

目 錄

壹、前言	1
1.1 研究緣起及目的	1
1.2 研究方法與流程	3
貳、文獻探討	4
2.1 資料庫的基本概念與設計	4
2.1.1 資料庫的定義	4
2.1.2 資料庫管理系統	5
2.1.3 為何要使用資料庫	5
2.1.4 資料庫的系統架構	6
2.2 資料庫管理系統的發展與特性	8
參、研究設計與實施	11
3.1 資料來源分析	11
3.2 實施	14
3.2.1 實施方法	14
3.2.2 需求說明	33
3.3 架構設計	38
3.3.1 搜尋整合系統架構	38
3.3.2 資料情報蒐集系統架構	51
3.3.3 權限管控系統架構	55
3.3.4 資料分析擴充功能	56
3.4 方案可行性分析	60
3.4.1 政治可行性	60
3.4.2 行政可行性	60
3.4.3 時程可行性	60

3.4.4 經濟可行性	60
3.4.5 法律可行性	61
3.4.6 技術可行性	61
肆、預期效益	62
伍、研究結果分析	63
陸、結論與建議	65
6.1 結論	65
6.2 建議	65
參考書目	66
中文部分	66
英文部分	66

圖目錄

圖 1-1：研究流程圖.....	3
圖 3-1：實施方法流程圖.....	15
圖 3-2：平行分散式架構圖.....	18
圖 3-3：資料情報蒐集系統架構圖.....	20
圖 3-4：資料情報蒐集系統原理.....	21
圖 3-5：分類流程圖.....	25
圖 3-6：查詢輔助工具應用圖.....	28
圖 3-7：政策中心 GRB 網站搜尋系統.....	29
圖 3-8：政策中心 GRB 網站搜尋輔助工具.....	30
圖 3-9：文官制度資料庫架構概念圖	33
圖 3-10：文官制度資料庫架構核心技術關聯圖.....	36
圖 3-11：文官制度資料庫系統應用流程圖.....	37
圖 3-12：行政院施政整合檢索系統首頁畫面參考.....	39
圖 3-13：動態分類提示詞.....	39
圖 3-14：關聯性知識地圖.....	40
圖 3-15：時間序列分析.....	41
圖 3-16：自動關鍵詞示意圖.....	41
圖 3-17：以文找文示意圖.....	42
圖 3-18：資料探勘查詢結果示意圖.....	42
圖 3-19：疾管局疫情報導網站.....	52
圖 3-20：疾管局國際疫情擷取系統.....	53
圖 3-21：頁面切割設定參考.....	54
圖 3-22：擷取後資料庫產生畫面參考圖.....	55
圖 3-23：知識分類訓練流程圖.....	55

圖 3-24：個人化主題蒐集專區瀏覽參考.....	56
圖 3-25：台灣地區地理事件分析圖參考.....	58
圖 3-26：時間與事件發展關係過程參考圖.....	59

表目錄

表 1-1：資料庫系統架構.....	7
表 3-1：國內資料庫來源網站.....	11
表 3-2：搜尋引擎與分類引擎.....	27

壹、前言

文官制度乃是一個國家對其公務人員之人事政策、法規及實務運作之總稱（考試院，2009：3）。二十一世紀，全世界面臨全球化、人口高齡化與少子化、資訊科技快速發展、經濟環境連動震撼、知識經濟時代來臨等等強烈的衝擊，任何國家不可能置身事外，有為的政府必須掌握此時代之脈動，有效因應人民之需求，造福全民，提供及時的政治經濟施政計劃及服務，並同時表現出高品質、高效能（考試院，2009：3）。

為使政府達到上述機能，文官制度之興革顯屬刻不容緩，在興革之政策思維上，首先理應遵循我國憲法所明定考試用人之基本規範，站穩憲政根基；其次，當今公私部門人事管理的普世通則—功績制原則，仍是起碼的規矩工具，不容忽略；再者，近二十年來各國政府再造帶動了文官制度改革之經驗，以及不斷成長之現代化管理技術，均宜擇其可行成功之處，加以參酌學習，但如何取得良好資訊的最快速途徑，使這些資料能有效地被應用並進而連結統計分析以形成資訊系統，提供決策時重要而有策略價值的參考工具，首要工作是必須建立一個規劃完善之整合性的文官制度資料庫。

1.1 研究緣起及目的

面對全球化、資訊化及知識經濟時代的快速發展，以及政經社會環境變遷的嚴峻衝擊，各國政府為了有效推動施政願景、政策理念，及增進國家競爭力，莫不致力於提昇施政能力及文官素質與效能，而其成敗關鍵，則在於有無健全的文官制度（考試院，2009：1）。

我國公務人員文官制度乃仿效西方國家的永業化、功績制及行政中立等原則所建立，然而隨著全球化、民主化、多元化、資訊化社會政經環境的變遷。在此大的潮流趨勢下，文官制度呈現了以下趨勢：1. 文官體系人員精簡。2. 文官體系的名稱改變（成為公務體系）。3. 文官制度不斷的進行改變。4. 文官體系管理由中央集權朝向部分分權。5. 文官終身雇用制度漸行解體，由定期契約取代。6. 高級文官的管理朝向中央統一管理。7. 文官來源朝向職位公開、開放競爭。8. 薪資制度朝向彈性、績效管理。9. 文官考選朝向部會分權。（OECD，2004，轉引自彭錦鵬，國家菁英季刊第一卷第三期，2005：52）應民主化發展趨勢、憲法保障人權及提昇社會公義之目標，考試院於中華民國九十八年一月八日第十一屆第十八次會議通過施政綱領。考試院第 147 次業務會報主席裁示：籌建「我國文官制度資料庫」之可行性；考試院研究發展委員會第 44 次委員會議決議：籌建我國「文官制度資料庫」之可行性研究。故本研究之動機，乃是依據上述全球潮流趨勢，因應我國文官制度變革之需及會議決議：籌建我國「文官制度資料庫」之可行性研究，俾規劃蒐集有價值的資訊並儲存下來，提供決策時重要而有策略價值的參考。

基於以上研究動機，本文之研究目的如下：

1. 探討制度評估的定義、功能及種類與何謂方案可行性分析。
2. 說明目前制度推動之概況，並分析資料庫平台建置之原因。
3. 藉由可行性分析對建置資料庫平台進行可行性探討，以瞭解其是否確實可行，如確實可行，資料庫平台之預期效益為何？
4. 在建置資料庫平台確實可行之前提下，對資料庫平台提出建議，以供規劃及完成階段之參考。

1.2 研究方法與流程

本研究主要是採文獻分析法及系統分析法，藉由蒐集相關資料庫系統操作評估及方案可行性分析等相關論著資料，予以歸納整理，以協助進行本研究之進行，進而對資料庫系統提出建議以供考試院規劃及完成階段之參考。至於本文之研究流程如圖 1-1：

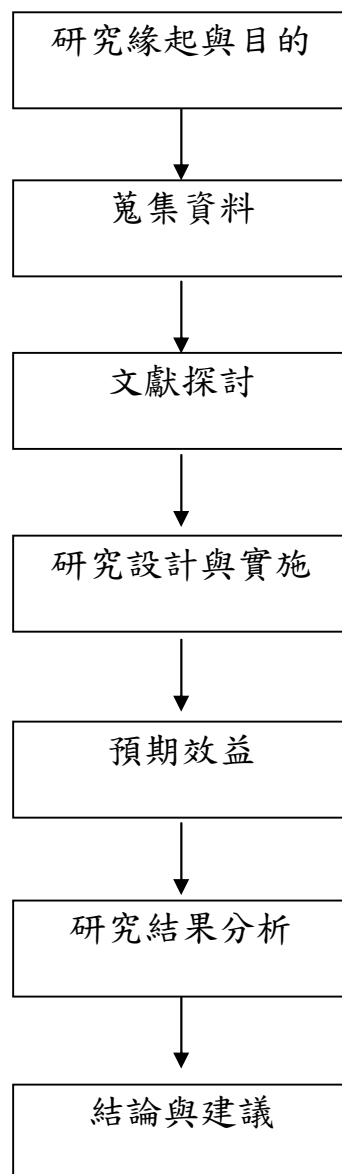


圖 1-1：研究流程圖

貳、文獻探討

2.1 資料庫的基本概念與設計

一個完整的資料庫系統包括四大部分：資料 (Data)、使用者 (User)、軟體 (Software) 與硬體 (Hardware)。

1. 資料：乃資料庫中的本體，依使用的狀況可分為「運算資料」與「異動資料」兩種。
2. 使用者：即資料庫系統主要的服務對象，包括直接使用者 (End User)、應用程式設計者 (Application Programmer) 以及資料庫管理者 (Database Administrator, DBA)。
3. 軟體：包括資料庫管理系統 (DBMS) 與應用程式 (Application Program)。
4. 硬體：包括電腦主機、磁碟機、光碟機與備份裝置等等 (鄧承根, 2006: 22)。

