

東海大學資訊管理研究所

碩士學位論文

建構以資訊視覺化為基礎之

專利關鍵字分析平台

**Constructing a Patent Keywords Analysis Platform based
on Information Visualization**

指導教授：余心淳博士

研究生：黃映瑋撰

中華民國 107 年 7 月

東海大學資訊管理學系碩士學位
考試委員審定書

資訊管理學系研究所 黃映瑋 君所提之論文

建構以資訊視覺化為基礎之專利關鍵字分析平台

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人：謝明名 (簽章)
委員：余心淳
張伊婷
謝明名

中華民國 107 年 7 月 19 日

謝誌

時光飛逝，兩年的研究生生活隨著論文的完成畫下了句點。首先，我要感謝我的指導教授余心淳博士，兩年來的循循善誘，孜孜不倦地教導，犧牲了許多寶貴的私人休息時間，費心的指導及努力的督促，無論是學業上的問題，或者生活中遇到的困難，老師都耐心的給予我寶貴的意見和鼓勵，讓我在研究所求學的階段能夠更有自信及目標向前，進而去完成學業及論文。

接著，我要感謝在百忙之中抽空來參加我論文口試的委員謝昶君教授及張伊婷教授，花費許多時間與精力給予我的論文意見，並提供我許多改善的空間，讓我的論文能夠更加完整。還有，在研究所中給予我許多幫助的張伊婷教授，教學內容上讓我受益良多，並私下花了許多時間指導我關於研討會文章的部分，不論是課業上及生活上，老師都願意傾聽我的困難並給予回饋，謝謝老師兩年來的陪伴與指導。

再來，要感謝實驗室的學長姊及同學們，因為你們的陪伴，兩年的碩士生活過得十分多采多姿，在課業上及生活上也獲得你們的協助，讓我能克服許多難題，也藉著跟大家的聊天獲得心靈上很大的慰藉；其中要特別感謝建達同學，在學業上給予我很大的協助，並在思想上給我新的見解，還有私下不厭其煩的聽我抱怨，並陪伴及解決我生活中遇到的問題。

最後，我要感謝我的家人及其他朋友，因為你們的支持，我才能專心的完成研究所的學位。兩年的碩士生涯轉眼結束，也代表著我即將步入社會，開始我人生新的旅程，我會謹記在研究所生活中的所學，在未來的工作上學以致用，繼續努力，將自己的能力及所學回饋於社會。

論文名稱：建構以資訊視覺化為基礎之專利關鍵字分析平台

校所名稱：東海大學

畢業時間：2018年07月

研究生：黃映瑋

指導教授：余心淳

論文摘要：

現今的資訊量過大，常常使人無法及時判讀重要的資訊。本文將以資訊視覺化的概念呈現，秉持著「最少的油墨下呈現最豐富的資訊」和「讓閱聽者眼球覺的有趣，增強互動性」的想法，將收集來的專利數據經過平台以視覺化的方式呈現，讓人能更有有效率的理解資料所傳遞的訊息，而各國專利資料庫多以條列式的方式提供使用者作為查詢，但是傳統條列式的專利資訊陳列方法並不能讓使用者在所查詢到的專利資料中快速讀取有用的訊息，因此本論文將提出一個以視覺化為核心的專利檢索資訊平台架構，並依其流程完成實作予以實現其功能，用以輔佐使用者進行智慧電網技術相關專利的檢索與分析。在專利的檢索處理上，所提出的架構先使用文字探勘的方法協助使用者將輸入的文章進行關鍵字排名，同時使用者可輸入所選擇的關鍵字並進行全球專利資料庫的比對分析，將關鍵字對標題、摘要及全文的占比以多樣視覺化的方式呈現給使用者，讓使用者能快速瞭解資料概況，協助使用者做為搜尋專利資料的輔助工具，以期提升專利檢索資料平台的使用效率。

關鍵詞：專利分析、資訊視覺化、關鍵字分析、文字探勘、智慧電網

Title of Thesis : Constructing a Patent Keywords Analysis Platform based on
Information Visualization

Name of Institute : Tunghai University, Graduate Institute of Information Management

Graduation Time : (07/2018)

Student Name : Ying-Wei Huang

Advisor Name : Hsin-Chun Yu

Abstract :

The amount of information available today is tremendous, making it impossible to interpret important information in a timely manner. This research will present the concept of information visualization, adhering to the idea of "the least amount of information presented under the ink" and "enhancing the interaction of the viewer's eyeballs", and the collected patent data will be passed through the platform in order to assist people understand the data more efficiently. Moreover, the patent database of each country provides users as a query in a categorical manner, but the traditional method of displaying patent information does not allow users to quickly read useful information in the patent information that is queried. The paper will propose a patent-based searching information platform architecture, and implement its functions according to its process to assist users in the search and analysis of patents related to smart grid technology. In the patent retrieval process, the proposed framework first uses the method of keywords extraction to peel off the articles, and the user can input the selected keywords and perform a comparative analysis of the global patent database. The keyword-to-title, summary and full-text ratios are presented to the user in a variety of visual ways, so that users can quickly understand the data profile and assist the user as a tool for searching patent materials, and improve the efficiency of the use of patent search data platform.

Keywords: Patent Analysis, Information Visualization, Keywords Analysis, Text Mining, Smart Grid

目次

摘要	I
Abstract	II
第一章、緒論	1
第一節 研究背景與動機	1
第二節 研究目的	5
第三節 研究架構	5
第二章、文獻探討	7
第一節 專利說明	7
第二節 資訊視覺化	15
第三章、研究方法	18
第一節 系統架構	18
第二節 資料蒐集流程	20
第三節 研究工具	21
第四章、系統設計	24
第一節 資料蒐集	24
第二節 資料處理	25
第三節 關鍵字分析	27
第四節 視覺化平台建置	29
第五節 前端平台呈現	31
第五章、結論與建議	36
第一節 結論	36
第二節 未來發展與建議	37
參考文獻	38
附錄	41

表次

表 2.1.3-1 CPC 主部技術代號介紹-----	13
表 2.1.3-2 Y 主部下之次部技術代號介紹-----	15
表 2.1.3-3 合作專利分類 Y04S 技術說明-----	41



圖次

圖 1.1 我國智慧電表總體架構圖-----	4
圖 1.3 研究架構圖-----	6
圖 2.1.3-1 CPC 與 IPC 階層結構比較-----	12
圖 2.1.3-2 CPC 階層結構介紹-----	13
圖 2.2-1 視覺化過程-----	16
圖 2.2-2 鑽石價格走勢圖-----	17
圖 3.1 系統架構圖-----	19
圖 3.2-1 專利檢索-----	20
圖 3.2-2 專利資料蒐集流程-----	20
圖 3.3.3 Django 架構圖-----	23
圖 4.1.1 PHP 之網路爬蟲程式-----	24
圖 4.1.2 爬蟲之網頁資料-----	25
圖 4.2.1 Python 之資料前置處理程式-----	25
圖 4.2.2 標題與摘要之前置處理-----	26
圖 4.2.3 內文之前置處理-----	26
圖 4.3.1 RAKE 之 Python 程式碼-----	28
圖 4.4.1 Stacked Bar Chart 之 Python 程式碼-----	30
圖 4.4.2 Donut Chart 之 Python 程式碼-----	30
圖 4.5.1 Django 文件架構-----	31
圖 4.5.2 後台管理之 Admin.py 程式碼-----	32
圖 4.5.3 模型檔建立之 Model.py 程式碼-----	32
圖 4.5.4 環境建置之 Setting.py 程式碼-----	32
圖 4.5.5 Django 資料庫欄位介面-----	33
圖 4.5.6 Django 資料庫介面-----	34
圖 4.5.7 瀏覽器顯示之 view.py 方程式定義-----	34
圖 4.5.8 網頁串接之 url.py 程式碼-----	34
圖 4.5.9 視覺化平台呈現圖-1-----	35
圖 4.5.9 視覺化平台呈現圖-2-----	35

第一章、緒論

第一節. 研究背景與動機

有科學研究指出[1]，人類百分之八十的經驗來自於視覺，相較於閱讀整篇的文字的廣告，人們更傾向觀看視覺化的訊息，也更容易對視覺化的資訊留下深刻的印象。而根據全球性資訊及市場測量媒體集團尼爾森（Nielsen Holdings PLC）調查顯示[1]，當使用者在瀏覽網站時，只會觀看網站中 28%的文字；而內容設計網站 Column Five 提出[1]，有三個條件讓資訊視覺化受到喜愛：吸引力、理解力及保留程度。吸引力是指在短時間內利用視覺化內容抓住使用者目光，理解力是指幫助大腦簡化數據資料的過程，保留程度則是為使用者留下難忘懷之經驗。

大數據的特性就在於大量、多元、動態變化等，但人類的五官及大腦辨識處理的能力是有限的，常常過大的數據量導致超出人體可消化的程度。不過，人體的感官中，視覺對於處理繁雜的訊息之反應能力最為優秀，在人類生活中，時時刻刻都在透過視覺獲取各種的資訊，所以在相同的資料中，能利用視覺化圖表，將更容易理解，降低解讀的困難及複雜度[2]。在資訊爆炸的時代，「略讀」已經成為一種快速吸收資料的趨勢，視覺化的內容也讓讀者可以快速理解複雜的觀念，讓作者想傳達的主題更加明瞭。

維也納社會與經濟博物館計畫主持人奧特·聶瑞曾說過一句名言：「與其忘記精準的數字，不如記得簡化的圖像」[2]。在傳達資訊時，最重要的目的便是將資訊清楚傳達讓對方理解，然而如果將所有資訊完整的提供給對方，便可能導致對方想要瞭解的意願降低，因此，如果能將數據完整做好整理，並有效利用資料視覺化的輔助，使分析的過程能及報告呈現更加有趣，使最終呈現的視覺化圖表能清楚傳達資訊的價值，讓人充分理解其內容及主題。

而專利文件的數量相當可觀，且每年都以倍數在成長，再加上專利文件都隱含大量的專業詞彙，若沒有有效的分析方法，將會花費許多人力及時間都無法有效閱讀及理解其中資訊，此外傳統的專利分類方法，不易理解其專利間之關係，近幾年所興起的資料視覺化與資料探勘技術進行專利文件之分析研究也逐漸受到重視，透過發現其中隱含的特徵、關聯或新穎有趣的模式，以提供分析人員相關

所需的資訊[3]。面對大數據的熱潮，大家都想能充分利用數據，而不是迷失在龐大的數據資料中，為了讓資料變得具有意義，將資料進行視覺化是個重要的過程，透過資訊的降維，能讓人更容易理解資料中所潛藏知識及意涵。

目前美國專利商標局在 2015 年推出 PatentsView 的新專利資料視覺化平台 (<http://www.patentsview.org/>) [4]，利用 1976 年至 2014 年 USPTO 所公開的大量專利文件，連結發明人、機構名稱、地區和整體專利申請活動，以視覺化工具、查詢工具和彈性應用程式介面，提供廣泛的使用者查找發明人申請專利的時間、地點和申請趨勢，由此可見利用視覺化的專利檢索平台開始成為一種趨勢。

而專利為企業無形的資產，專利文件包含了許多有用的資訊，包括能瞭解新技術的發展動向與機會、評估研發的動向、探測競爭對手的動向…等，專利權使公司的研發成果受到保護，對於技術的創新、研發還有未來的開發方向皆有重大的影響[5]。

Leu[6]認為企業使用專利分析來減少研發成本、計算及分析出相關產業技術，強化研發成果，並且保護相關的智慧財產權，透過專利分析的方式更能容易的解釋出技術發展的趨勢及現況。近年來，專利分析之重要性因高科技領域的創新過程中變得複雜，導致創新研發時間週期所短，導致市場的預測很難準確，因此專利分析的價值提高，另外許多公司利用專利分析來預測技術發展的趨勢，期望透過專利分析鎖定市場趨勢，進而擁有創新的表現能使公司在市場上獲利[7] [8]。

能源是文明進步不可或缺的原動力，在近幾年來世界環境變遷及氣候異常的影響下，使能源及電力議題更加受到重視，各國除了開始致力於尋找新能源，希望能夠取代傳統能源外，也開始著手與新科技結合，想找出更多替代方案，減少資源的浪費。因此，在這樣的背景下，人類開始面對許多能源相關的議題，以潮汐、風力及太陽能等可再生能源來替代傳統的石油、天然氣等不可再生能源發電，儼然成為世界各國進行電力建設的新課題。但是利用再生能源發電具有不穩定的特性，且過去所使用的傳統電網無法應付現今的電力建設需求，所以許多先進國家提出了升級電網的計畫，並同時進行推動智慧電網 (Smart Grid) 的技術開發與設置，如果能透過智慧電網，以更好的狀態生產及配送電力，來應付各種不同的狀況，安全的控制能源的使用，將能使資源受到更有效的利用，進而節省不必要的浪費。

美國總統歐巴馬曾在 2009 年所提出的美國復甦與再投資法案(American Recovery and Reinvestment Act, ARRA)中[9]，提及將針對智慧電網及相關設備之投資額上達 45 億美元，補助對象包括智慧電表、智慧電網區域實證及儲能實證等。歐巴馬政府並將智慧電網視為綠色新政 (Green New Deal)中一環，希望透過智慧電網技術的活躍及投資的增加，能夠帶動經濟的穩定及創造更多就業機會。

歐洲則是最早跨國電網整合的地區，歐盟以能源、環境及永續發展為議題，在 1998 年到 2011 年間開了三次架構計畫(Framework Programme, FP)，分別為 FP5、FP6 及 FP7，而在這三次的架構計畫中確立歐洲將以永續發展做為歐洲科技發展之策略，並於 2005 年成立智慧電網技術平台 (European Smart Grid Technology Platform, ETP Smart Grids)，在此後還發表了歐洲未來電網構想與策略、歐洲未來電網策略性研究議程及歐洲未來電網戰略部屬文件[9]，進而確立未來歐洲智慧電網的發展方向。

