應用及需求層面的趨勢，希冀可將智產的再利用性最大化，並考量貼近企業需求並具有良好之運作效能，因此基於圖形結構為檢索基礎，以及多階層智產資料庫架構，並配合文字搜尋比對演算法，提出了一個具有良好效率、發揮智產可再利用性，並容許智產內容隨時增刪修改的智慧型智產檢索系統。

我們亦根據相關領域之文獻探討，歸納出了目前智產搜尋系統的不足之處，並因應提出數項創新的方法，試圖克服此領域幾項已知的問題：

- (1) 利用階層化的智產資料庫架，以及人工後設建立之智產組合關聯，構建出以搜尋智產解決方案為導向的檢索思維。

- (2) 藉由良好的人工後設智產文件關聯資訊，串連起各個智產文件的關聯屬性，因此不但強化了檢索時的符合程度，更避免掉了智產方案單獨存在的孤島效應，充分應用了現有的智產資源。

- (3) 本研究提出的實務模型，設計之初即以大量資料之比對為目標，因此亦具有良好的資料擴充性及專業適用性，可供後續實務導向應用之便。

由於本研究採用多階層之智產架構，並包含了多種類型的圖形及文字內容比對原則，因此本研究的未來方向，將會著重於圖形比對以及文字內容比對的權重調整上，以期能更符合檢索之智產庫階層架構，除了增加智產資訊本身的再利用性外，也希望能更貼近使用者的需求，提升智產檢索服務的層次。

參考文獻

- [1] B. Carlson, T. Graser, D. Mariner, J. Palof, P. Tamminga, *Targeted asset capture, identification, and management*, United States Patent, No. 7080355 B2, dated Jul. 18, 2006.
- [2] Chien Lin, Ci-Wei Lan, Jiunhau Ye and Yen-Ching Wu, *A Design of Smart Enterprise Asset Management*, The 10th IEEE International Conference on e-Business Engineering (ICEBE 2013), Coventry, United Kingdom.
- [3] D. Mukherjee, S. Mani, V. S. Sinha, R. Ananthanarayanan, B. Srivastava, P. Dhoolia, P. Chowdhury, *AHA: Asset harvester assistant*, 2010 IEEE International Conference on Services Computing, pp. 425-432. 2010.
- [4] H. He. A. Singh, *Closure-Tree: An Index Structure for Graph Queries*, Proceedings of International Conference on Data Engineering, 2006.
- [5] H. Shang, S. Zhang, X. Lin, *Taming Verification Hardness: An Efficient Algorithm for Testing Subgraph Isomorphism*, Proceedings of the VLDB Endowment, 1(1), 2008.
- [6] H. Shang, X. Lin, Y. Zhang, J. X. Yu, W. Wang, *Connected substructure similarity search*, Proceeding of the ACM SIGMOD International Conference on Management of data, 2010.
- [7] K. Gouda, M. Hassaan, *Compressed Feature-based Filtering and Verification Approach for Subgraph Search*, Proceedings of International Conference on Extending Database Technology, 2013
- [8] L. Zou, L. Chen, M. T. Özsu, *Distance-join: Pattern match query in a large graph database*, Proceedings of the VLDB Endowment, 2(1), 2009.
- [9] L. Wang, X. Ming, J. You, *The steps and methodology of identifying master data from business process*, 2009 WRI World Congress on Engineering, pp. 329-333. 2009.
- [10] Navarro, Gonzalo. *A guided tour to approximate string matching*. ACM Computing Surveys 33 (1): 31-88. doi:10.1145/375360.375365, 2001.
- [11] N. Noy. *Semantic integration: a survey of ontology-based approaches*, ACM Sigmod Record, 33(4): 65-70, 2004.
- [12] O. Levina, O. Holschke, J. Rake-Revelant, *Extracting business logic from business process models*, 2010 The 2nd IEEE International Conference on Information Management and Engineering, pp. 289-293. 2010.
- [13] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wiley, 1999.
- [14] S. Amidi, *Apparatus and method for capturing information during asset inspection in a processing or other environment*, United States Patent, No. 8059882 B2, dated Nov. 15, 2011.
- [15] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman, *Indexing by latent semantic analysis*, Journal of the American Society of Information Science, 41(6):391-407, 1990.
- [16] S. Kondo and K. Ohkushi, *Software asset systemizer*, United States Patent, No. 5361355, dated Nov. 1, 1994.
- [17] T. Ohta, Y. Tsuruoka, J. Takeuchi, et. al., *An intelligent search engine and GUI-based efficient MEDLINE search tool based on deep syntactic parsing*. In Proceedings of the COLING/ACL on Interactive presentation sessions, pp. 17-20, Association for Computational Linguistics, 2006.
- [18] T. Le, H. Elghazel, M. Hacid, *A relational-based approach for aggregated search in graph databases*. Proceedings of Database Systems for Advanced Applications, 2012.
- [19] W. Fan, J. Li, S. Ma, H. Wang, Y. Wu, *Graph homomorphism revisited for graph matching*, Proceedings of the VLDB Endowment, 3(1-2), 2010.