2.1.1 資料庫的定義

資料庫主要具有兩種意涵：一為多個資料的組合，而成為一個資料庫。而「資料」舉凡個人的生日、住址、姓名等都可以稱做是資料，進一步更可以說資料是「以一或是多個形容詞，描述的特定事物 (桂思強, 2004)」。通常會將資料整合在一個或是多個檔案中加以分類組織，以利於儲存、修改與取回相關的資訊 (Christopher J.M., Philip K.C., Gregory, 1993); 一則泛指可用以管理資料的軟體或是應用程式。而更確切來說可以稱為電子資料庫 (Electronic database)，因為每一筆資料都是以電子形式來作為儲存 (黃三益, 2005)。

而資料庫以往扮演著儲存資料的重要角色，但是從資料庫衍生出的「資料倉儲」與「資料探勘」等技術，可以讓專家從這些繁雜的「資料」中分析出重要的「資訊」來協助機關的營運發展，可以當作制定決策的依據，並且將過往的經驗轉化成有用的知識，分享給每個被授權的使用者。

2.1.2 資料庫管理系統

資料庫管理系統指的是在某個目的與需求下，由一或多個程式組成以提供使用者透過軟體來管理資料庫。一般而言，資料庫管理系統必需具備有以下的功能或是達成以下的工作目標；同樣地，有些也在 Certification Magazine (2005) 中被提出：

1. 維持資料的一致性：消除不適當的重複資料，使得各個應用的系統能夠相互的共享資料庫內的資料。
2. 查詢處理：可以經由輸入特定條件取得所需要的資料與報表。
3. 使用權限：當機關的所有資料整合在一個資料庫當中時，每一個使用者應當僅能使用其權責範圍內能夠擷取個人授權存取的資料，並且集中管理執行資料的安全管制，以維護資料的安全性與機密性。讓個別使用者可以看到整體統計的資料，而看不到個別細部的資訊；也就是讓每個人看到自己該看到的資料。
4. 備份及復原：備份指的是將資料轉存某一形式的檔案或是任何型態的資源；復原則是能夠將備份的資料重新回復至資料庫內。
5. 資料獨立性與完整性：移除程式的擷取邏輯，讓資料保持獨立性；且明確的表示各系統間與資料的關連，保持資料的完整性。
6. 異動處理：指將一連串的回存動作視為一個整體的單位，若是少了一個步驟或是其中一個步驟失敗則會回到第一步驟。
7. 資料變更：能夠去新增、刪除與修改資料庫內部的資料，並且是受到監督、稽核與可追溯的。

2.1.3 為何要使用資料庫

資料庫能夠將資訊整合過後成為能夠利用的資訊，資料庫使用的目的也就於此。而資料庫更具備有以下的優點讓我們有信心去使用資料庫：

1. 有效的基本管理：有效管理基本的人、事、物。人包含了員工與來往的客戶；物則是指產品或服務；事就是人與物互動所產生的記錄與資料。
2. 掌握即時訊息：依照各資料庫設計的目的延伸的應用，可以更快的掌握訊息。
3. 準確的決策分析：這也是建立資料庫最大的目的所在，常見的應用則是銷售排行、績效排行等等。
4. 資料共享：資料庫系統依照使用者的需求，可大可小。若是需要大到在網路上運作，則資料庫資料的共享也是此應用的一大目的。

2.1.4 資料庫的系統架構

近年來，有許多學者不斷地提出各種資料庫的系統架構，但是其基本架構還是不外乎與美國國家標準協會提出的 ANSI/SPARC 的架構一致，大致上可以將資料的描述區分成三個層次：外層（External Level）、概念層（Conceptual Level）和內層（Internal Level）。

1. 外層：指的是使用者所看到的資料庫的長相，又稱為「外部資料綱目」（External Schema）。也就是利用檢視表（view）的技巧讓每一位使用者依權責看到不同範圍的資料，無法看到全部的資料庫。
2. 概念層：將所需要的資料集中，去除重複與矛盾，正規化資料，並且使用資料庫描述語言來說明這樣的模式。也可稱之為設計概念綱目（Conceptual Schema）。
3. 內層：決定實體資料庫在磁碟中的儲存策略。如：資料要用哪種索引技術？資料要放在哪個磁碟目錄？

整理如表 1-1：

表 1-1：資料庫系統架構

美國國家標準協會 (ANSI/SPARC)	各層所代表的意義	運用的技術
外層 (External Level)	使用者所看到的外觀。 如：檢視表、資料表	檢視表、資料表
概念層 (Conceptual Level)	描述資料庫的邏輯架構。如：用資料表來呈現資料庫設計的邏輯架構。	實體關係模型 資料的正規化 資料庫的設計邏輯
內層 (Internal Level)	儲存資料庫資料與索引的實體結構。如：記錄使用者資料庫資訊的系統目錄	索引、檔案結構 資料庫實體的設計

資料來源：張真誠、蔡文輝、林敏惠（2003：1-9），轉引自田子龍（2006：5）

2.2 資料庫管理系統的發展與特性

所謂資料庫，就是針對特定目標進行資料的蒐集及組織後將其儲存在電腦內。傳統的處理方式常是每個資料庫各自獨立，而各資料庫都是因應不同的需求所設計，此種處理方式在程式設計上比較單純，但也因衍生了許多困擾和問題，例如相同的資料重複出現在不同的資料庫，不但造成記憶體空間大量浪費，若想更新某筆資料就必須更正所有包含該資料的資料庫，否則就會造成資料的不一致，因此反而需要花費更多時間、精力和成本來維護資料庫；又由於不同資料庫之間的資料往往有密切關聯，倘資料分散存放於各個不同的資料庫，卻無法予以整合，資料之存取處理便會顯得非常複雜，同時更難保證所查詢到的資料是否完整，而資料安全性也堪慮。

近幾年來，資訊的應用突飛猛進，為改進以往缺點，做到減少重複、維持資料完整性及資源共享的特性，同時實現安全性限制，而發展出資料庫管理系統 (Database Management System, 簡稱 DBMS)。DBMS 的觀念源自於 1960 年代開始，自 1970 年起相關的理論才蓬勃發展。DBMS 不僅能維護資訊及依需求擷取資訊，更重要的是能對資料庫進行控制及管理。欲達到有效率的管理，在建置資料庫之前，應該清楚地決定哪些資料是我們需要記錄的，而哪些是可以被忽略的，他們之間的關係又是如何。這個重要程序稱為資料模式化 (data modeling)，藉此產生一個資料模式 (data model)。資料模式的設計宗旨是必須將所有訊息予以整合而不分散，對使用者提供有效且完整的資訊。為了讓使用者可不經深思而很輕鬆地進行資料的查詢擷取，建立資料模式所使用的語言必須精準且簡單：精確的語言模式可避免因使用者之解釋不同而產生誤解，簡單的語言模式易懂方便傳達聯繫。

從資料庫管理系統的歷史，可依照資料的結構及操作方式，將其加以分類。

遠從資料庫之前的檔案 (files) 起，資料庫的系統模式經歷反轉串列 (inverted-list)、階層式資料庫 (relational database)。例如時下流行的 Sybase 的 SQL Server、Oracle 的 Universal Server、Infomix 的 Universal Server、Microsoft 的 SQL Server、Microsoft 的 Access 97 及 Visual FoxPro 等均屬關聯式資料庫管理系統。關聯式資料庫管理系統 (簡稱 RDBMS) 的基本概念為闡述資料間的關係以關聯式的方式來組織，並且應用程式不應與資料的內部儲存結構有任何相依關係 (此集資料獨立的概念—應用程式不依賴於任何一種特定的儲存結構或是存取策略)。在此模式下，應用程式不會因為將來資料內部結構的修改而有所牽連，而會因應新的儲存結構而做大幅度的改變，因此，應用程式的維護較為簡單。關聯式資料庫的基本結構形式為：

1. 使用者所看到的都是資料表 (table) 形式，整個資料庫分數個資料表來儲存資料，這些資料稱之為「關聯資料表」，每個關聯資料表由一筆或多筆紀錄 (record) 所集合而成，而每筆紀錄可包含一至多個資料項目 (item，相當於資料欄位)。
2. 各資料表之間存在某些限制條件，以維持資料完整性。
3. 使用者可使用的運算如資料新增、刪除、擷取，都是從舊有的資料表中產生新的資料表。