日本在智慧電網的推動上則主要是在強調在利用可再生能源、確保電力系統能有效的運作，及從家庭到城市之能源規劃及管理。在 2009 年，日本召開一系列的次世代能源社會系統協議會[9]，以智慧社區 (Smart Community) 做為主要核心概念，想將現有的分散式能源進行整合管理，以在未來足夠支持居民的生活所需。

台灣的能源政策中，主要是以推動節能減碳計畫及提高再生能源之使用為重點項目，並將智慧電網納入在國家綠能低碳總行動方案的項目中，以推行智慧電表之基礎建設發展、規劃智慧電網及使用智慧電力之服務，而在 2010 年經濟部更研擬智慧電表基礎建設推動方案，發展我國智慧電網之整體規劃架構[9]，如圖 1.1 所示。

我國智慧電網系統推行主要有以下四項目標[9]：確保穩定供電、促進節能減碳、提高綠能之使用及引領低碳產業，預計在 2021 年能夠完成配電自動化的推動，使再生能源可以納入供電系統中進行管理，而發展成為區域型態的儲能系統。



圖 1.1 我國智慧電表總體架構圖

資料來源：經濟部能源局

第二節. 研究目的

經由上一節敘述，可以瞭解本研究之研究動機。在過去專利分析主要是找出相關領域進行一次性分析，一般而言判斷專利文件需要深度閱讀其內容，才可以明確判斷出該類別，因而需要耗費龐大的人力與時間成本。

本研究將利用經濟部智慧財產局所提供的國家專利資料庫之現今世界各國都重視的智慧電網專利資料並結合資訊視覺化的介面進行研究，將對所蒐集的專利資料進行資料前置處理，包含過濾資料內容，並使用文字探勘方式對資料進行斷詞處理，再進行資料統整分析及篩選，最後建置一個優化的視覺化關鍵字分析平台，將使用者所輸入的關鍵字之佔比（包含標題、摘要、內文）利用視覺化圖形方式呈現，讓使用者面對複雜且大量的專利文件資料時，能夠迅速判讀，並可以快速的選取出所需要的資料。

第三節. 研究架構

本研究架構如圖 1.3 中所視，共分為五個章節，說明如下：

1. 第一章緒論：闡述本研究之研究背景與動機、研究目的及研究架構。
2. 第二章文獻探討：閱讀相關文獻，以國內外論文、相關期刊、著作…等，加強定義主要研究項目所使用的工具與方法，本章主要介紹專利背景與發展概況、資料視覺化、專利視覺化及關鍵字分析。
3. 第三章研究方法：透過上述文獻的整理與回顧，制定研究流程，並說明專利資料檢索之方法及蒐集、平台建置之流程及方法。
4. 第四章研究結果與分析：針對本研究進行研究結果與分析，包含將所蒐集的資料以視覺化界面的呈現方式，及關鍵字分析，並加以探討結果。
5. 第五章結論與建議：針對本研究之結果加以說明及歸納，並給予後續相關研究的方向及建議。

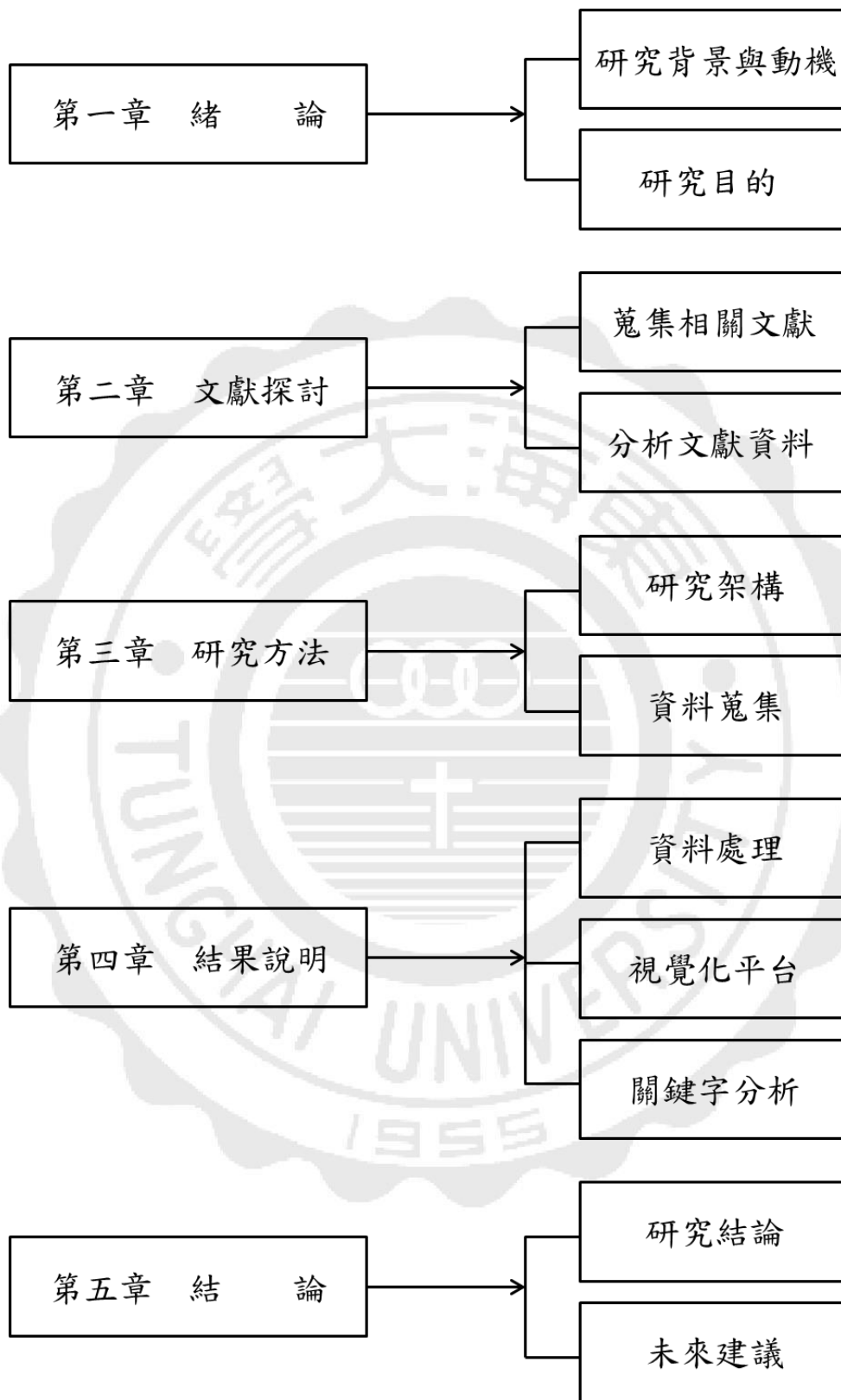


圖 1.3 研究架構圖

第二章、 文獻探討

本章節將討論專利及資訊視覺化之國內外相關研究。第一節首先，將介紹專利在各國的定義，再分別介紹世界著名的專利檢索系統，最後，針對專利中的分類及本研究中用到的專利類別進行介紹。第二節主要針對資訊視覺化的概念部份進行國內外文獻的研讀討論，再對國內外專利檢索網站之視覺化部份進行比較。

第一節 專利說明

一、 專利的定義

專利，其主要目的是在於鼓勵大眾從事發明，主要由各國政府分別設立專利制度，以保護發明人之權益，並讓專利權人與大眾之間以合法及適當的方式利用發明，由此促進產業的發展。以經濟部智慧財產局在 2014 年公布的資料中[10]，將專利定義為：發明人對其創作或發明所提出之申請，經審查符合規定以後，國家將其技術公開並給予專利權，並賦予其在一定時間內具有保護權益之作用。而所謂的具有專利權，是指在一定期限內及國家境內，賦予專利擁有者可以製造、使用、販賣或排除其他人「進口」其發明至該國家境內之行為。

目前世界各國對於專利權皆有保護的共識，但目前各國對於專利權的類型、保護年限及申請規定皆有不同，而專利大致上可分為發明專利、新型專利及新樣式專利三種。發明專利主要為保護思想的創作，著重在功能、技術、製造及使用上的便利性的改進，範圍涵蓋了無一定空間型態之物質、有一定空間型態之物品、方法、生物材料等，皆可以申請為專利。新型專利與發明專利相似，但重點主要在於物品的結構如形狀、構造或組合之創作。新樣式專利為保護對物品之形狀、花紋、色彩等透過視覺的創作，重點為表面視覺呈現的方式，與技術無關。以下將根據 2012 年眾律國際法律事務網站之專利工程師陳弘易所整理資料[11]，對世界主要專利大國之專利種類及相關規定進行簡介：

1. 美國：美國將專利分為設計專利(Utility Patent)以及新式樣專利(Design Patent)兩種。美國的設計專利相當於我國的發明專利制度，兩專利皆為自申請日或主張之優先權日起十八個月後進行公開，但國際優先權部分設計專利期限為 12 個月，而新式樣專利期限則為 6 個月，兩專利權有效期限皆為申請日起 20

年以及公告日起 14 年[11] [12]。

2. 歐洲：歐洲地區的專利協定包含歐洲專利公約（European Patent Convention, EPC）的發明專利及歐盟的歐盟智慧財產局（Office for Harmonization in the Internal Market, OHIM）新式樣專利。EPC 專利為一個共同審查平台，但並不直接授與專利權，EPO 透過專利之審查結果，可直接被締約國所接受，而無需在重新審查。EPC 之發明專利自申請日或主張之優先權日起十八個月後進行早期公開，國際優先權期限為 12 個月，專利權有效期為自申請日起 20 年[11] [13]。
3. 中國大陸：大陸將專利分為發明專利、實用新型專利及外觀設計專利三種，發明專利自申請日或主張之優先權日起十八個月後進行早期公開。國際優先權部分發明專利與實用新型專利為 12 個月，外觀設計專利為 6 個月，而發明專利之專利權有效期為自申請日起 20 年，實用新型專利之專利權有效期為自申請日起 10 年，外觀設計專利之專利權有效期為自申請日起 10 年[11] [14]。
4. 日本：日本將專利分為特許專利、實用新型專利及意匠設計專利三種，而日本的特許專利相當於我國的發明專利制度，特許專利自申請日或主張之優先權日起十八個月後進行早期公開。國際優先權部分特許專利與實用新型專利為 12 個月，意匠設計專利為 6 個月，而特許專利之專利權有效期為自申請日起 20 年，實用新型專利之專利權有效期為自申請日起 10 年，意匠設計專利之專利權有效期為自申請日起 15 年[11] [15]。
5. 南韓：南韓將專利分為發明專利、新型專利及新式樣專利三種，發明專利自申請日或主張之優先權日起十八個月後進行早期公開。國際優先權部分發明專利與新型專利為 12 個月，新式樣專利為 6 個月，而發明專利之專利權有效期為自申請日起 20 年，新型專利之專利權有效期為自申請日起 10 年，新式樣專利之專利權有效期為自申請日起 15 年[11] [16]。
6. 中華民國：我國將專利分為發明專利、新型專利及新式樣專利三種，發明專利自申請日或主張之優先權日起十八個月後進行早期公開。國際優先權部分發明專利與新型專利為 12 個月，新式樣專利為 6 個月，而發明專利之專利權有效期為自申請日起 20 年，新型專利之專利權有效期為自申請日起 10 年，新式樣專利之專利權有效期為自申請日起 12 年[11] [17]。

二、 各國專利檢索系統介紹

專利資料庫的建置對於專利之檢索、分析與研究都是十分的重要，各國或相關機構也都建立有資料庫，若能善加利用，對於科技之發展與專利的申請都是很有幫助的，以下將對世界主要的資料庫作簡介。

1. 美國專利商標局 (US Patent and Trademark Office, USPTO)

美國專利商標局主要提供 1976 年以後美國專利之全文檢索，但 1790 年至 1976 年之專利僅能以專利號及分類號作為查詢，還提供近兩年內申請美國專利之案件查詢，而資料庫為每週更新一次[12] [18]。

USPTO 資料庫的主要特色為[12] [18]：

- 提供參考文獻 (References Cited) 連結的功能，可立即連結至參考文獻中所引用和被引用的美國專利中。
- 提供專利分類號之輔助檢索的介面，以協助使用者檢索專利分類號系統。
- 免費提供美國專利授證之專利全文影像檔。
- 檢索介面簡單易懂適合所有人，並有指令式檢索供專業人士使用。

2. 歐洲專利局資料庫 (Espacenet)

歐洲專利局資料庫為歐洲專利局 EPO 所建置，以提供檢索歐洲專利局及歐洲組織成員國之許可專利文獻、PCT 專利文獻及全世界 90 多個國家的專利文獻，可說是全球最大的專利資料庫網站，並提供簡易檢索、進階檢索與合作專利分類號 CPC 檢索[13] [18]。

Espacenet 資料庫的主要特色為[13] [18]：

- 可以提供查詢及鏈結至專利家族 (Equivalent Patent)。
- 提供專利分類號之輔助檢索。
- 免費提供約 50 多國以英文撰寫的專利文獻，並且可以瀏覽約 20 多個國家的專利全文影像檔。
- 檢索介面的設計簡單易懂，另外設有輔助說明，適合所有人。

3. 中國國家知識產權局專利信息檢索系統(State Intellectual Property Office, SIPO)

中國專利信息檢索系統為中國國家知識產權局和中國專利信息中心共同開發，具有一定的權威性和可靠性，且收集了中國自 1985 年實行專利制度以來全部中國專利資訊，包含全文影像檔與法律狀態等，且亦可搜尋近 100 多個其他國家、地區與組織之專利資料，其資料庫同為每週更新[14] [18]。

SIPO 資料庫的主要特色為[14] [18]：

- 提供所有中國專利之全文影像檔。
- 提供專利之法律狀態資訊。

4. 日本特許專利資料庫 (Japan Patent Office, JPO)

日本特許廳專利檢索網站提供日本之特許專利檢索，並使用日本專用之專利分類號 FI 及 F-Terms 作為檢索介面的輔助，以提供使用者對於專利分類號系統檢索之輔助，且由於日本市場對全球企業乃是至關重要，所以日本特許廳同時也提供英文版日本特許專利檢索網站以讓國內外人士方便查詢，但英文版網站僅提供自 1976 年以來的日本公開特許（發明申請公開）之日本專利的英文摘要數據庫（Patent Abstract of Japan, PAJ）提供查詢，而 PAJ 從 1993 年 1 月起也開始包括提供法律狀態信息[15] [18]。

JPO 資料庫的主要特色為[15] [18]：

- 提供日本特許專利之全文影像檔。
- 提供日本專利分類號 FI 及 F-Terms 的檢索介面輔助，以協助非專業人士使用專利檢索系統能更加便利。

5. 南韓專利檢索系統 (Korea Intellectual Property Rights Information Service, KIPRIS)

南韓專利檢索系統之專利說明書全文主要以韓文為主，而資料庫涵蓋自 1979 年以後之專利資料包含專利書目、專利說明書及圖檔可供使用，並提供英文查詢，摘要部分則會以英文顯示，也提供使用者將相關資訊由韓文翻譯為英文之服務[16] [18]。

KIPRIS 資料庫的主要特色為[16] [18]：

- 提供將韓文之韓國專利公報及經由韓國申請 PCT 的相關資料翻譯為英文。

6. 中華民國全球專利檢索系統 (Global Patent Search System, GPSS)

全球專利檢索系統由經濟部智慧財產局所建置，主要為了使我國各產業及中小企業能夠有效快速掌握產業技術的發展趨勢。而此檢索系統整合全球五大專利局 (美、日、歐、韓、中國大陸)、WIPO 與本國之專利資料，以提供國內外使用者便於跨國檢索，並提供簡易、進階、布林、表格式等多種檢索方式搜尋，更可令使用者精準掌握所需資料，並了解產業間的技術發展狀況，為企業的專業技術研發之專利做好布局[17] [18]。

GPSS 資料庫的主要特色為[17] [18]：

- 以平行分散式運算為基礎搭配上非同步顯示，使在資料搜尋的過程中仍能同時觀看資料，且檢索不受到查詢資料筆數上限的限制。
- 系統可提供中英日韓語言搜尋，並另外提供中英文之片語查找功能。
- 提供自動繁體與簡體字轉換及整合同音異體字，讓使用者不論輸入繁體或簡體中文都能有效檢索對應之資料。
- 提供號碼檢索、布林檢索、進階檢索及表格檢索四種檢索模式，及便於使用之檢索指令。
- 提供專利之影像資料包含公開公報、公開說明書、公開全文、專利公報、公告說明書、公告全文以及新型專利技術報告。
- 提供使用者標記清單功能，令使用者勾選所需的專利資料，便於閱讀並可讓資料輸出及列印。

三、 合作專利分類與智慧電網專利技術類別介紹

為了讓使用者及各國專利局能有效及快速方便的檢索各種資料，而建立專利分類，是一種專利文件檢索的方式，以用來確定專利申請案件中的新穎程度，並加以進行評估其專利相較過去專利之進步幅度及稀有程度，目前國際間較為熟知的專利分類系統，包括國際專利分類 (International Patent Classification, IPC)、合作專利分類 (Cooperative Patent Classification, CPC)、美國專利商標局所採用的美

國專利分類 (United States Patent Classification, USPC) [12]、歐洲專利局所採用的歐洲專利分類 (European Classification, ECLA) [13]及日本專利局所採用的 FI (File Index) 分類和 F-term (File Forming Term) 分類[15]，而本研究主要資料為 CPC 分類中 Y04S 之類別的資料，因此本節將只針對 CPC 分類及 Y04S 進行說明。

1. 合作專利分類號 (CPC)

由美國專利商標局局長 David Kappos 與歐洲專利局局長 Benoît Battistelli 在 2010 年 10 月 25 日所簽屬的協議[19]，兩國之專利局將開發以歐洲專利分類系統 (ECLA) 為基礎，並納入兩局最佳分類實務的共同分類系統，而將此分類命名為「合作專利分類 (Cooperative Patent Classification, 簡稱 CPC)」，CPC 分類系統以 ECLA 為基礎，但涵蓋了美國專利分類系統 (USPC) 包含將近 25 萬個分類，並於 2013 年 1 月 2 日正式宣布啟用，而該系統為全球性的文件分類系統。

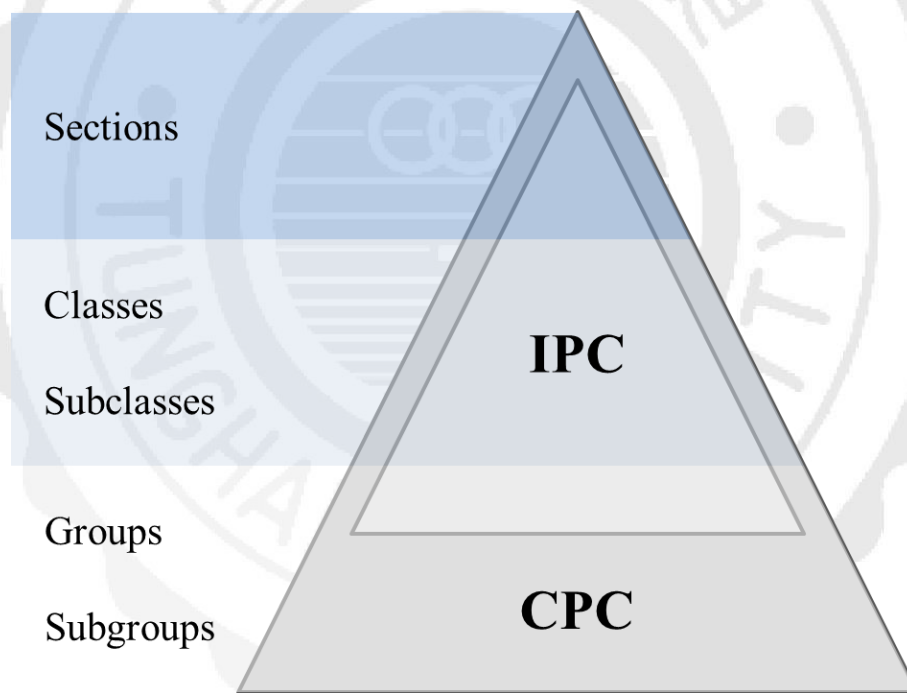


圖 2.1.3-1 CPC 與 IPC 階層結構比較

參考資料：經濟部智慧財產局，2016

SIPO 在 2013 年 6 月 3 日與 EPO 簽署備忘錄[19]，打算將 CPC 系統引進以作為 SIPO 內部的分類系統，而在接受過 EPO 專家訓練後，從 2016 年 1 月起，對專利新申請之案件開始進行 CPC 分類；KIPO 則是於 2013 年 6 月 5 日及 2014 年 6 月 3 日，分別與 USPTO 及 EPO 簽署備忘錄[19]，自 2015 年 1 月起對所有專利

新申請案件以 CPC 進行分類。世界五大專利局 (IP5) 中，僅有 JPO 仍繼續維持使用其自行開發的 FI 及 F-term 進行分類，其他成員都已陸續將 CPC 納入其專利分類系統，CPC 以有成為全球專利分類系統的趨勢，而各國專利分類制度也邁入了新的里程碑。

CPC 與 IPC 具有相同的階層結構，如附圖 2.1.3-1 所示，都是由部到目，在絕大部分的情況下，CPC 都比 IPC 更精緻化，而相較於 IPC，CPC 是具有更多更仔細的文字附加及分類表可供檢所查閱。

表 2.1.3-1 CPC 主部技術代號介紹

主部代號	主部名稱
A	人類生活必需品
B	作業操作、運輸
C	化學、冶金
D	紡織、造紙
E	固定建築物
F	機械工程學包含照明、加熱、武器、爆破
G	物理學
H	電學
Y	新興發展技術

參考資料：USPTO 網站

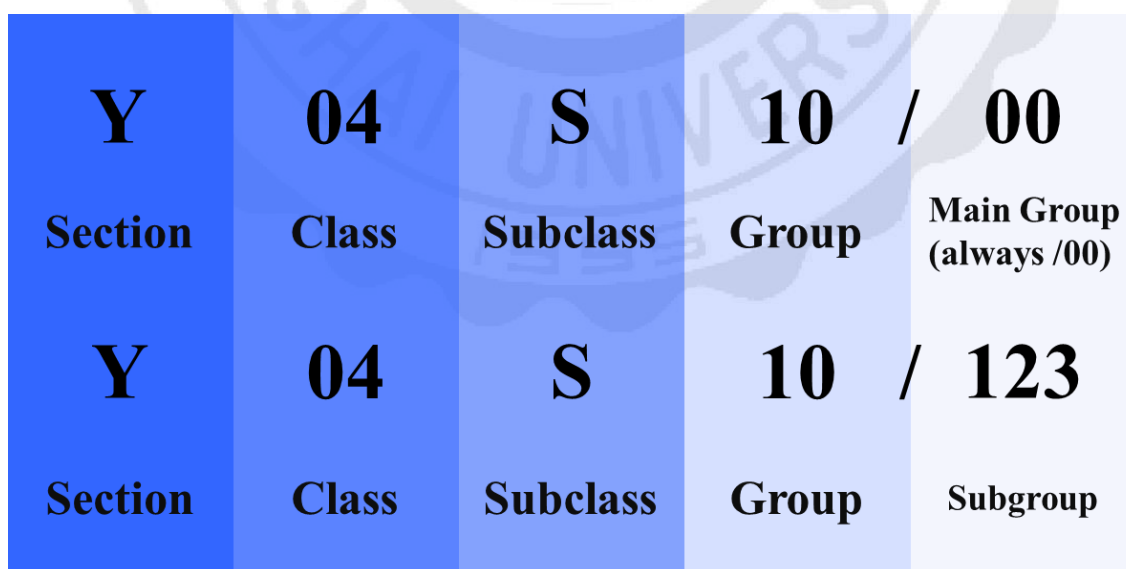


圖 2.1.3-2 CPC 階層結構介紹

資料來源：本研究自行整理

CPC 分類號主要分為五階階層結構[12]，分別為主部(Section)、主類(Class)、次類(Subclass)、主目(Main Group)及次目(Subgroup)，其中主部為英文大寫字母 A~H 與 Y，主類為兩位數字，次類為主類之後單一英文字母，主目為次類後 1~4 位數字，而次目為主目之後 2~6 位數字。其中主部英文字母所代表之技術如表 2.1.3-1 所示，A 為人類生活必需品，B 為作業操作及運輸，C 為化學及冶金，D 為紡織及造紙，E 為固定建築物，F 為機械工程學，G 為物理學，H 為電學及 Y 為新興發展技術，共計 9 大主部。如下圖 2.1.3-2 所示，以分類 Y04S10/123 為例，Y 為主部，04 為主類，S 為次類，10 為主目，123 為次目。

2. Y04S 介紹

「Y 部」為 CPC 所特有，也是索引碼的一種，Y02 分類機制原本是 2009 年 EPO 和聯合國環境署 (United Nations Environment Programme, UNEP) 為進行專利趨勢研究而訂定[12]，其作為補充 CPC 分類表中 A 到 H 部之附加資訊，內容整合 USPTO 商業方法 (Business Methods) 分類架構、交叉參照技術文獻次分類 (Cross-Reference Art Collections) 與暫時性的分類註 (Digests) 至分類表之次類中，或是用於標記 (Tagging) 橫跨多個部的新興技術 (Emerging Technologies)。目前包括 9 個次分類如表 2.1.3-2：氣候變化 (Y02A)、建築物 (Y02B)、溫室氣體捕集 (Y02C)、減少使用 ICT (Y02D)、能源 (Y02E)、工農業 (Y02P)、納入交通運輸的減排技術 (Y02T)、廢棄物 (Y02W)，再加上關於智慧電網類 (Y04S)，已涵蓋了完整的永續科技 (Sustainable Technologies) 領域。

Y04S 作為 Y 主部下面之次部，其申請的專利內容為與智慧電網 (Smart Grids) 有關技術，包含以改善電力發電、傳輸、分發、管理或使用之系統整合與電網運行、通信或信息技術相關的技術，而 Y04S 下之主目及次目之細項資料詳細情況請參考附件一，表 2.1.3-3。

表 2.1.3-2 Y 主部下之次部技術代號介紹

次部代號	次部技術
Y02A	適應氣候變化的技術
Y02B	與建築有關的氣候變化減緩技術
Y02C	溫室氣體的捕獲、儲存、分類或處置
Y02D	減少信息和通信技術中的氣候變化減緩技術[ICT]
Y02E	減少與能源發電、輸電或配電有關的溫室氣體排放
Y02P	生產或加工貨物的氣候變化減緩技術
Y02T	與運輸有關的氣候變化減緩技術
Y02W	與廢水處理或廢物管理有關的氣候變化減緩技術
Y04S	系統整合與電網運行、通信或信息技術有關的技術，以改善電力發電、傳輸、分發、管理或使用，即智慧電網 (SMART GRIDS)

參考資料：USPTO 網站

第二節 資訊視覺化

資料視覺化 (Data Visualization)，是指利用不同的統計圖表、立體模型等圖形化工具，從複雜且龐大的資料庫中萃取出有價值的資料，並將其轉換成容易理解及閱讀的訊息[20]。學者 Regine 在 2008 所發表的 Visualizing : Tracing an aesthetics of data 文章中[21]，對資料視覺化的定義為：「資料是一門交錯的科學，其利用圖像的廣大傳達功能，為科學與社會研究的進程中所發現大量抽象資訊事件，其所蘊含的意義，導致原因及之間的依存關係，提供人們一個亦於理解的解釋」。