關聯式資料庫雖具有有效整合文數字資料，形成單純的表格及簡易的查詢語言 (SQL，唯一種世界標準的結構查詢語言—structure query language) 的能力，但無法滿足圖形、影像、多媒體等各項資料型態的需求。最近由於物件導向技術 (object-oriented program，簡稱 OOP) 的開發，提供解決這項難題。所謂物件導向，是將資料的基本概念抽象化，視資料處理的對象為一個物件 (object)，並允許使用者自定資料型態，除了文數字型態，包括圖形、程序、及函數等都可依附在欄位或變數上，也可定義其他類別，因此豐沛了可用的型

態。在資料庫管理系統內所具備名稱的個體皆可被視為一個物件，如資料表、查詢、表單、報表、巨集、模組、視窗、控制按鈕等。根據物件導向程式設計原則所建立的關聯式資料庫，是目前資料庫管理系統的主流。

參、研究設計與實施

3.1 資料來源分析

文官制度資料庫的建立乃是為提昇公務人員行政效率、強化知識管理，在面臨有關人事政策之決議事項時，能夠適時提供關鍵資訊做為決策分析，因此所需涵蓋的的資料庫層面必須廣闊，主要整合及蒐集有價值的資料庫，重新分析組織整理，成為一個知識轉換、分享、創造、加值及利用之知識庫平台，藉由本資料庫的建立可提昇智慧資本，創造有利組織學習的環境，以強化決策過程的幕僚作業能力，並針對所需的資訊情報加以整合與運用，進而提昇決策及行政的效率和品質。

從我國文官制度現況分析，文官制度的主要內容包括：人事、考選、銓敘、保障、培訓、撫卹、退休等，並且各項內容皆有訂定制度及法規做為依據標準，因此，文官制度資料庫整體資料來源主要以文官制度內容為需求，大致上作成或取得資料涵蓋包括：人事、法規、期刊、論文(研究報告)、新聞、公報、全球資訊網、公文系統、立委質答詢、院會議事、知識管理等，資料來源取得方式以現有公部門資料庫整合並透過情報蒐集系統針對國內外特定網站及特定資料下載整合。

資料來源及取得資料庫分析參考如表 3-1：

表 3-1：國內資料庫來源網站

類別	資料庫來源網站名稱	機關	備註
人事	<ul style="list-style-type: none">■ 行政院暨所屬各機關職員錄編印系統■ 公務人力資料庫	人事行政局	此部份資料庫取得可透過公務人力資料庫下載作業，再導入本系

			統架構資料庫作為整合查詢
法規	<ul style="list-style-type: none"> ■ 全國人事法規釋例資料庫 ■ 人事行政局最新法規 ■ 全國法規資料庫 ■ 銓敘部法規輯要檢索系統 ■ 公務人員保訓會法規輯要 ■ 考選部法規輯要檢索系統 ■ 退撫基金法規彙編及釋例 	<p>考試院</p> <p>人事行政局</p> <p>法務部</p> <p>銓敘部</p> <p>保訓會</p> <p>考選部</p> <p>退撫基金</p>	可透過搜尋引擎文字探勘技術，針對該資料庫網站建置索引整合查詢
期刊	<ul style="list-style-type: none"> ■ 考試院網站(出版品) ■ 考選部網站(出版品) ■ 銓敘部網站(公務人員月刊) ■ 人事行政局網站(研究與發展成果) ■ 保訓會網站(文件典藏) ■ 文官培訓所網站(下載專區) ■ 國家圖書館 	<p>考試院</p> <p>考選部</p> <p>銓敘部</p> <p>人事行政局</p> <p>保訓會</p> <p>文官培訓所</p> <p>國家圖書館</p>	可透過搜尋引擎文字探勘技術，針對該資料庫網站建置索引整合查詢，除國家圖書館只能查詢書目及書名。
論文 研究報告	<ul style="list-style-type: none"> ■ 考試院網站(考試院圖書館) ■ 考選部網站(出版品) ■ 銓敘部網站(論文與研究報告) ■ 人事行政局網站(研究與發展成果) ■ 保訓會網站(文件典藏) ■ 文官培訓所網站(下載專區) ■ 退撫基金網站(研究報告) 	<p>考試院</p> <p>考選部</p> <p>銓敘部</p> <p>人事行政局</p> <p>保訓會</p> <p>文官培訓所</p> <p>退撫基金</p>	可透過搜尋引擎文字探勘技術，針對該資料庫網站建置索引整合查詢，除國家圖書館只能查詢書目及書名。

	■ 國家圖書館	國家圖書館	
新聞	■ 國內各大報網站及電子媒體網站		可鎖定主要網站相關資料、透過資料情報蒐集系統自動下載資料，此部分可能需要付費或取得授權。
公報	■ 政府公報資訊網	國家圖書館	可透過搜尋引擎文字探勘技術，針對該資料庫網站建置索引整合查詢。
內部資料	<ul style="list-style-type: none"> ■ 公文系統 ■ 立委質答詢系統 ■ 院會議事系統 ■ 知識管理系統 	考試院	僅針對考試院內部非對外公開資料庫管理系統。

資料來源：本研究整理

3.2 實施

文官制度資料庫之架構建立，主要透過資訊技術將資料整合並透過資訊處理方式轉化成為知識源，知識的來源非常廣泛，並以不同的方式存在，以目前網際網路(Internet)的廣泛應用及發達的情況下，網際網路(Internet)自然成為一個最大的知識供應源，並且超乎目前常用的資料庫許多。因此如何有效的擷取有價值的知識並整合目前公部門資料庫網站知識庫，並轉化成為文官制度資料庫，為本研究主要建立目標。

3.2.1 實施方法

本資料庫建置實施方法主要可分成四個階層包括：1. 資料來源、2. 資料分類、3. 資料查詢、4. 資料分析。如圖 3-1：

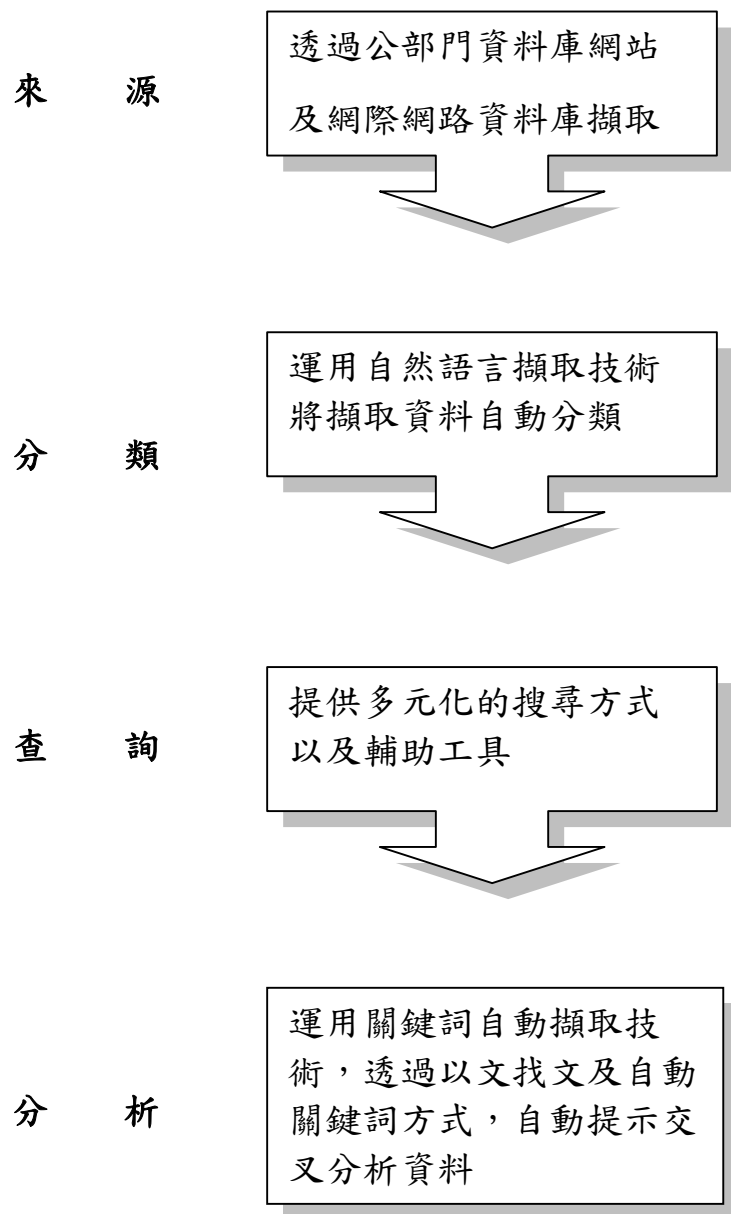


圖 3-1：實施方法流程圖

3.2.1.1 資料來源

1993 年管理大師彼得杜拉克在「資本主義社會」一書中，強調「知識」已改變傳統之生產要素，而且已經成為支配資源、維持未來競爭優勢的主要元素。而如何運用知識提升競爭力及開創事業，甚至創造經濟繁榮，是二十一世紀「知識工作者」最重要的課題。