視覺對於人類是一項重要的感官，人類能夠運用此來接收大量資訊，當眼球所視大量文字資料時，則不易接收其資料所傳達的訊息。對於視覺化的方式，有許多人將呈現方始使用各種類型的圖表呈現，也依據資料型態不同的圖表，或以顏色等多種方式呈現資料間的差異，藉由資訊展現的方式也可以藉由視覺上容易凸顯的重點來讓使用者更關心。

人類的視覺系統視有效率與細微事務的模式搜尋器。大腦跟視覺皮層組成了一個大規模的平行處理器，提供資訊到人類認知中新的高速管道，再處理高層次方面，感知與認知是彼此緊密且相關的[22]。

人是視覺性的動物，而視覺更佔據了大腦一半以上的訊息處理資源，因為人腦對於視覺的屬性有較敏感的詮釋，使用視覺表現來探索資料，使資料更具意義[23]。資料的視覺化的關鍵是針對各式各樣的資料類型，使用形狀、大小、顏色等適當的視覺屬性，透過這些方法，使其成為易閱讀、容易理解的資訊。

隨著科技的發展，資料量也跟著增加，視覺化技術便是重要的數據處理方法，資訊視覺化（Information Visualization）即將大量的資料處理後，在以人類可以較容易接受的圖形或表格，來傳達訊息[24]。而人們在面對大量資料的時候需要花費較多的時間與精力來理解，為了讓使用者能有效的進行分析與瞭解內容資料，透過圖片即可達到一張圖勝過千言萬語的效果[25]。

一個成功的視覺化能方便使用者快速掌握即吸收所傳達的訊息[26]。資訊視覺化可透過四階段來做規劃[25]，分別是數據的收集和儲存、預先處理階段、數據視覺表示及人類感知，其視覺化過程是意圖如圖 2.2-1。

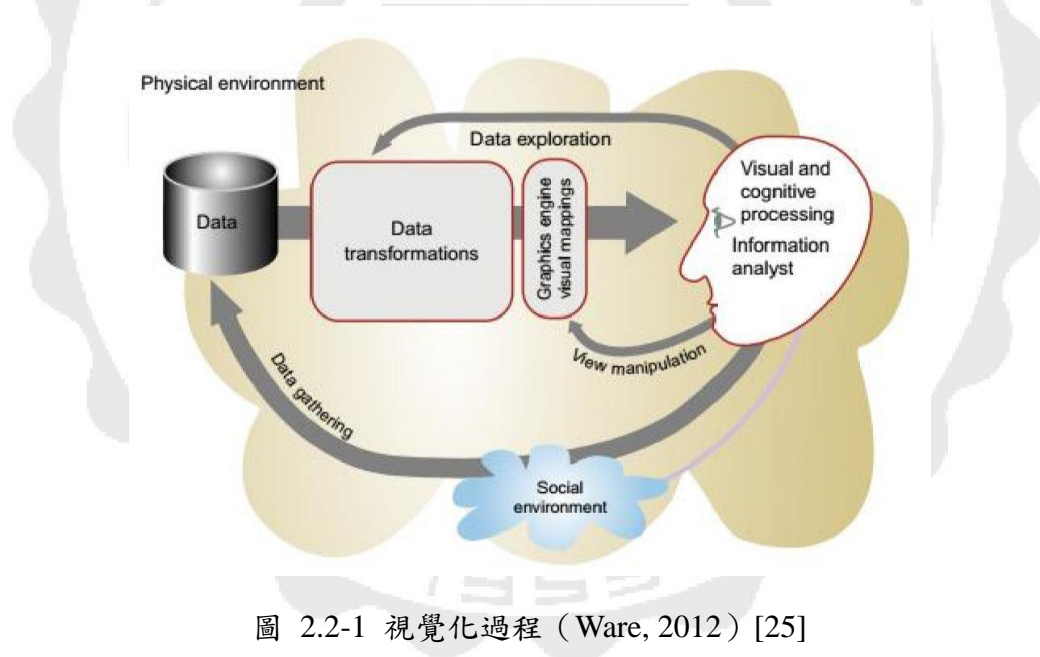


圖 2.2-1 視覺化過程（Ware, 2012）[25]

康仕仲提出要有效呈現一個高品質的資訊視覺化圖表必須考慮到以下四點[27]：

1. 深度：在資料中融入呈現專家的角度，讓大眾能透過專家的角度，更理解其資訊要呈現之內容。
2. 極簡：僅呈現有用的資訊、符號及文字，避免過多的雜訊干擾。
3. 好用：將資料用圖像表達出正確想傳遞的訊息，使的好的資料像說故事一般，

能讓人理解更深層的訊息內容。

4. 忠實：將資料的正確程度完整的呈現。

而對於資料視覺化的呈現方式，學者概略來說可分為以下兩派意見[20]：

- 敘事型視覺化：主張表現資訊的正確性，應以呈現「最詳盡資訊」為目的。
- 研究型視覺化：主張強調圖表的設計感，讓「閱聽者印象深刻」為目的。

敘事型學者中的經典代表 Tufte 教授[28]主張良好的視覺化作品應該以少量簡易的圖像去表達出最多而完整的訊息，因此他定義了一個「Data-Ink Ratio」指標來代表圖表中的資訊量及使用的油墨量的比值，而比值越高的作品代表更能有效傳遞資訊，也代表這作品是傑出的視覺化作品 [29]。而研究型的學者[22]則主張強調圖表背後所代表的故事性，能否引起讀者興趣，使之覺得有趣，並一眼就能快速明瞭圖表中所要傳達的意涵。

而資訊圖表在 80 年代正開始興起，Tufte 便引用了在 1982 年時代雜誌著名設計師 Nigel Holmes 一個圖表作品[29]（圖 2.2-2 所示），而定義了一個叫圖表垃圾（Chartjunk）的名詞，因為在 Tufte 教授提出的 Data-Ink Ratio 的角度中，這幅作品使用了過多的油墨，只為了將圖表做裝飾美化，所以他覺得 Nigel Holmes 的作品算不上是好的視覺化作品，但是，這張過多華麗裝飾的圖表，卻比最初的單純折線圖更讓人印相深刻，大家都記住了這女性的腿部曲線對應的是 80 年代初期的鑽石價格，而正是研究行學者的主張。

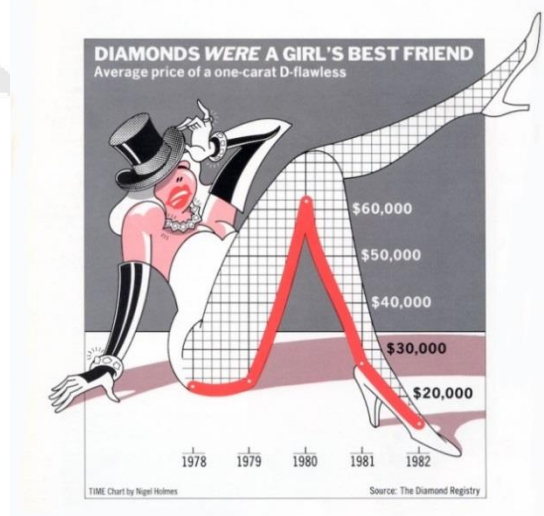


圖 2.2-2 鑽石價格走勢圖（Nigel Holmes）

第三章、 研究方法

本研究將提出建置一個輔佐使用者搜尋專利資料的關鍵字分析平台，並著重在以視覺化的方式呈現關鍵字檢索資料，以利於使用者迅速判讀檢索之資料。我們採用文字探勘技術做為關鍵字推薦的排序，再以視覺化的方式呈現關鍵字搜尋的資料概況。本章在第一節描述系統架構的部分；第二節為介紹建置系統所使用的模組及其功用簡介。

第一節 系統架構

以圖 3-1 之系統架構圖所示，整體系統建置流程共分為四大部分，分別為資料蒐集、資料處理、關鍵字分析及視覺化平台的建置及呈現。

1. 資料蒐集

以資料爬蟲的方式，將全球專利檢索系統中我們所需要的 Y04S 資料，進行蒐集。

2. 資料處理

將上述蒐集到的專利資料進行轉檔並進行前置處理，使專利說明書之內文與標題分離，並利用文字探勘的方式進行斷詞切字，方便後續之關鍵字分析使用，最後將標題、摘要及內文依照年份排序，以利後續視覺化平台建置之使用。

3. 關鍵字分析

利用前置處理後的專利說明書內文資料，以 Rake 演算法進行關鍵字提取及分析，以利使用者在鍵入關鍵字時有參考依據。

4. 視覺化平台建置及呈現

此部份主要提供視覺化結果予使用者使用，並透過介面所呈現出來之資訊，讓使用者更容易判讀專利資料之訊息。

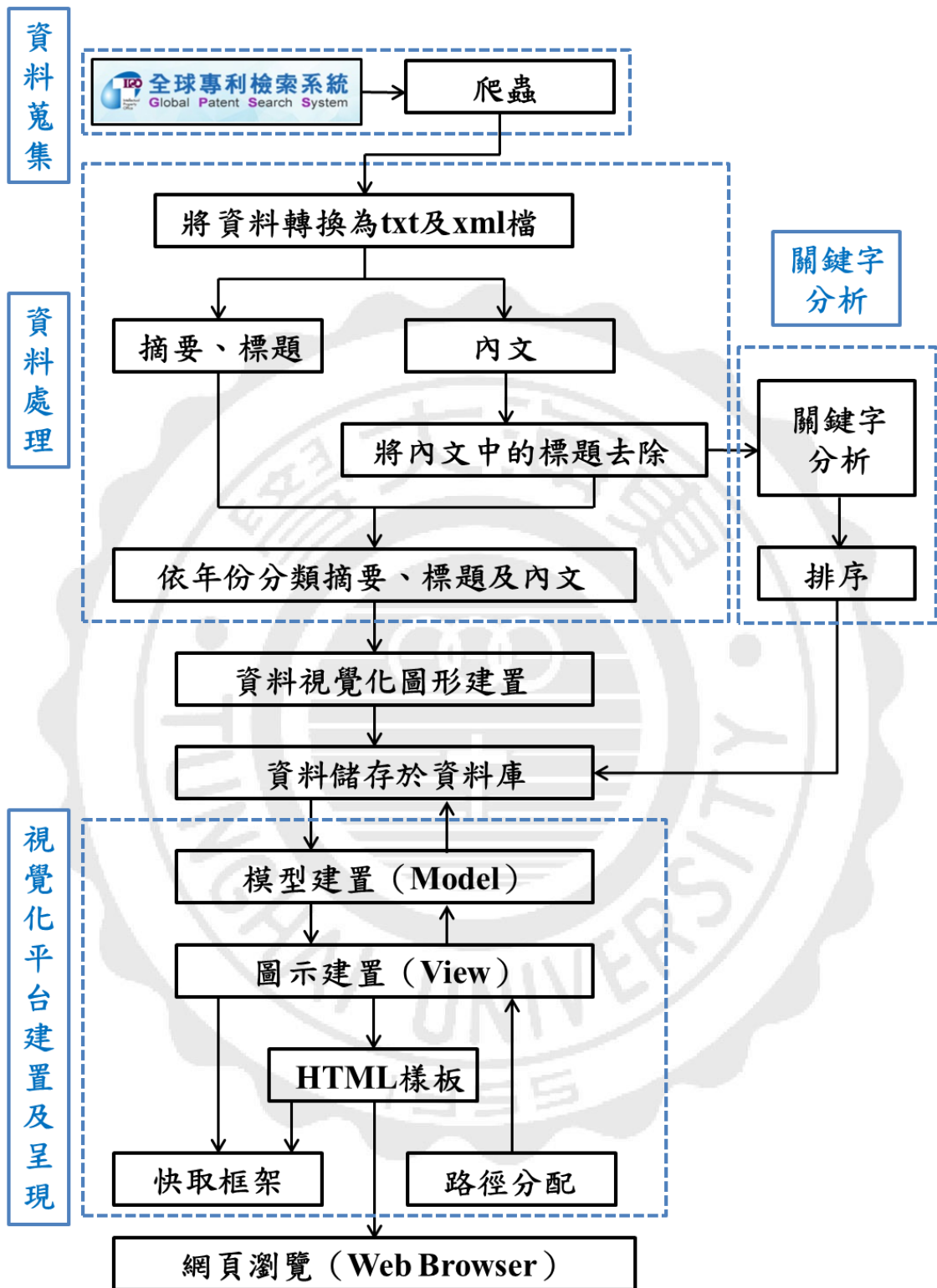


圖 3.1 系統架構圖

第二節 資料蒐集流程

本論文研究資料來自於中華民國經濟部智慧財產局全球專利系統資料庫，以進階檢索的方式，選取CPC代碼為Y04S這個類別項目下，從2007/1/1至2017/12/31，近10年來所有的資料，之後再選取資料庫中美國公開日為基準的資料共8296筆，進行爬蟲，其檢索代碼為ID=20070101:20171231 AND (CS=Y04S*)，詳細網站資訊如圖3.2-1所示。

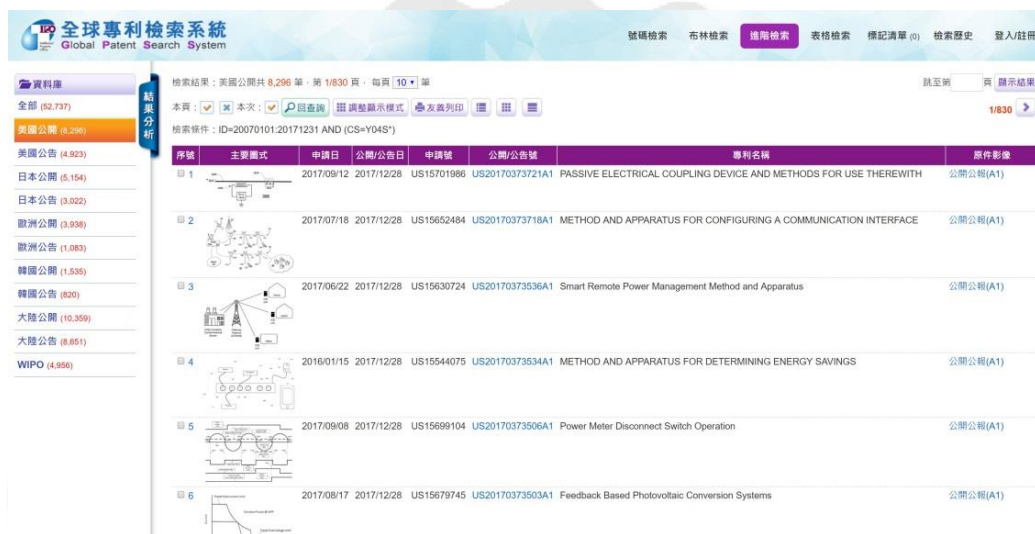


圖 3.2-1 專利檢索

資料來源：經濟部智慧財產局全球專利檢索系統

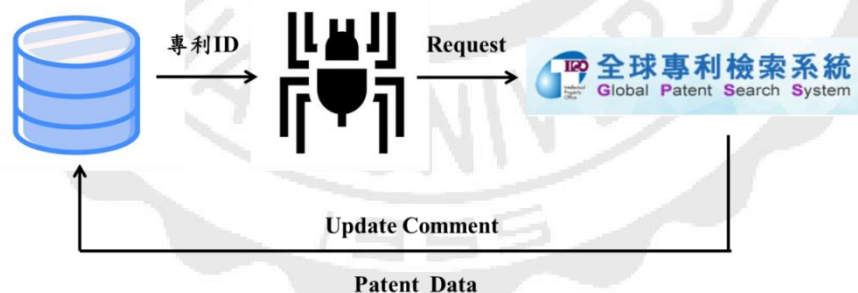


圖 3.2-2 專利資料蒐集流程

資料來源：本研究自行繪製

而在蒐集流程上，首先必須設置爬蟲 (Crawler)，抓取專利系統資料庫中專利資料儲存至資料庫，並且給予一個辨識號 (ID)，而後每天更新專利資料至資料庫，而爬蟲的設定則是使用排程方式，每一周定期抓取新資料樣本，詳細流程如圖 3.2-2 所示。

第三節 研究工具

以下將分別針對資料處理、關鍵字分析及視覺化平台的建置及呈現所使用到的工具及方法進行說明。

一、 資料處理

前置資料處理的工作，分別為針對資料進行刪除重複及內容不全之檔案、格式的轉換、斷詞及刪除常用字 (Stop Word)、詞性還原 (Stemming) 及最後將已經處理好的資料依照年份進行分類。藉著這些資料的前置處理，過濾雜訊，以便後續分析及系統建置之使用。

(一). 刪除重複及內容不全之檔案

本研究蒐集經濟部智慧財產局全球專利系統資料庫中 CPC 代碼為 Y04S 之專利資料共 8296 筆，經刪除資料內容不全及合併專利內容相同之專利文件後，最後共計 8228 筆資料。

(二). 格式的轉換

本研究經過爬蟲所抓取的檔案格式為 txt 檔，但因為專利全文在做文字探勘部分需要使用 xml 檔，所以利用 Python 進行轉檔，以便後續使用。

(三). 斷詞、刪除常用字及詞性還原

利用文字探勘的方式，將專利資料中的全文部分進行斷詞、刪除常用字及詞性還原，以下是會使用到的 Python 套件：

1. 自然語言處理工具包 (NLTK)

是 Python 的自然語言處理包，為了讓電腦能夠理解我們所使用的自然語言，而研發的套件，目的是讓電腦能夠分析、轉換集生成自然語言重要技術，分析主要是透過統計機率模式、機器學習…等方式，去分析文字，瞭解文件所傳達的意義，轉換是將文字資料轉換成另一種形式，比方說自動翻譯，而生成是指讓電腦透過應用可以自己生成是當的語句[30] [31] [32]。

2. Shutil

Shutil 是 Python 中的文件操作模塊，Shutil 提供了大量的文檔操作，特別針對文檔的拷貝與刪除，而其最主要的功能是目錄和文檔操作及壓縮操作[33]。

3. OS

OS 模組是 Python 的操作模組，用來處理像是文件和目錄這種需要手動的操作，OS 模組提供了一個統一的操作系統的接口函數，而這些函數是由平台指定的，而且能在不同的操作系統平台中之特定函數間自由切換，而實現跨平台操作[34] [35] [36]。

(四). 資料儲存

最後將前置處理好的檔案部分，依照專利標題、專利摘要及專利全文分別以各年度資料分類儲存，以便後續視覺化繪圖時能方便使用資料。

二、 關鍵字分析

以下將針對 RAKE 提取關鍵字方式進行說明。

在 2010 年由 Stuart Rose, Dave Engel, Nick Cramer and Wendy Cowley 發表的 Automatic keyword extraction from individual documents 論文中所提出的方法[37]，RAKE 算法是用來作為關鍵字 (Keyword) 提取的方法，在英文中關鍵字通常涵蓋多個單字，RAKE 法先利用標點符號將文章劃分成若干個句子，而句子之間再透過空格分化成多個單字，排除 and, the, of 等不具有意義的單字，最後提取出關鍵字。

三、 視覺化平台建置與呈現

1. Python

Python 是由 Guido van Rossum 在 1989 年所開發的新的直譯式語言[38]，直到 1991 年 2 月他將 Python 0.9 發布在 alt.sources 這個新聞論壇後，才開始廣為人知，接著在 1994 年 Python 1.0 正式發表，並且成立專屬討論區” comp.lang.python ”，奠定了 Python 日後發展的第一步。

而近年來 Python 受到許多歡迎，是因為單以一種語言就能在許多領域上應用，包含機器學習、App 撰寫、遊戲程式設計、大數據分析等，可說是一種萬用的程式語言。而讓 Python 成為萬用語言的最大原因是 Python 含有豐富的函式庫 (Library) 可供使用[39]。

但是儘管 Python 能夠使用的領域廣，但他有一個致命缺點就是執行緩慢，Python 的架構工程師為了讓 Python 在各種場域發揮得宜，做出的取舍使 Python 不

能最佳化，犧牲了效能[39]。

而本研究的後台部分運用許多 Python 的函式庫進行關鍵字前置處理及視覺化平台的建置，主要的後台架構皆由 Python 撰寫而成，而前端也是連結 Python 的網站函式庫 Django，可說是本論文所建置的系統中最重要使用工具。

2. Django

Django 是一個使用 Python 為基礎開發的網頁應用程式框架 (Framework)，是一種 MTV (Model-Template-View) 軟體架構模式[40]，在 Python 的網頁開發中被廣泛運用，因為注重組件的重用性和可插拔性，使開發複雜、資料庫驅動的網站變得更簡單架設。

如圖 3.3.3 所示，在 Django 的框架的 View 是特定網址 (URL) 的回呼函式 (Callback Function)，而回呼函式會決定哪一個資料被獲取，所以 View 描述的是"你看到的訊息"。進一步來說，Django 採用樣板 (Template) 的方法，將內容 (Content) 從展現的視圖中分離，而 View 所指定的 Template 則會描述"你如何看到、長什麼樣子"，即 View 和 Template 都是關於資料顯示的部分，Django 收到請求之後將決定適當的 View。

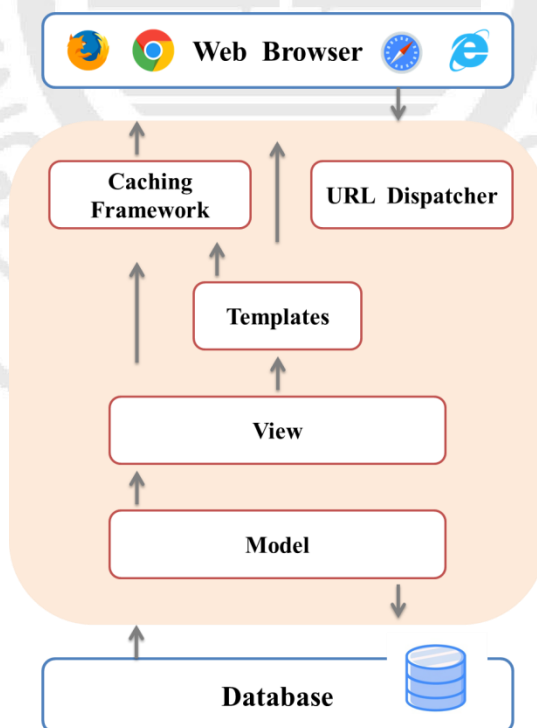


圖 3.3.3 Django 架構圖

資料來源：Django 介紹-<https://djangogirlstaipei.gitbooks.io>

第四章、系統設計

本研究設計一套專利分析比對視覺化平台，提供使用者能了解某項當前的專利趨勢，而資料視覺化的部分可以讓使用者能更有效率且清楚了解資料分布。同時也提供文章關鍵字分析，讓部分使用者可以更快速知道查檢搜尋方向，避免因為關張細節關鍵字不清楚而浪費時間大範圍查詢。本章節首先介紹資料收集，其中包含資料來源選定與爬蟲技術導入；第二部分將會說明資料收集後的處理細節與分類方式；在第三部分將會介紹平台所提供的關鍵字萃取分析；最後兩小節將會介紹前後端串接與結果畫面呈現。

第一節 資料蒐集

本研究利用全球專利檢索平台搭配爬蟲技術產生資料以供分析，將抓取專利資料存取至後台資料庫。本文使用 PHP (Hypertext Preprocessor) 作為爬蟲程式撰寫語言，目的在於抓取專利文件中的專利標題、摘要與內文等三項資料，部分程式語言之建置如圖 4.1.1 所示。而網路爬蟲程式在抓取單篇網頁內容後暫存為文字檔 (txt)，之後可以隨時轉成為需要之檔案格式，資料儲存結果如 4.1.2 所示。

```
$url1 = $linkarray[$count1];
$html1 = get_url($url1);

// echo var_dump($html1);

$dom2 = new simple_html_dom();
$dom2->load($html1);

$div1 = $dom2->find('div[@class="divsum_AB"]', 0);
$td1 = $dom2->find("td", 8);
$td2 = $dom2->find("td", 10);
$td3 = $dom2->find("td", 12);
$td4 = $dom2->find("td", 28);
$td5 = $dom2->find("td", 30);
$div2 = $dom2->find('div[@class="panel-body"]', 3);

// echo var_dump($html1);
if (!file_exists('opendata')) {
    mkdir('opendata', 0777, true);
}
$filename = "opendata\patent".$filecount;
$fp = fopen($filename, 'w');
fputs($fp, "Abstract\t".$strip_tags($div1->innertext)."\nApplied Date\t".$strip_tags($td1->innertext)."\nPublished Date\t"
.strip_tags($td2->innertext)."\nApplied Number\t".$strip_tags($td3->innertext)."\nInternational Classification Number\t"
.strip_tags($td4->innertext)."\nCooperate Classification Number\t".$strip_tags($td5->innertext)."\nDetail\t".$strip_tags($div2->innertext));
fclose($fp);

$filecount++;
$count1++;
```

圖 4.1.1 PHP 之網路爬蟲程式

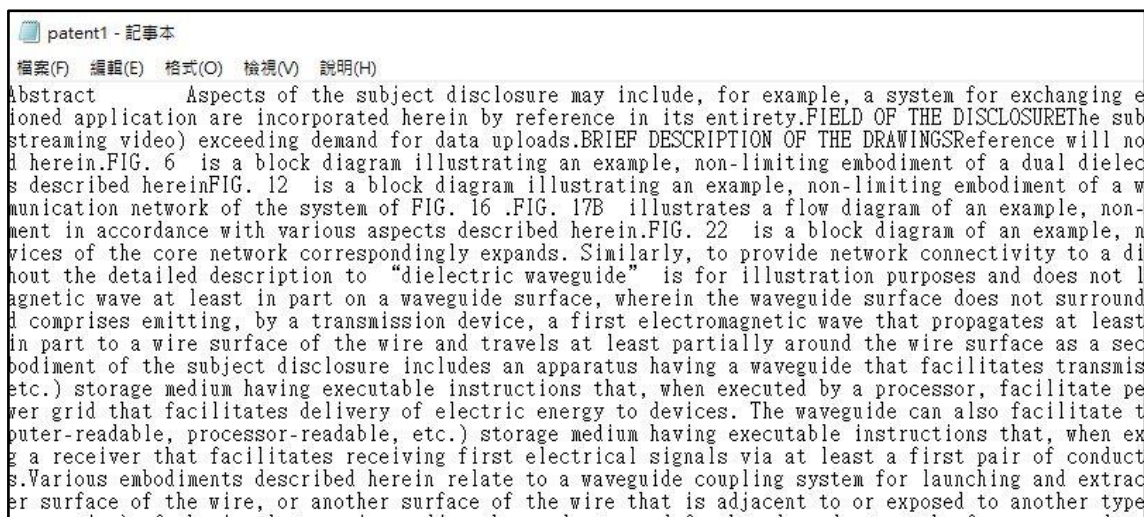


圖 4.1.2 爬蟲之網頁資料

第二節 資料處理

在資料處理部分，因為爬蟲資料結果裡包含標題、摘要與內文，在同步分析中會有效率及準確性問題。因此本文將標題與摘要轉存為 Xml 檔，另一方面撰寫程式將標題和摘要去除，使原本的文字檔案只保留內文的部分，而資料前處理是使用 Python 組成，部分程式語言之建置如圖 4.2.1 所示。

```
import os

source = "C:\Python27\patent\opendata"

for subdir, dirs, files in os.walk(source):
    count = 1
    for file in files:
        result_file = open("C:\Python27\patent\datawithnoheader\\"+file+".txt", 'w+')
        output = open("C:\Python27\patent\datawithnoheader\\"+file+".txt", "w")
        sample_file = open(source+"\\"+file)
        text = sample_file.read()
        output.write(text.split("Detail",1)[1])
        print "File:" + str(count) + "finished!"
        count +=1
```

圖 4.2.1 Python 之資料前置處理程式

在處理完標題及摘要去除與轉存後，後台分析軟體可以透過欄位分佈去分析標題與摘要部分，另一方面，關鍵字分析程式也能更準確的萃取文章重點。Xml 之標題及摘要與前處理後文字檔結果如圖 4.2.2 與圖 4.2.3 所示。

	A	B
1	title	abstract
2	METHOD AND APPARATUS FOR CONFIGURING A COMMUNICATION INTERFACE	Aspects of the subject disclosure may include, for example, a system for exchanging electrical signals and guided electromagnetic waves between customer premises equipment and service provider equipment to provide uplink and/or downlink communication services. Other embodiments are disclosed.
3	METHOD AND APPARATUS FOR DETERMINING ENERGY SAVINGS	A method for evaluation, measurement and verification of energy savings achieved by an installation of an energy saving device, such as an advanced power strip or energy hub device. The method includes an installer installing an energy saving device in a premises supplied with electricity by an energy supplier. Installation information is recorded and to a monitoring entity, usage data describing the ongoing use and power consumption of the energy saving device is then transmitted to the monitoring entity.
4	Power Meter Disconnect Switch Operation	Techniques are disclosed herein for improved power meter disconnect switch operation, which may include opening and/or closing of the disconnect switch. In particular, for reasons such as reduction of electromechanical stress on the disconnect switch, the disconnect switch may be operated based, at least in part, on a voltage at the load side of the disconnect switch. For example, in some cases, the disconnect switch may be opened slightly before a zero crossover of a waveform corresponding to the load side voltage. As another example, in some cases, the disconnect switch may be closed slightly before or slightly after a zero crossover of a waveform corresponding to the load side voltage.