「知識工作者」所欠缺的，不是資訊；而是欠缺時間來將資訊進行整理、分類、分析及行動。過去傳統的知識管理概念都是透過建立一個管理平台，提供「知識工作者」透過平台將所要的資料進行整理刊登，此方式必須耗費工時及人力，並不符合二十一世紀的知識時代，其實我們所要的知識資訊已經在我們日常所建立的文件、檔案、資料庫、電子郵件、網站、網際網路之中，如何透過資訊技術將現有眾多的資料源整合並轉化成為知識庫，才是符合經濟價值的知識管理。

依據憲法規定：考試院為國家最高考試機關，掌理考試、公務人員之銓敘、保障、撫卹、退休、公務人員任免、考績、級俸、陞遷、褒獎之法制事項。考試院暨所屬機關分別已經建立了許多文官制度所需的資料庫，如何將目前這些資料庫整合並轉化成為文官制度資料庫，為本資料庫所建立之資料來源方法。

除了現有制度所建立資料庫之外，另外需要考量其他資料源的蒐集，透過其他資料源的蒐集並整合現有制度資料庫，進而擴大知識經濟的價值。

資料來源方法主要可分成「資料整合」以及「資料蒐集」二個階層，茲分別就這二個階層作分析說明。

3.2.1.1.1 資料整合

本資料庫整合研究方法主要運用搜尋引擎原理，所謂搜尋引擎原理並非是一般人所認知的透過入口網站，輸入關鍵字。搜尋引擎自動會到網際網路中的網站去搜尋比對關鍵字進而查詢到所要查詢的資料，例如：如果你到 Google 或是 Yahoo 的網站中去搜尋，你會發現 Google 跟 Yahoo 所查詢的結果資料數量會不同，並且也會發現並非所有在網路上的網站資料都可以查詢的到，最主要的原因在於搜尋引擎的原理是透過軟體機器人，也就是建立索引的方式來產生自己的索引查詢資料庫，例如：Google 每隔十天都會透過自己的軟體機器人連結全世界的網站已經登錄在 Google 裏面的網址，並將這些網站中的文字擷取下來並且儲存在自己的電腦當中，透過索引的建立並產生關聯性，因此當我們在 Google 中輸入關鍵字查詢，實際上我們是在 Google 的電腦主機中查詢他已經建立的索引資料庫，再透過索引資料庫的關鍵字比對，找到他的關聯性網站。

因此，本資料庫的採用整合方法就是透過搜尋引擎中的資料索引建立方式，針對目前所要整合的相關網站資料庫建立索引，再將建立索引的資料庫網站擷取該網站資料，將擷取下來的網頁資料從非結構資料轉化為結構資料，並可儲存至文官制度資料庫中，透過搜尋方式提供資料查詢。

建立索引可分成二種方式，一種是直接對網站網址 <http://www.xxx.xxx.xx> 建立遠端索引，並且可設定索引建立範圍，例如：全網站或是某一頁網頁等，另外一種方式就是針對資料庫主機直接建立索引，並採「平行分散式架構」(Distributed Architecture)可整合不同區域網站主機資料庫整合查詢。

■ 「平行分散式架構」 (Distributed Architecture)

一般搜尋引擎提供二個模組，分別的功能為「Indexer(建索引)」、「Searcher(查索引)」。「Indexer(建索引)」模組，可以放在檔案伺服器 (File Server) 或資料庫伺服器(DB Server)；而「Searcher(查索引)」模組可以放在應用伺服器(AP Server)或 Portal Server，以達到平行分散式之目的。如圖 3-2，位於台北的之使用者，只需針對「台北 1 Server」進行檢索，此時該系統代理使用者至其他的伺服器查詢，同時合併所有伺服器的結果，回饋給使用者；同時，高雄之使用者也可以得到同樣的服務。

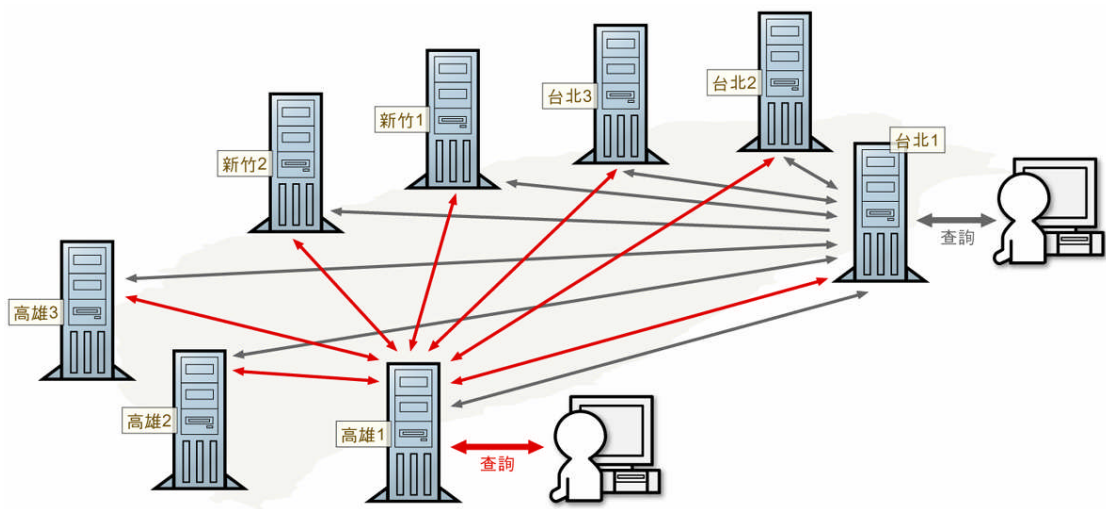


圖 3-2：平行分散式架構圖

因此，平行分散式架構的主要功能，可將分散於各異質平台上的資訊，以任何伺服器為入口，均可查詢分散各地之伺服器全文，提供平衡負載(Load Balance)與風險分擔(Risk Sharing)功能。

3.2.1.1.2 資料蒐集

在資料氾濫的今日，我們不是要「資料」，而是要資料組織及結構後的「知識」；因此「知識資源管理」才是未來競爭優勢的主要核心元素。

身為一個知識工作者，主要的工作就是平常資料的蒐集、分析，並整理出可用的資料以協助決策者做最後的決策。

隨著「網際網路」Internet 的普及與發達，網際網路已成為全球最大的知識庫，如何從幅員遼闊的網際網路中蒐集有價值的資料轉化成為知識庫，已成為重要知識來源。

當然知識工作者不可能二十四小時透過 Google 或是 Yahoo 等入口網站，從上千萬筆以上的資料去篩選過濾找出有價值的資料，因為這樣所需耗費的人力並不符合經濟價值。

因此，可建置「資料情報蒐集系統」，如圖 3-3，針對全球與文官制度有關的網站或是新聞知識庫過濾篩選，並自動定時下載資料或網頁，再透過「知識探勘」技術，將所下載的網頁資料分析、過濾，轉化成為具有價值的知識庫。