圖 4.2.2 標題與摘要之前置處理



圖 4.2.3 內文之前置處理

在資料前處理完後，本文依照年份進行分類，由 2007 至 2017 和總年份共分成 12 組資料集，因本文所建置平台會呈現往年每年相關專利比對結果，讓使用者明白是否有上升或下跌趨勢。

第三節 關鍵字分析

資料探勘中的關鍵字分析與萃取技術在現今社會已行之有年，從監督式到非監督式都已有不凡且成熟的成果，而現在常見的關鍵字萃取方式可以分為 TF-IDF 與 RAKE 兩種技術，分別詳述如下：

1. TF-IDF (Term Frequency - Inverse Document Frequency)

TF-IDF 是一種用在訊息檢索與文字探勘中的加權技術，用字詞出現的頻率來評估對於文件集或資料庫中某份文件的重要程度[41]，是目前很常見的關鍵字詞頻與逆詞頻之計算方式，其中 TF (Term Frequency) 目的在衡量一個詞在某文檔中出現的頻率次數，一個文檔中多次出現的詞大多數具有特殊意義，但並非所有都具有代表性，例如常見的你、我、他，雖然出現頻率高但卻不具備任何意義，而 IDF (Inverse Document Frequency) 為逆向檔案頻率即逆詞頻，就針對在所有文檔中出現頻率過高的某一詞去做一個參數調整。TF-IDF 公式如下所示：

$$TF_{i,j} = \frac{n_{ij}}{\sum_k n_{k,j}}$$

i = 特定文件中所出現的其中一種特定單詞

j = 某一特定文件

n_{ij} : 關鍵字詞在檔案中出現的次數

$\sum_k n_{k,j}$: 檔案中所有字詞出現次數之和

$$IDF_i = \log \frac{N}{DF_i}$$

N : 文件資料庫之文件總數

DF_i : 關鍵字詞 i 出現在資料庫中的次數

最後將 TF 值乘以 IDF 值，對每個字詞都算出一個數值，就可以得到 TF-IDF，而如果特定檔案中之 TF 值為高詞彙頻率，而 IDF 值為低檔案頻率，則可得到高權

重之 TF-IDF，而 TF-IDF 傾向濾掉常見之詞彙留下特殊重要的詞彙。

$$TF\text{-}IDF = TF_{i,j} \times IDF_j$$

```
for subdir, dirs, files in os.walk(source):
    #initialize the rake_object
    rake_object = rake.Rake(stoppath)

    for current_dir in dirs:
        #open the folder and save all existing files direction in a list
        #current_dir = Y04S 10_10...those subdir name
        path = "C:\Python27\patent\testdata\\"+current_dir
        #print path

        #create subdir txt result file
        result_file = open("C:\Python27\patent\testdata\\"+current_dir+".txt", 'w+')
        output = open("C:\Python27\patent\testdata\\"+current_dir+".txt", "w")
        print current_dir
        folder_dirs = os.listdir(path)

        #rake process
        for file in folder_dirs:

            name = path+"\""+file
            del keywords_result[:]
            #keywords_result.append(name)
            print name
            sample_file = open(name)
            text = sample_file.read()
            keywords = rake_object.run(text)

            #save keywords of each file processed
            i = 0
            while i < 15:

                keywords_result.append(keywords[i][0])
                keywords_result.append(str(round(keywords[i][1],5)))
                #print keywords[0]
                i+=1

            output.write(name)
            output.write("\n")
            #print len(keywords_result)
            j = 0
            while j < len(keywords_result):
                #print keywords_result[j]
                output.write("Word: ")
                output.write(keywords_result[j])

                j+=1
                output.write("    Score: ")
                output.write(keywords_result[j])
                output.write("\n")
                #print "saved"
                j+=1
            output.write("\n")
```

圖 4.3.1 RAKE 之 Python 程式碼

2. RAKE (Rapid Automatic Keyword Extraction)

RAKE 是另一種能夠精準計算關鍵字權重的方式，除了基本的詞頻之外，RAKE 在關鍵字權重計算上還加入了候選詞選擇、特性計算和對關鍵詞評分選擇。字詞出現的順序也會影響權重大小，例如若字詞在首尾出現則代表其重要性越高。與 TFIDF 最大的不同在於 RAKE 的關鍵字不受限於一，可以是由兩字以上的字組成的詞彙。另外一方面 RAKE 的特性在於適用單篇文檔分析，在做單一文章關鍵字萃取時有很大的優勢。

總結以上特點，本文再謹慎比較後決定採用 RAKE 技術於關鍵字萃取模型上，因為在分類時可能會出現某一類別專利只存在單一文檔，若導入 TF-IDF 則會造成所有字的權重一樣，會讓使用者不知何從選擇。RAKE 特性在於能做單篇文檔分析，且在平台即時性上也佔很大優勢，不需要像 TF-IDF 中的 IDF 計算需要逐一比較某詞在資料集中所有檔案的出現頻率，所以我們最終選定以 RAKE 作為本論文中關鍵字分析之方法，而 RAKE 關鍵字萃取之開發程式碼如圖 4.3.1 所示。

第四節 視覺化平台建置

收集資料整理完後，本文採用 Python 做為後台視覺化圖像產生的程式語言，選用 Python 原因在於其圖像設計套件的多樣性與完整性。另外一方面，Python 是開放性語言，在圖像設計上的套件選擇也會較多且彈性，而本研究將導入 matplotlib 結合 numpy 等工具來呈現圖像。在圖像設計上，本文設計 Donut Chart 與 Stacked Bar Chart 來視覺化呈現資料，其中 Donut Chart 將用於關鍵字組合的專利比較，而每年的統整圖則將使用 Stacked Bar Chart。兩種設計圖的程式碼如下圖 4.4.1 與 4.4.2 所示。

```

46 def content_function_get_count (string1):
47     txt_count = 0
48     for subdir, dirs, files in os.walk(txt_source):
49         for file in files:
50             txt_output = open(txt_source+"\\"+file, "rb")
51             text = str(txt_output.read())
52             if string1 in text:
53                 txt_count+=1
54             txt_output.close()
55     return txt_count
56
57 def content_function_get_index(string1):
58     count = 0
59     ary = []
60     for subdir, dirs, files in os.walk(txt_source):
61         for file in files:
62             txt_output = open(txt_source+"\\"+file, "rb",encoding='utf-8')
63             text = txt_output.read()
64             if string1 in text:
65                 count+=1
66                 name = txt_output.name
67                 name = name.replace(subdir, "").replace("\patent","").repla
68                 ary.append(int(name))
69             txt_output.close()
70     return ary
71
72 def title_function_get_count (string1) :
73     table = xml_process()
74     result_array=[[[]]]
75     title = []
76     for i in range(table.nrows):
77         title.append(table.cell_value(i,0).lower())
78
79     title_count = 0
80     title_counter = 0
81     title_list = []
82     for j in title:
83         if string1 in j:
84             title_count +=1
85             title_list.append(title_counter)
86             title_counter +=1
87     #print title_list
88     result_array.append( title_count)
89     result_array.append(title_list)
90     return result_array
91

```

圖 4.4.1 Stacked Bar Chart 之 Python 程式碼

```

50 def make_pie(total_count, kw_count, kw, fig_index):
51
52     cm = plt.get_cmap("tab20c")
53     cout = cm(np.arange(2)*4)
54     fig, ax = plt.subplots()
55     width = 0.35
56     explode = (0,0.1)
57     pie, _, _ = ax.pie([total_count-kw_count,kw_count],
58                       radius=1,colors=cout, shadow='true', autopct = '%.1f%%', pctdistance = 0.8,explode = explode)
59     plt.setp( pie, width=width, edgecolor='white')
60     ax.text(0, 0, select_type, style='italic',size=15, ha='center', weight = 'bold')
61     cin = cm(np.array([2,6]))
62     labels = list(map("".join, ['Others',string1]))
63     pie2, _ = ax.pie([total_count-kw_count,kw_count], radius=1-width, labels=labels,
64                    labeldistance=0.5, colors=cin, shadow=True,explode = explode)
65     plt.setp( pie2, width=width, edgecolor='white')
66     plt.savefig(fig_index, transparent = True) #worked!!
67     plt.show()
68     return
69
70 for i in string_array:
71     string1 = i
72     for j in type_array:
73         select_type = j
74         if select_type == "abstract":
75             type_count = abstract_function_get_count(string1,path_xml)[1]
76         if select_type == "title":
77             type_count = title_function_get_count(string1,path_xml)[1]
78         if select_type == "content":
79             type_count = content_function_get_count(string1,txt_source)
80         #change year here!
81     fig_path = "figure\\"+years+"\\"+string1+"_"+select_type+".png"
82     make_pie(total,type_count,string1,fig_path)
83     print (j)

```

圖 4.4.2 Donut Chart 之 Python 程式碼

第五節 前端平台呈現

在關鍵字萃取和視覺化圖像生成完成後，本文將導入 Django 作為前端平台建置架構。Django 是一個使用 Python 開發 Web 的應用程式框架，Django 屬於一種大而完整的 MTV 架構，Django 平台套件會負責在合適的時機呼叫定義的方程式，圖 4.5.1 為本文平台開發之 Django 文件架構。

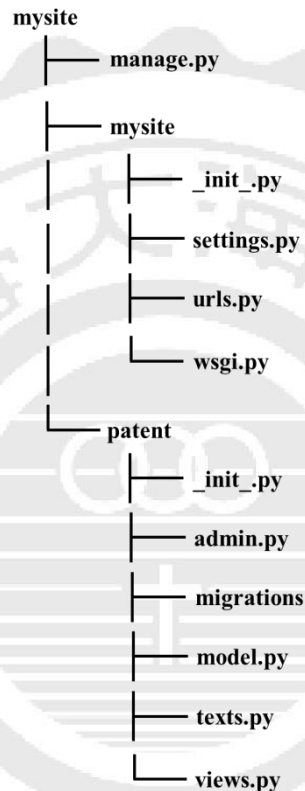


圖 4.5.1 Django 文件架構

在此架構下，可以將資料庫（SQLITE3）、後台（Python）與瀏覽器（Google Chrome）串接起來，以下將針對資料庫儲存設計與瀏覽器結果呈現做詳細介紹。

一、 資料庫儲存設計

有別於其它資料庫結構設計（如：MySQL），SQLITE3 不需要額外設定欄位名稱、屬性和值，SQLite3 的介面設計均可在 Model.py 文件中修訂切更新效率快，同時在輸入值的部分可採用介面化的方式。因此本文選用 SQLITE3 作為資料庫格式，依照研究目的將前面資料分類後的結果儲存在此資料庫。在 Django 中設計資料庫相關的文件包含以下：

1. 後台管理-Admin.py

Django 透過 Admin.py 去控管所有資料夾，只要將設計好的資料庫名稱寫進即可完成串接。另外一方面此文件會記錄使用者權限設定。在本研究中有設定一個超級使用者 (SuperUser) 帳號密碼以及一個資料庫名稱 (Post)，部分程式碼如下圖 4.5.2 所示。

```
from django.contrib import admin
from .models import Post

admin.site.register(Post)
```

圖 4.5.2 後台管理之 Admin.py 程式碼

2. 建立模型檔-Model.py

Model.py 主要是 Django 架構中用來宣告 Database Class 的文件，可以在此文件中定義或同步資料庫名稱、欄位及屬性，Django 會依據這些設定建立資料夾以供儲存，資料庫宣告程式碼如下圖 4.5.3 所示。

```
from django.db import models

# Create your models here.
class Post(models.Model):
    title = models.CharField(max_length=100)
    content = models.TextField(blank=True)
    abstract = models.TextField(blank=True)
    photo = models.URLField(blank=True)
    created_at = models.DateTimeField(auto_now=True)
```

圖 4.5.3 模型檔建立之 Model.py 程式碼

3. 環境建置-Setting.py

在 Django Setting.py 中可以設定資料庫類型以及資料庫路徑設定和名稱，相關程式碼如下圖 4.5.4 所示。

```
# Database
# https://docs.djangoproject.com/en/1.8/ref/settings/#databases

DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.sqlite3',
        'NAME': os.path.join(BASE_DIR, 'db.sqlite3'),
    }
}
```