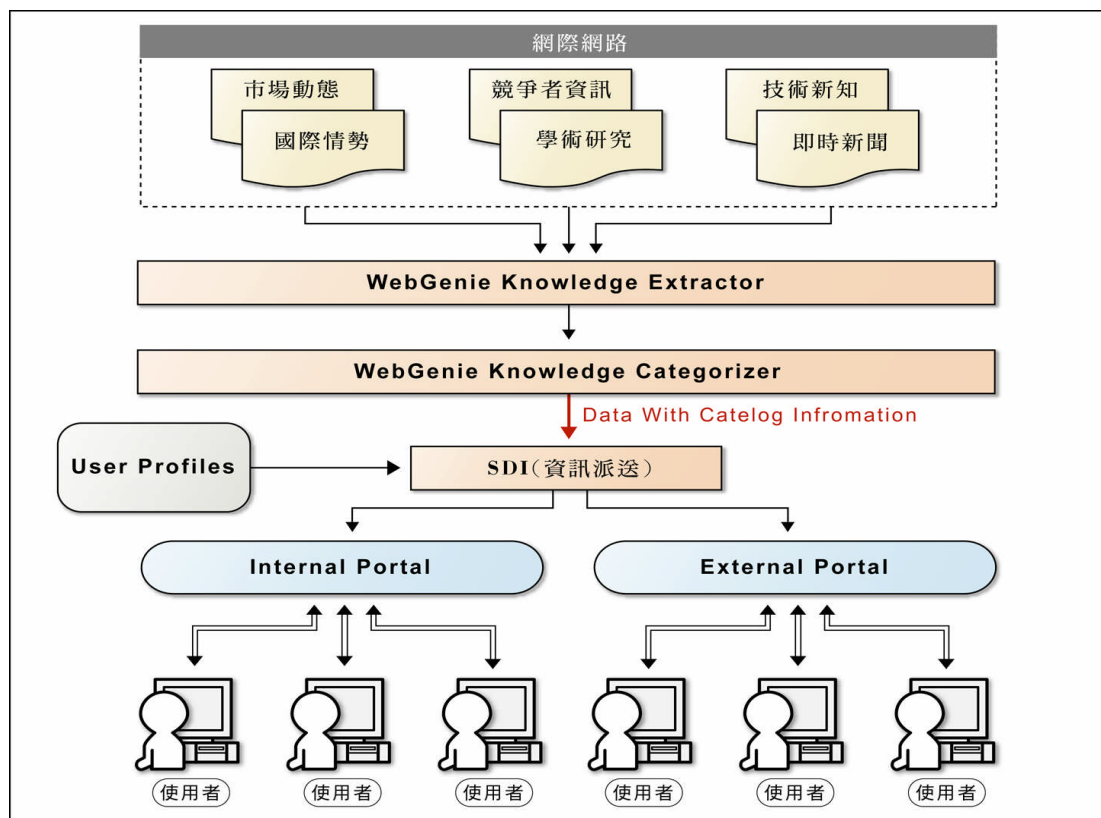


圖 3-3：資料情報蒐集系統架構圖

「資料情報蒐集系統」所採用的原理乃是透過「資訊代理人技術」提供使用者一個功能強大、使用容易的介面，使用者可以不受限制地設定資訊擷取來源，針對任何網站進行資料的擷取。它能將網路上非結構化的網頁資料轉換為結構化資料，如圖 3-4，存入指定的資料庫中的特定欄位。不論是動態或靜態網頁皆能轉換，協助快速有效率地擷取資料，並將資料轉換為資訊，提升其可用性。利用系統自動化代替人工做例行性的資料擷取動作。因此，使用者可依照需求可以定義出多個資訊代理人，每日定時將需要的資源都收集到資料庫中，不僅節約人力時間，取得的資料更能作為往後分析、統計的依據，做進一步的應用。



圖 3-4：資料情報蒐集系統原理

■ 知識探勘 (knowledge discovery, KD)

「知識探勘」(knowledge discovery, KD) 是擷取隱晦、有用、未被發掘、有潛在價值的規則、資訊或知識的一種過程。(Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, *Advances in Knowledge Discovery and Data Mining*, AAAI Press/ The MIT Press, 1996.) 在實務上，此過程需要運用資訊組織與分析等探勘技術，透過與使用者的互動，來反覆探索資料庫或文件庫，以發現新的、有趣的訊息或規律，再經由人工解讀結果，讓發現的規律訊息變成有用的資訊或知識。

根據資料特性的不同，「知識探勘」可分為「資料探勘」(data mining, DM) 與「文字探勘」(text mining, TM)。前者處理結構化 (structured) 資料，即每筆資料有共同欄位可記錄於資料庫者，而後者處理非結構化 (unstructured) 資料，即每筆資料沒有共通的結構性可言，經常為長短不一、記載訊息的自由文字。

知識探勘的步驟大致分為：

資料蒐集

資料清理

資料轉換

探勘技術運用

結果呈現與解讀

而知識探勘採用的方法，主要有：

關聯分析 (association)

分類 (classification)

歸類 (clustering)

摘要 (summarization)

預測 (prediction)

序列分析 (sequence analysis)

由於資料特性的不同，資料探勘 (DM) 與文字探勘 (TM) 在步驟與方法的技術細節上都有所差異。

文字探勘 (TM) 運用的技術，幾乎都跟詞彙的頻率與出現篇數有關，但這兩項資訊在資料探勘 (DM) 中極少用到。

資料探勘 (DM) 主要運用於大型資料庫上，提供資料庫管理系統額外的資料分析與統計功能；而文字探勘 (TM) 主要運用在大量的文件庫上，供作資訊搜尋、訊息過濾、事件關聯、趨勢預測、犯罪分析、案例追蹤、知識萃取、知識管理、決策輔助等之用。(曾元顯，2004)

3.2.1.2 資料分類

知識管理的核心的概念，便是「分享」，由於「分享」讓組織的智慧可以獲得最大的綜效。然而事與願違，由於資料分類的架構不清楚，因此找不到所要的資料。而這些資料不但無法協助我們獲得知識分享，而且還浪費許多寶貴的時間在於尋找資料。

資料分類是根據資料內容或主題給定類別的工作。例如，新聞資料可按其報導的內容，給予「政治」、「外交」、「娛樂」、「運動」等類別。通常，這些類別都是事先定義或選定，以符合管理者的需求與期望。而給定類別的工作，傳統上都由人工閱覽資料，根據其主題大意，給予適當的類別標示。

資料分類的目的，在對資料進行分門別類的加值處理，使得資料易於管理、利用。資料分類可將非結構化的資料，轉換成結構化的資訊，是資訊組織、主題分析與知識管理的重要工具。

因此資料分類成為知識分享中重要的瀏覽指標，如何將整合之後的資料庫透過適當的分類，在知識管理中是重要的事項，當然，如果要透過人力來作資料分類自然準確率會較高，但是相對也需要耗費很長的時間來完成。

近年來，拜資訊技術普及運用之賜，政府單位的數位文件資料不斷快速累積，數量大到難以有效的管理與利用，資料分類的需求也就因應而生。為此，如何利用「自動化的分類」技術，快速有效的協助人工分類，來應付大量暴增的分類需求，是現今資訊服務與知識管理的重要課題。

資料分類自動化後，會帶出更新、更便利的應用方式，除了提供館藏瀏覽（collection browsing）、主題檢索（topic-based retrieval）、文件管理（歸檔、調

閱、分享)外，還可應用在網頁過濾、電子郵件過濾、資訊選萃 (SDI, Selected Dissemination of Information)、資訊配送 (information filter or routing)、甚至是文字探勘(text mining)、新知發掘(knowledge discovery)、知識管理(knowledge management)等領域。跟資料檢索一樣，舉凡牽涉到非結構化資料的處理，都有資料分類的應用。

資料分類，需要瞭解資料的主題大意，才能給定類別，因此是相當高階的知識處理工作。要將資料分類自動化，必須先整理出分類時的規則，電腦才能據以執行。然而，多數的分類工作，其分類規則通常難以用人工分析歸納獲得。因此，機器在做自動分類之前，還必須加以訓練，使其自動學習出人工分類的經驗與知識。

現今自然語言理解的技術，還無法讓電腦瞭解任意的自由文句。因此機器在做資料分類時，常將資料分解成一個個語意較小的單位，通常為資料的關鍵詞彙，或稱「特徵詞彙」，再從這些詞彙與類別中找出對應的關係。有時分類的問題，簡單到只要資料的某個欄位中出現什麼特徵詞，就分到什麼類別去。但大部分的情況都沒那麼簡單。例如，「文化大革命」這個類別，如何界定哪些詞彙一定是屬於這個類別，哪些不是？顯然某些詞彙對這個類別較重要（比較有鑑別力），其他的則較不重要（比較不具鑑別力）。如何決定這些詞彙在每個類別的權重，正是機器學習 (Machine Learning) 可以派上用場的地方。

應用機器學習與自然語言處理的理論與方法，經過數年的資料自動分類研究、實驗與調整，資料自動分類成效已達人工分類水準，並且可同時進行上千個類別、上千篇文件的每日即時分類。目前為止，這些發展出來的技術與方法，已應用於數種產業領域，每日進行例行的資料自動分類任務。

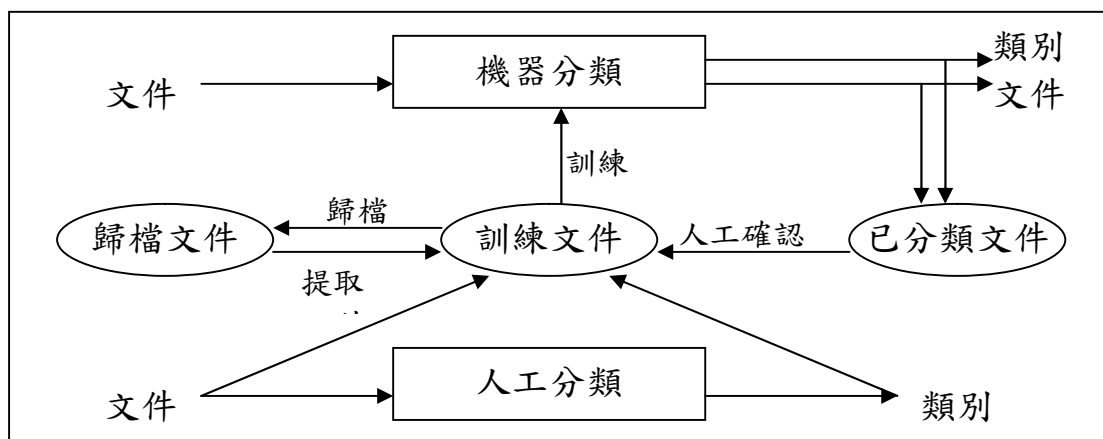


圖 3-5：分類流程圖

假如沒有機器分類，那麼人工就要仔細閱讀、分析該篇文件，然後為了標出類別，還要熟悉各個類別的意義及其在分類架構中的相對範圍。假若使用的分類架構很大，例如有上百個類別，一一記住每個類別的內容與範圍，以比對出文件的類別，常常需要熟練且具備耐心的人員才能勝任。

有了機器分類後，自動分類的結果對人力而言，可以視為是一種分類提示，對減輕人力分析文件、決定類別的努力，有相當大的助益。

尤其對新手而言，此種分類提示猶如有專家在旁協助，分類工作將更快能步上軌道。實際應用上，自動分類可以節省一半以上的人力。

自動分類也可以縮短分類人力的職前教育訓練成本與在職者的熟悉摸索期間。因此，人員流動造成的損失，可大幅降低。(曾元顯，2002)

3.2.1.3 資料查詢

根據一項調查顯示：一個知識工作者每天平均花 15%~35% 的時間，在搜尋與工作相關的資訊。而只有低於 50% 的機率，找到他們所需要的資訊！

為何會花如此高比率的時間在搜尋？原因之一是「沒有採用較智慧的工具」來輔助使用者，以致於影響工作效率。

政府機關的資訊來源，主要可分為二大類。一是「結構化資訊」二者為「非結構化資訊」。前者儲存在於單位組織的內部資料庫或試算表中，藉由系統化的語言（如統計工具或 OLAP 等等）將知識整理出來。而後者儲存在於單位組織內外之 Internet/Intranet 網頁及單位組織內部文書檔案、個別員工之專門技術檔案目錄或電子郵件中。

由於「非結構化資訊」較難以處理，因此市面上開發了許多的文字探勘工具(Text Mining Tools)來加強其使用效率。其中以「搜尋引擎」(Search Engine)是一種較為普遍的文字探勘工具。「分類引擎」(Categorization Engine)亦是一種方法，可建立起文件之間的關聯性。

「搜尋引擎」都是假設使用者很清楚在找什麼，但在大部分的情形下，使用者是：當我看到時，我就知道我要什麼。所以花了很多時間來做搜尋及瀏覽的工作；「分類引擎」可以主動告訴使用者，有類似相關的資訊在這個文件集中，是不是比較可能找到所需資訊！

知識工作者通常會用二個方法來找資料：

1. 明確地知道搜尋的答案，只是要找到更多相關資訊。
2. 不清楚真正的答案，只希望藉由搜尋及瀏覽來找到答案。

前者的需求，可透過「搜尋引擎」來滿足知識工作者的需求；而後者，則必須要有一個類似圖書館員的角色，引導使用者在正確的文件集中，藉由與使用者的互動，重覆地搜尋及瀏覽資料，以便找到適當的資料。

因此，我們可以說資料查詢的便利性必須同時具備「搜尋引擎」與「分類引擎」，因為二者是相輔相成的關係。其主要的使用時機如表 3-2：

表 3-2：搜尋引擎與分類引擎

搜尋引擎	分類引擎
<ul style="list-style-type: none">• 當你確定採用某些【關鍵詞】可以找到你要的文件• 當你【確定知道】要找那些文件時	<ul style="list-style-type: none">• 當你不确定要用那些【關鍵詞】時• 當你有直覺【一定在這群文件中】時• 可以協助使用者在特定分類下進行較小範圍的搜尋時• 當你想要提供類似入口網站的【目錄服務】時

一般傳統的資料查詢主要是以布林邏輯 (Boolean Logic) 精確比對的模式提供資料查詢。此種模式提供檢索條件之間的交集 (AND)、聯集 (OR)、差集 (NOT) 的運算，甚至於後切截比對 (truncation) 以及檢索詞之間鄰近條件 (proximity) 的過濾功能，對專業的檢索者而言，這些功能其實是相當有效的查詢工具。

然而由於一般的使用者對布林邏輯的運用較不熟悉、查詢結果未能依照符

合程度排序、以及查詢字串要求精確無誤等原因，布林邏輯的比對模式易導致較高的檢索失敗率（search failure）與資訊溢檢率（information overloading），從而造成資料查詢系統不易使用的情形。

從以上觀點可知，除可增加查詢條件讓使用者可以多元化選擇查詢方式，另外可透過查詢輔助工具及文字探勘技術，自動提示「查詢關鍵字」及「關聯性文件」，可簡化查詢方式及縮小查詢範圍，例如：自動關鍵詞、動態分類提示詞、以文找文、熱門查詢詞、K-map 知識地圖、時間序列、資料源選取、簡繁對譯等，輔助工具不僅可以幫助使用者快速找到所要查詢資料，更可以提升查詢的精確度。如圖 3-6：

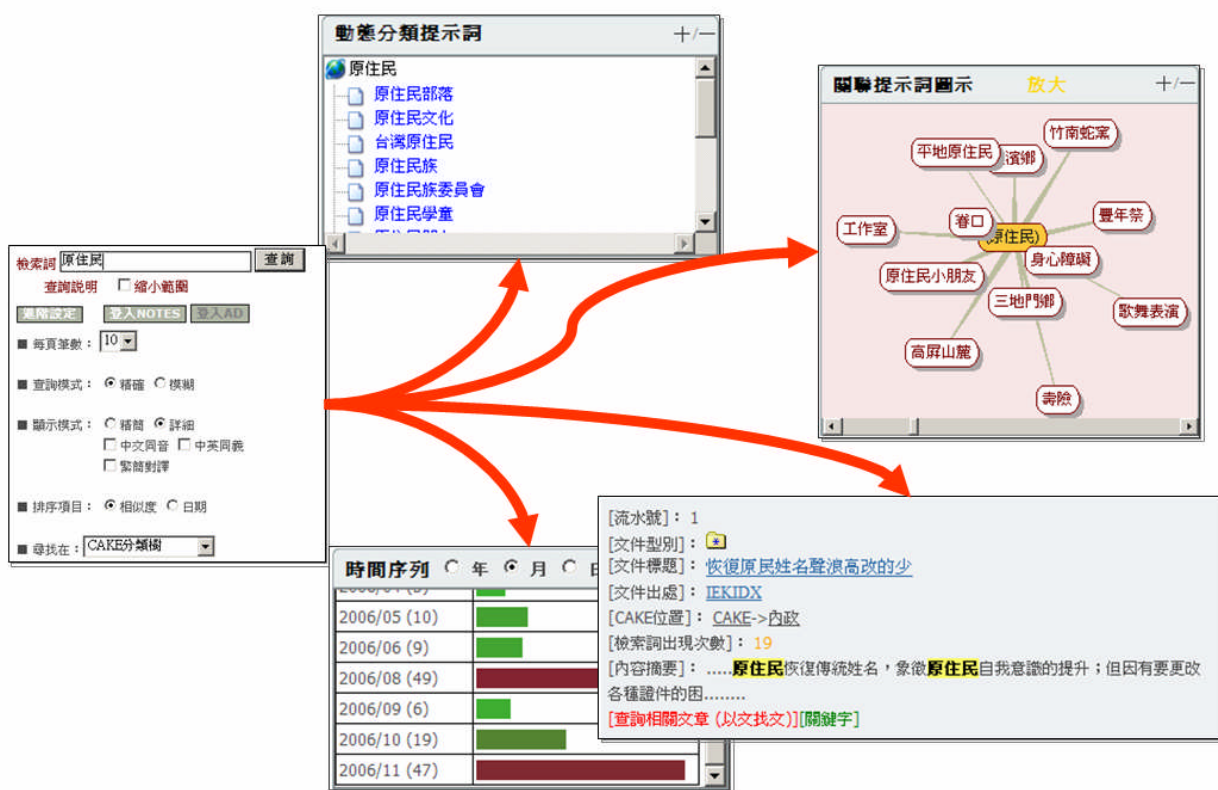


圖 3-6：查詢輔助工具應用圖

■ 政策中心 GRB 網站(參考網站搜尋功能分析)

STPI

GRB 智慧搜尋系統

Government Research Bulletin

一般查詢 / 進階查詢 查詢說明

字串查詢：
 自然語言 精準查詢 模糊查詢
 研究計畫 研究報告 全部(含全文)
每頁筆數: 10
排序依: 相似度 遞減
編號查詢:
 從結果中查詢 同義詞

年度: 之後 大於 千元 之前 小於 千元
研究性質: 忽略 基礎研究 自行研究 委託研究 應用研究
研究方式: 忽略 自行研究 委託研究 應用研究
研究領域: 所有研究領域
主管機關: _____
執行機構: (輸入代碼) _____ 查機構代碼 清除