圖 4.5.4 環境建置之 Setting.py 程式碼

4. Urls.py

此檔案在 Django 架構中主要負責用來管理後台路徑連結，將創建好的資料庫路徑串接好以供前端瀏覽器可以訪問。

在撰寫好資料庫相關的 Python 程式碼後，將透過 cmd 指令方式來執行所有相關動作包含建立、刻入欄位名稱與特徵和設定使用者權限。其指令順序如下：

```
python manage.py makemigrations >> python manage.py migrate >> python manage.py createsuperuser
```

最後創立的資料庫結果如圖 4.5.6 所示，而其中儲存資料介面如圖 4.5.5 所示。

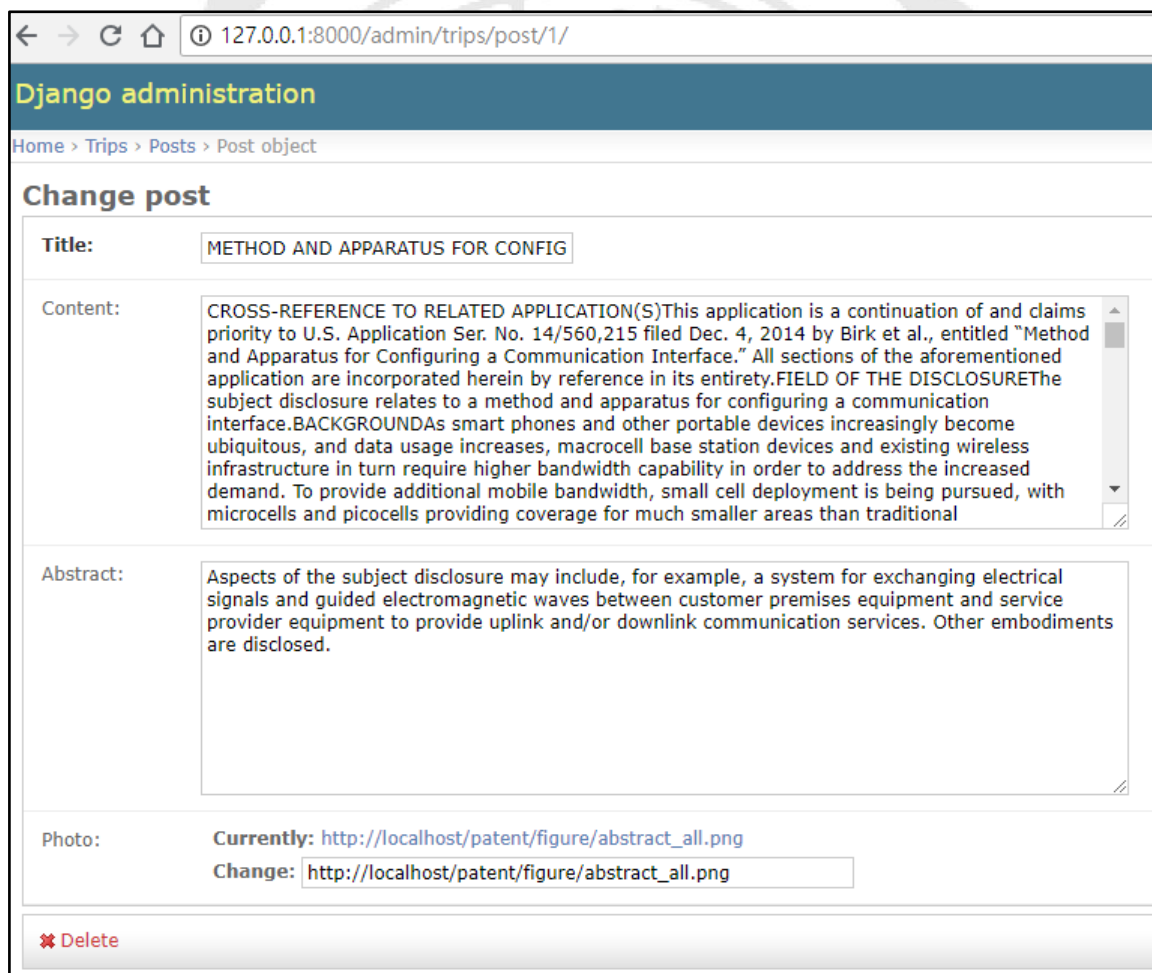


圖 4.5.5 Django 資料庫欄位介面

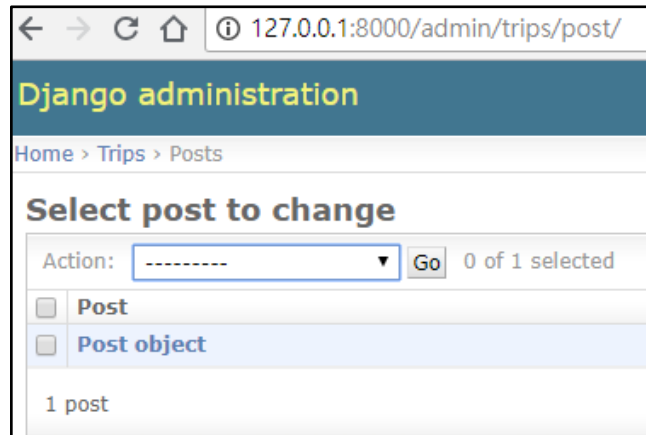


圖 4.5.6 Django 資料庫介面

二、 瀏覽器結果呈現

在完成串接資料庫與視覺化圖形製作後，Django 將透過 view.py 和 url.py 和 html 來達到瀏覽器呈現的效果。其中 url.py 負責將 html 檔案中呼叫 Python 變數的部分串接到相對應的事件或方程式中，而定義對應事件或方程式的工作將在 view.py 中完成。圖 4.5.7 為 view.py 部分程式碼。

```

614
615
616 def main123():
617     #main function call
618     if len(kw_array) == 1:
619
620         draw_one_word()
621
622     if len(kw_array) == 2:
623         draw_two_word()
624
625     if len(kw_array) == 3:
626         three_word_draw()

```

圖 4.5.7 瀏覽器顯示之 view.py 方程式定義

而在連接呼叫 url.py 的程式碼也如下圖 4.5.8 所示。

```

15 from django.conf.urls import include, url
16 from django.contrib import admin
17 from trips.views import hello_world, home, post_detail
18
19 urlpatterns = [
20     url(r'^admin/', include(admin.site.urls)),
21     url(r'^hello/$', hello_world),
22     url(r'^$', home),
23     url(r'^post/(?P<pk>\d+)/$', post_detail, name='post_detail'),
24 #     url(r'^test/$', read_text, name='test'),

```

圖 4.5.8 網頁串接之 url.py 程式碼

最後在全部部屬完成後，在瀏覽器（Google Chrome）輸入本地端的 url 即可呈現。視覺化平台呈現如圖 4.5.9 所示。

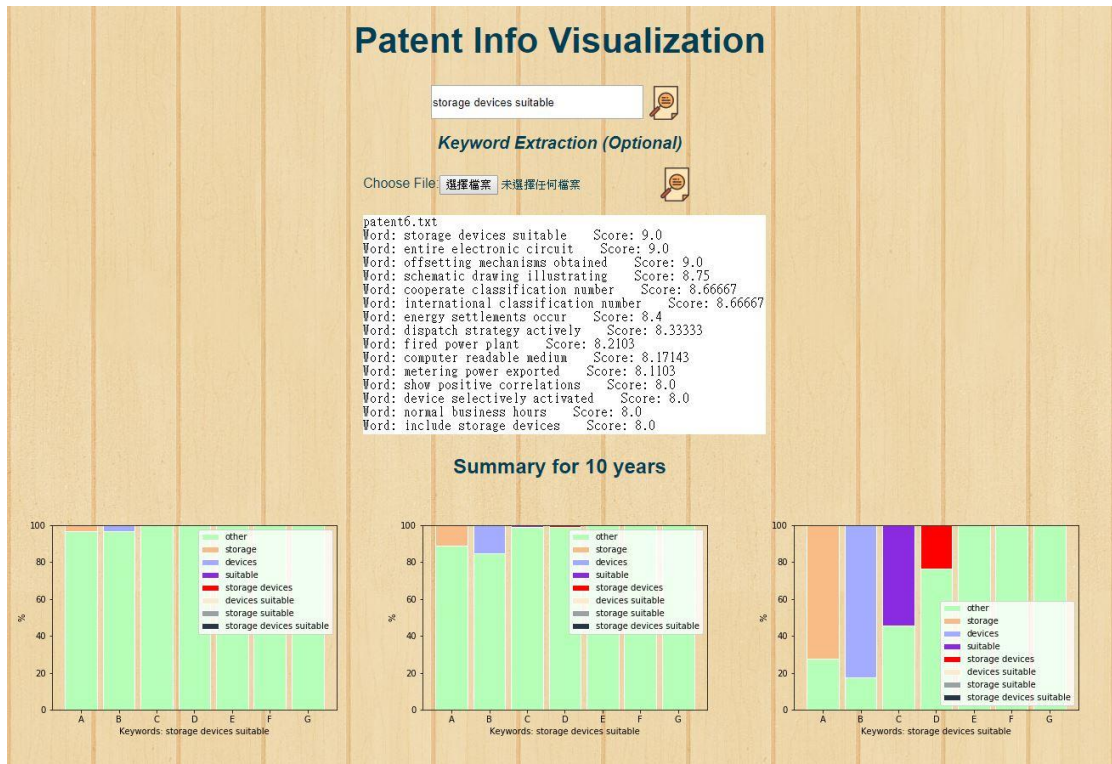


圖 4.5.9 視覺化平台呈現圖-1



圖 4.5.10 視覺化平台呈現圖-2

第五章、 結論與建議

本章節將對本論文進行總結及討論，並提出未來研究方向，第一節將整理本論文所實作之系統平台的相關議題給予結論，第二節則為未來建議將探討未來相關研究方向。

第一節 結論

我們都可以瞭解專利對一個企業來說是重要的智慧產權保障，而面對技術更新的過快，使專利資料日益龐大，難以整理判讀，而世界各國目前較多人使用的五大專利檢索平台都沒有進行視覺化的介面，在檢索的過程中皆必須逐一閱覽才能理解資訊，而逐一閱覽的方式花費太多人力資源，若能利用視覺化的方式先行呈現出所查詢的專利資料概況，提供使用者利用圖形先進行資料總體的判讀，將更有效的讓使用者瞭解資料的重點，降低閱讀資料所需花費的時間，以提升專利檢索的效率，帶動提高平台的使用率。

本論文提出建置一套以視覺化為基礎之專利關鍵字分析之平台系統，輔佐使用者在專利檢索中，能迅速的掌握專利資料的概況，提升檢索效率。而本研究所建置的系統特色為：

1. 關鍵字推薦：一般專利檢索系統普遍都有提供關鍵字檢索的部分，但如非專業人士，對於相關領域之關鍵字，並不能有效判斷，因此本研究主張提利用文字探勘的方式，讓使用者上傳專利資料全文檔案，提供關鍵字分析，並將關鍵字詞彙做排序，以提供使用者更精準的關鍵字，以利選擇所要鍵入的詞彙，避免所鍵入之詞彙無法找到所對應的專利資料。
2. 視覺化圖形：目前世界前五大檢索系統中缺乏視覺化的介面呈現資料，只有 USPTO 推出的 Patents View 用以讓使用者查找發明人申請專利的時間、地區與申請趨勢，因此本研究提出將使用者所輸入的關鍵字先以視覺化的圖形呈現，讓使用者初步了解該關鍵字在歷年之專利標題、專利摘要及專利全文出現頻率，有別於過去專利資料以條列式的方式呈現資料，讓人得逐一判讀才能吸收資訊，視覺化的圖形能夠使讀者先簡易判讀所檢索之資料概況。

本論文將所提出的視覺化關鍵字分析系統利用 Python 進行實作，雖然在資料

處理速度上尚未能有所提升，但本論文實際建置系統之結果符合當初的預期及目的，但考量到使用者需求，還有許多能夠改善及加強的部分，一個視覺化的關鍵字分析平台對於使用者來說是相當有效率的輔助工具。

第二節 未來發展及建議

本研究實作建置系統後，發下以下四個要點，值得做為未來專利視覺化關鍵字分系統平台建置時的改善方向：

1. 美化平台介面：本研究因資料分析為關鍵字對應專利標題、摘要及專利全文之占比，其所能呈現的視覺化圖形有所限制，且因為每個人對於審美的標準及概念並不相同，所以建議未來可依照需求，將介面及目前的視覺化圖形加以美化，也將其他視覺化圖形，如：泡泡圖、環狀圖、雷達圖等，結合其他專利資訊的分析應用在系統中，使系統更加完善。
2. 資料範圍：本研究主要只蒐集全球專利分析系統中美國公開部分從 2007 年至 2017 年之 CPC 分類中的 Y04S 次類資料，但這僅是專利資料中的小小一環，在未來建議可以加入其他分類系統及其他項目資料，不會受限於智慧電網之專利資料檢索時才能夠使用，使的整體系統更加完備。
3. 提升處理速度：本研究因專利數據資料量多且複雜，其中部分資料之單篇專利全文內容量過於龐大，而導致程式在執行關鍵字排序推薦及視覺化圖形呈現時，比對及繪製圖形所需要耗費的時間過於攏長，進而建議未來能更優化程式及資料庫，以降低在網頁中呈現資料分析所需要耗費的時間。
4. 系統上線：雖然目前系統已經初步建置完成，但系統要能上線使用還要面對實際的環境情況加以修改，並能夠符合使用者的期待，所以目前離上線使用還有許多方向仍要改進，期許未來能夠改善成功完成上線。