檢索 清除

查詢得到 0 筆結果

財團法人國家實驗研究院科技政策研究與資訊中心
(10636)台北市和平東路二段106號14~16樓 TEL: 02-2737-7796 FAX: 02-2737-7669
隱私權保護及網站安全宣告

圖 3-7：政策中心 GRB 網站搜尋系統

資料來源：<http://grbsearch.stpi.org.tw/GRB/>

國家實驗研究院政策中心擔任國家科技資訊提供者的角色，陸續引進開發各項資料庫與相關電子資源，而該中心的各項資訊中包括了提供相關資料庫內容的服務，但因日益龐雜的網路資料，使用者欲進行資料檢索及過濾搜尋，往往需要進入各個網路服務網站逐一進行搜尋工作，不僅耗費大量時間並且針對相同的資料無法過濾與區隔。

因此透過文字探勘技術建置智慧型搜尋系統，如圖 3-7，整合該中心所有 GRB 的資料庫，使內部和外部的使用者能夠透過統一介面，找到相關資料。該網站除了資料搜尋整合外，更重要的部分是提供個人化服務和網站整合單一登入介面，透過該系統協助研究人員快速且有效的掌握各專業領域資訊，來幫助使用者進行更有效的主動分析與客戶服務。



圖 3-8：政策中心 GRB 網站搜尋輔助工具

並提供輔助工具針對搜尋的結果進行各種相關性統計分析功能，包括相關文章推薦、關聯地圖、事件時間序列分析等功能，幫助使用者能更精確地檢索、更完整地搜尋到所有需要的資料。如圖 3-8：

3.2.1.4 資料分析

西哲培根(Francis Bacon)說：『知識就是力量。』(Knowledge is Power.) 也就是說一個人所知愈多愈能解決其所遭遇的問題，由此可知現代的知識工作者必須不斷的充實知識與研究，方能解決其問題並有效的提升競爭力。

知識工作者常面臨的窘境在於遭遇問題時，如何從龐大知識庫當中找到關聯性文件作為決策分析，過去往往需要人工分析與整理，透過人工製作關聯性索引典的優點是正確性高，缺點則是成本大、建構速度慢、維護不易、以及事先選用的詞彙可能與後續或其他新進的文件無關。過去資訊檢索實驗的研究指出，一般目的 (general-purposed) 的索引典運用在特定領域的文件檢索上，會出現無法提升檢索效能的情形。

因此，透過「知識探勘」(Knowledge discovery or text mining)的技術，可以改善解決人工問題，「知識探勘」是一種運用人工智慧的資料分析方法，根據文獻本身的主題及內文，自動且即時產生關聯性，「知識探勘」的技術核心所產生應用包括：「自動關鍵詞」、「以文找文」、「動態提示分類」、「自動分類」、「關聯性知識地圖」等。

例如：在文官制度資料庫中，如果我們要搜尋所有資料有關【文官改革】，我們可以在檢索系統中輸入【文官改革】關鍵詞，系統可以針對該關鍵詞先篩選過濾找出文章標題或內容有【文官改革】的資料，在龐大資料庫中可能會出現上百筆以上搜尋結果，但並非代表所有的文件都是我們所要的知識內容，這當中可能包括法規、新聞、公告、論文等，當我們從其中一篇文件資料內容中瀏覽確認是我們所要的關鍵資料文件，但是可能還有上百筆文件我們無法確認。

因此，「以文找文」就可以幫助我們從這一篇關鍵資料文件，自動篩選交叉分析上百筆文件中與這篇關鍵資料文件內容相似度 90%的文件，找出關聯性並提供建議與提示。

此外，每一筆關聯性資料也可透過「以文找文」方式針對人才資料庫同時建立索引庫，除提供資料作為參考並可自動提示針對此文件的相關專家人才提示。

另外，運用「關鍵詞自動擷取」方法，還可以協助我們提示有關的關鍵詞或是關聯性動態提示詞，這些技術都可以協助知識工作者在作為決策分析時提供必要的知識文件建議。

「關鍵詞自動擷取」是一種辨認數位文件內有意義且具代表性字串(string)、片語(key phrases)、詞彙(keywords)、或內容片段(key segments)的自動化技術。

由於關鍵詞是呈現文件主題意義的最小單位，因此大部分對非結構化文件的自動處理，如自動索引、索引典自動建立、自動摘要、自動分類、自動歸類、相關回饋、自動過濾、事件偵測與追蹤、知識探勘、資訊視覺化、概念檢索、檢索提示、關聯知識分析、自動化權威控制、自動化詢答系統等，都必須先進行關鍵詞擷取的動作，再進行其他的處理。可以說，關鍵詞擷取是所有文件自動處理的基礎與核心技術。(曾元顯，1997)

3.2.2 需求說明

3.2.2.1 建置目標

1. 整合跨部會文官制度相關資訊來源及全球網際網路相關資訊蒐集，透過強大的知識搜尋探勘系統整合查詢考試院及相關部會之間相關資訊系統及資料庫，協助使用者迅速檢索目標文件。
2. 透過整合性檢索單一介面，達到資訊搜尋一站服務(one stop service)，提升文官制度資訊完整度。
3. 運用智慧型相關性與概念動態連結等搜尋技術，建立各資料源關聯性，提供決策者就單一事件或主題作多面向解析。
4. 建立個人化主題搜尋訂閱，便利使用者隨時檢閱最新之相關人事法規及資訊，以利分析、整理及組織相關文件。如圖 3-9：

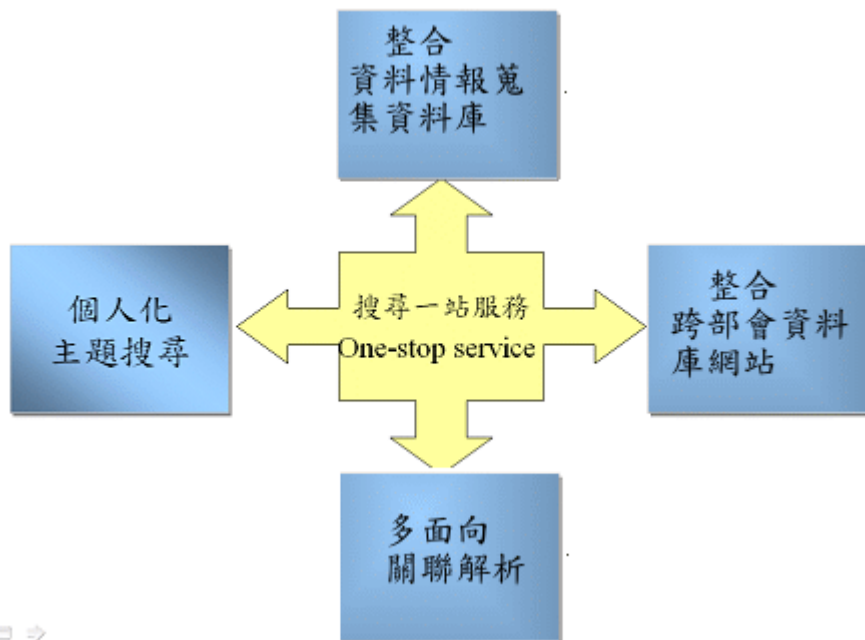


圖 3-9：文官制度資料庫架構概念圖

3.2.2.2 需求建議

文官制度乃是一個國家對其公務人員之人事政策、法規及實務運作之總稱。也因此文官制度資料庫更需要整合與跨部會的知識流通，並且搭配與新聞、輿情、學術研究進行整合性的諮詢與研討，雖然透過各項文官制度資訊系統的建置已大幅降低人工處理的時程，但是因應未來更趨龐大的業務資訊以及為強化個別業務單位或個人對資訊深度或廣度的掌握，規劃此一文官制度資料庫與資料探勘搜尋引擎及情報蒐集系統，以期創造知識型政府的綜效，並建立決策支援輔助的基礎建設，從需求面分析如下：

1. 人事資訊綜合查詢與關聯解析

我國文官制度為提昇行政效率、強化知識管理，已陸續建置完成相關資訊系統，例如：全國人事法規釋例資料庫、銓敘部法規輯要檢索系統、公務人員保訓會法規輯要、考選部法規輯要檢索系統、退撫基金法規彙編及釋例、公務人力資料庫、行政院暨所屬各機關職員錄編印系統等。

多年來已累積相當龐大的文官制度資訊。由於考試院的業務分工，承辦人員在蒐集所需要的業務資訊時，常常需要操作不同的應用系統、輸入不同的帳號/密碼及面對不同的操作界面的困擾。再者，承辦人員從不同的系統蒐集到資料後，仍須要花大量時間瀏覽這些文件，整理後才能知道它們彼此間的關係，以及判斷每一份文件的重要性，方能進一步整理出該主題之脈絡，以做為決策之參考依據。