以上四點為目前所建置之視覺化關鍵字分析平台尚未完成之工作，若能從中改進，對專利視覺化分析平台將會有很大的助益。

參考文獻

- [1] 謝佳穎 (2016) , <smartM-社群內容行銷, 視覺化就是關鍵> , 線上資料 ,
<https://www.smartm.com.tw/article/32323834cea3>
- [2] 彭其捷 (2017) , <資料視覺化的奇幻之旅-探索未來的秘密武器> , 線上資料 ,
<https://findit.org.tw/newsPage.aspx?pageId=332>
- [3] Blake, Catherine. 2011, "Text Mining," pp. 123–155 in Blaise Cronin (ed.), Annual Review of Information Science and Technology, Vol. 45. Medford, NJ: Information Today.
- [4] 智慧財產權月刊 VOL.203, 2015 年 11 月, 線上資料 ,
<https://www.tipo.gov.tw/public/Attachment/5102313525691.pdf>
- [5] Gupta, V. K., "Technological Trends in the Area of Fullerenes Using Bibliometric Analysis of Patents," *Sciencometrics*, Vol. 44, No. 1, 1999, pp. 17-31.
- [6] Leu, H. J., Wu, C. C., & Lin, C. Y. 2012, "Technology exploration and forecasting of biofuels and biohydrogen energy from patent analysis." *International Journal of Hydrogen Energy*, Vol. 37, No. 20, pp. 15719-15725
- [7] Scherer, F. M., 1982, Inter-industry technology flows in the United States. *Research Policy*, 11, 227-245.
- [8] Paci, R., Sassu, A. & Usai, S., 1997, International patenting and national technological specialization. *Technovation*, 17(1), 25-38.
- [9] 經濟部能源局智慧電網總體規劃方案簡報, 2012 年 9 月, 線上資料 ,
[http://www.smartgrid.org.tw/file/%E6%99%BA%E6%85%A7%E9%9B%BB%E7%B6%B2%E7%B8%BD%E9%AB%94%E8%A6%8F%E5%8A%83%E6%96%B9%E6%A1%88\(%E7%B0%A1%E5%A0%B1\).pdf](http://www.smartgrid.org.tw/file/%E6%99%BA%E6%85%A7%E9%9B%BB%E7%B6%B2%E7%B8%BD%E9%AB%94%E8%A6%8F%E5%8A%83%E6%96%B9%E6%A1%88(%E7%B0%A1%E5%A0%B1).pdf)
- [10] 經濟部智慧財產局, 2014 年 1 月, 線上資料 ,
<https://www.tipo.gov.tw/ct.asp?xItem=502690&ctNode=7633&mp=1>
- [11] 陳弘易 (2012) , 眾律國際法律事務網站, 線上資料 ,
<http://www.zoomlaw.net/files/15-1138-12808.c1148-1.php>
- [12] 美國專利局, <http://patft.uspto.gov>
- [13] 歐洲專利局, https://worldwide.espacenet.com/advancedSearch?locale=en_EP
- [14] 大陸專利局, <http://www.sipo.gov.cn/>

- [15] 日本專利局，<https://www.j-platpat.inpit.go.jp/web/all/top/BTmTopPage>
- [16] 韓國專利局，<http://eng.kipris.or.kr/enghome/main.jsp>
- [17] 中華民國專利資訊檢索系統-經濟部智慧財產局，
<https://gpss.tipo.gov.tw/gpsskmc/gpssbkm?@@@0.08006131467753974>
- [18] 財團法人國家實驗研究院科技政策研究與資訊中心-科技產業資訊室，2003 年，
線上資料，http://iknow.stpi.narl.org.tw/Post/patentsearch.aspx?&Area_No=6
- [19] 智慧財產權月刊 VOL.217，2017 年 1 月，線上資料，
<https://www.tipo.gov.tw/public/Attachment/6123015284251.pdf>
- [20] Hsinlan Chen (2016)，<什麼是資料視覺化 (Data Visualization) ?>，線上
資料，<https://www.inside.com.tw/2016/07/29/what-is-data-visualization>
- [21] Regine., 2008, <Visualizing: tracing an aesthetics of data>, we make money not art,
Available: http://we-make-money-not-art.com/_map/
- [22] Jason Lankow et al., 2012, Infographics: The Power of Visual Storytelling., Wiley.
- [23] Iliinsky, N., 2013. Choosing Visual Properties for Successful Visualizations, New
York: IBM Corporation.
- [24] 陳欣怡 (2011)，《Web-based 虛擬實境校園資訊系統視覺化元件設計》，銘
傳大學資訊管理學系碩士論文，未出版。
- [25] Colin Ware, 2012, Information Visualization: Perception for Design, Morgan
Kaufmann
- [26] Georgios A. Pavlopoulos, Dimitris Malliarakis, Nikolas Papanikolaou , Theodosis
Theodosiou , Anton J. Enright and Ioannis Iliopoulos, 2015, Visualizing genome
and systems biology: technologies, tools, implementation techniques and trends,
past, present and future. Articles.
- [27] 康仕仲 (2014)，決策者偏頭痛-淺談資料視覺化，線上資料，
https://docs.google.com/presentation/d/1B2DFcFKTK4mCQOz9iwBJX_We9Vq4bd8CpNVCGWbNaRo/edit#slide=id.ga8c7c5ed_038
- [28] Edward R. Tufte., 1983, The Visual Display of Quantitative Information (1rd ed.),
USA : Graphics Pr
- [29] 絕對別忽略的三個視覺化關鍵要素，2015 年 4 月，線上資料，
<http://blog.infographics.tw/2015/06/three-keys-to-visualization/>

- [30] NLTK- <http://nlp.csie.ncnu.edu.tw/~shin/nlp.and.mt.html>
- [31] NLTK- <https://ithelp.ithome.com.tw/articles/10193224>
- [32] NLTK- <https://www.ibm.com/developerworks/cn/linux/l-cpnlk/index.html>
- [33] Shutil-
<https://hk.saowen.com/a/064c74926f08e9b6328f7127d26212e30a623f658966597e65fe25262616be28>
- [34] OS- <https://blog.csdn.net/SeeTheWorld518/article/details/47981925>
- [35] OS- <http://www.iplaypy.com/module/os.html>
- [36] OS- <http://pythonnote.blogspot.com/2012/09/python-os.html>
- [37] Stuart Rose. et al., Automatic keyword extraction from individual documents, Text Mining: Applications and Theory, John Wiley & Sons Ltd, 2010
- [38] Python-
<http://www.goinglearn.com.tw/programming-languages/144-python-%E5%84%A%E9%BB%9E%E8%88%87%E7%BC%BA%E9%BB%9E>
- [39] Python- <https://zh.wikipedia.org/wiki/Python>
- [40] Django- <https://sites.google.com/site/djangowebapp/django-framework>
- [41] Salton, G; McGill, M. J. (1986). Introduction to modern information retrieval. McGraw-Hill. ISBN 978-0070544840.

附錄

表 2.1.3-3 合作專利分類 Y04S 技術說明

Y04S	與電力網絡運行，通信或信息技術有關的系統集成技術，用於改進電力生成，傳輸，分配，管理或使用，即智能電網。
Y04S 10/00	支持發電，輸電或配電的系統。
Y04S 10/10	以監控，控制或運行的電網元件或設備為特徵的系統。
Y04S 10/12	這些元件或設備屬於或涉及能量發生單位，包括分散式發電[DER]或負載側發電。
Y04S 10/123	能源生產單位屬於或涉及可再生能源。
Y04S 10/126	能源生產單位是或涉及基於電力的車輛，即電動車輛[EV]或混合動力車輛[HEV]（遠程或合作收費 Y04S 30/12;與運輸部分的互操作性相關的細節，例如車輛識別、身份驗證、身份證明或賬單 Y04S 30/14）。
Y04S 10/14	該元件或設備屬於或涉及儲能單位（用於包含不間斷電源或備用發電機 Y04S 20/12 的系統）。
Y04S 10/16	元件或設備屬於或涉及電力變電站。
Y04S 10/18	元件或設備是或包括開關，繼電器或斷路器，例如：智能電子設備[IED]。
Y04S 10/20	元件或設備是或涉及保護元件，裝置或系統。
Y04S 10/22	該元件是靈活的交流輸電系統[FACTS]或功率因數或無功功率補償或校正單位。
Y04S 10/24	元件或設備是或涉及電壓調節單位。
Y04S 10/26	元件或設備是或涉及測量單位。
Y04S 10/265	元件或設備是或涉及相量測量單位[PMU]。
Y04S 10/30	以狀態監測為特徵的系統，例如 故障，溫度監測，絕緣體監測，尖端放電。
Y04S 10/40	以顯示信息為特徵的系統，例如：數據或控制。
Y04S 10/50	支持電力網絡運行之管理的系統或方法，涉及與負載端終端用戶應用程序的某種程度的交互作用（未使用）。
Y04S 10/52	停電或故障管理（未使用）。
Y04S 10/522	故障檢測或位置。
Y04S 10/525	電力恢復。
Y04S 10/527	電能質量監測。
Y04S 10/54	操作方面的管理，例如：規劃，負載或產能預測，維護，建設，擴張。
Y04S 10/545	計算有效之低碳管理或電力系統運行的方法或系統。
Y04S 10/56	供應鍊或物流。
Y04S 10/58	財務方面。
Y04S 10/60	報告、信息提供、統計或分析。

Y04S 20/00	支持終端用戶固定應用管理或操作的系統，包括配電的最後階段以及地方一級的控制，監控或操作管理系統。
Y04S 20/10	系統之特點為監控、控制或操作的終端用戶元件或設備（未使用）。
Y04S 20/12	在最後的配電階段涉及的儲能裝置，不斷電[UPS]系統或備用、應急發電機的元件或設備（涉及發電，輸電和配電的儲能裝置 Y04S 10/14；不斷電系統或備用、應急發電機作為最終用戶應用 Y04S 20/248）。
Y04S 20/14	元件或設備涉及保護元件，開關，繼電器或斷路器。
Y04S 20/16	元件或設備屬於或涉及電源插頭，插座，變壓器或電源板。
Y04S 20/18	一個或多個元件是直流電網、電網或配電線路。
Y04S 20/20	終端用戶應用程序控制系統（未使用）。
Y04S 20/22	以控制為目標的系統（未使用）。
Y04S 20/221	一般電源管理系統。
Y04S 20/222	電力需求響應系統，例如 卸載，尖峰用電調節。
Y04S 20/224	降載、電力中斷、零售電價需求響應。
Y04S 20/225	系統進入節能模式，即睡眠、低功耗或待機模式。
Y04S 20/227	智能家居技術或建築物自動化設備管理系統。
Y04S 20/228	涉及家庭自動化通信網絡為特徵。
Y04S 20/24	該系統是由終端用戶應用程序為主（未使用）。
Y04S 20/242	終端用戶應用程序是家用電器或涉及家用電器。
Y04S 20/244	家用電器正在或涉及加熱通風和空調[HVAC]單位。
Y04S 20/246	該系統涉及遠程操作燈具或照明設備。
Y04S 20/248	涉及不斷電[UPS]系統或備用/緊急發電機的最終用戶應用程序（用於不間斷電源系統或最後配電階段 Y04S 20/12 的備用或緊急發電機）。
Y04S 20/30	智慧電表。
Y04S 20/32	遠端讀表系統。
Y04S 20/322	靜態型產品。
Y04S 20/325	移動型產品。
Y04S 20/327	遠端讀表系統包括打開/關閉電源的機制。
Y04S 20/34	判定用戶行為對環境影響的系統。
Y04S 20/36	用於檢測或報告異常的方法或設備，例如：故障、停電、洩漏。
Y04S 20/38	通過分析電流或電壓波形來識別單個負載。
Y04S 20/40	顯示關於時間的使用情況，例如：監測使用情況的變化、使用情形與天氣監測。
Y04S 20/42	聯網的公用水、電表，例如：在一棟建築物內。
Y04S 20/44	顯示水電費價格或成本。
Y04S 20/46	遠端顯示儀表度數。
Y04S 20/48	判斷將網路中個主機資料分散到各主機的方法。

Y04S 20/50	改裝已安裝的儀表。
Y04S 20/52	計算生成能量或電力的系統。
Y04S 20/525	監測可再生能源發電系統，例如：太陽能電板之性能。
Y04S 30/00	交通領域特定終端用戶應用的系統。
Y04S 30/10	支持電動或混合動力車輛互用性的系統（未使用）。
Y04S 30/12	遠端或合作充電。
Y04S 30/14	與共同操作相關的細節，例如：車輛判別、認證、識別或計費。
Y04S 40/00	通信或信息技術支持電力生成、傳輸、分配或終端用戶應用程序管理的特定方面。
Y04S 40/10	具體的通信技術方面（未使用）。
Y04S 40/12	監控、控制或管理單位與監控、控制或操作電氣設備之間的傳輸結構或支持之細節。
Y04S 40/121	使用電力網絡作為傳輸的支持。
Y04S 40/122	使用脈衝信號。
Y04S 40/123	使用修改網絡功率信號的參數。
Y04S 40/124	使用數據傳輸線。
Y04S 40/125	使用電話線。
Y04S 40/126	使用無線數據傳輸。
Y04S 40/127	借助移動電話。
Y04S 40/128	使用互聯網。
Y04S 40/14	與數據或信號的處理或調節有關的方面（未使用）。
Y04S 40/143	通過專用傳輸進行通信。
Y04S 40/146	通過電網傳輸進行通信。
Y04S 40/16	監控、控制或管理單位與被監控、控制或操作的電氣設備之間的覆蓋通信網絡間之管理細節。
Y04S 40/162	與配置或供應有關。
Y04S 40/164	與網絡拓撲有關。
Y04S 40/166	與故障管理有關。
Y04S 40/168	與進行性能監控有關。
Y04S 40/20	信息技術特定方面（未使用）。
Y04S 40/22	計算機輔助設計[CAD]、模擬、建模。
Y04S 40/24	網絡安全、保護計算機或計算機系統免遭未經授權的活動之排程，例如：網絡安全。
Y04S 50/00	市場活動涉及與電力網絡運營和通信或信息技術有關的技術的系統的運作。
Y04S 50/10	能源交換，包括從終端用戶應用流向電網。
Y04S 50/12	結算、開發票、購買或出售交易或其他相關活動，例如：成本或使用評估。

Y04S 50/14	市場營銷，即市場調查和分析、測量、促銷、廣告、買方分析、客戶管理或獎勵。
------------	--------------------------------------

參考資料：USPTO 網站