因此，文官制度資料庫應導入一個整合內外資訊的綜合查詢(federate search)的系統並且透過文字探勘的應用，強化使用者的檢索策略與關聯性的分析，使其快速達到搜尋目標文件或歸納主題文件群的效益。

2. 院外資訊綜合查詢與關聯解析

考試院為國家最高的考試機關，因此在業務上，更需要統整考試院暨所屬機關之業務資訊，以利決策之依據。在過往，當考試院需要所屬機關提供相關資訊時，常需公文往返，曠日費時，影響行政決策的效率。

因此，文官制度資料庫應導入一個可支援分散索引/分散查詢的檢索系統，跨部會資料庫及異地資料庫，同時可與前述之院內系統在資料的搜尋上做整合，更進一步可建立彼此間的關聯性，將更有效的對資訊做全面的掌握。

3. 個人化主題蒐集與分類

考試院各業務單位或個別的業務同仁對資訊的掌握必須由點延伸至面，增加深度與廣度，必須要掌握單一事件或主題所有關聯性整合資訊，方能回應民眾需求。並且需要在業務主題的蒐集上更簡便，傳承上可建立脈絡進而可做有效益之分享。

因此，文官制度資料庫應導入一個主題專卷系統(個人化主題搜尋)之導入，可由各業務單位分別建立專屬的知識樹，整合文官制度知識分類架構，利用簡單的定義，將相關分類的主題資訊自動化的持續蒐集整理，同時各主題分類也可成為前述整合查詢的資料源，在權限的範圍內更方便分享個別業務單位或同仁所累積的成果。

4. 創造知識型政府的綜效，建立決策支援輔助的基礎建設

透過關聯性的建立與分散式的應用，可跨平台、跨部會的將個別系統串連起來，使政府內的知識工作者得以充分將資訊穿透的效益轉變成行政

知識的能量，並且配合深度的文字探勘能力，可將資訊做不同維度或多面向的解析，如地理事件圖、時間事件圖等等，據此以作為未來決策支援的輔助應用。如圖 3-10：

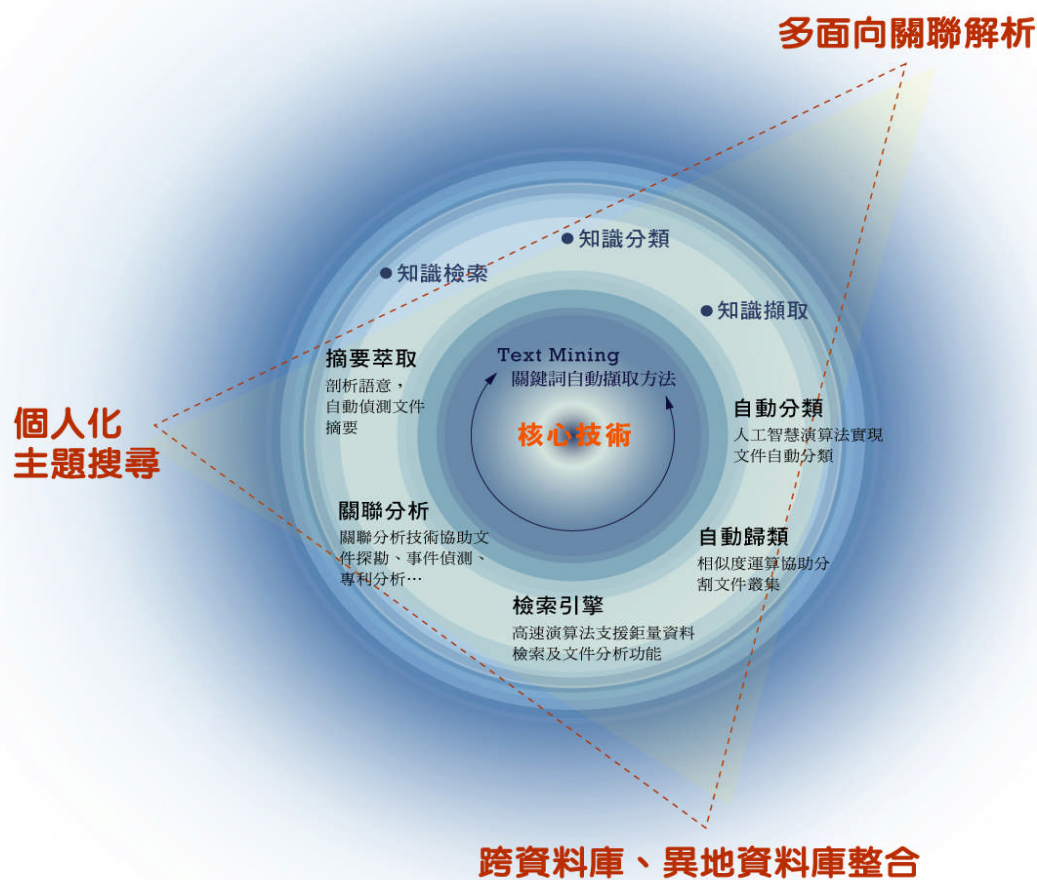


圖 3-10：文官制度資料庫架構核心技術關聯圖

3.2.2.3 整體系統分析

本系統主要整合院內及院外單位各式的資料庫、異質的資料種類，並提供一個智慧型的統一搜尋介面，可幫助使用者，依照工作的目標，對多面向的文官制度資料庫，進行資料整合及提供查詢，以挖掘龐大政府組織內外部的隱性資訊，並獲得多種角度、來源的相關資訊。可節省使用者學習、習慣多樣搜尋介面的時間。

同時亦可利用此整合後的共用知識庫，依照使用者需求，成立主題專卷。可進一步將主題相似的文件，依照主題，進行自動化分類，形成一個主題專卷。以達到組織內部知識流通、加值，提升公務人員行政效能。

進一步藉由此系統，建立跨部會(考選部、銓敘部、公務人員保障暨培訓委員會、人事行政局...)資訊及全球網路資料搜尋相關情報之整合平台，期能達到電子化資訊的附加價值增加、行政服務有效能的知識型政府。如圖 3-11：

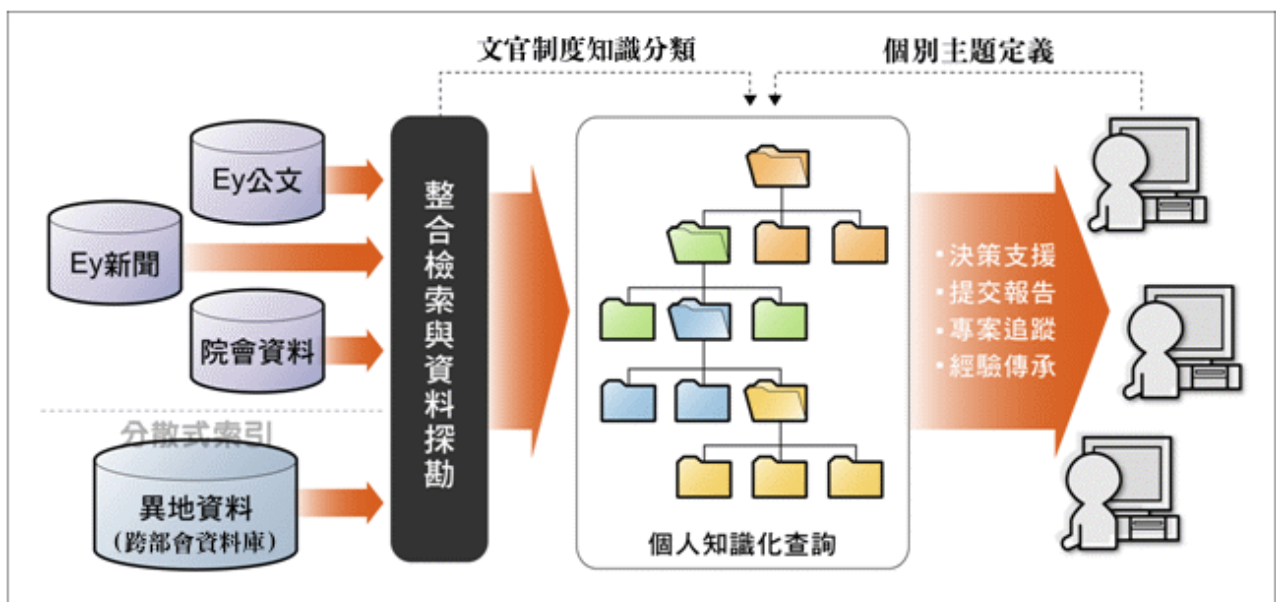


圖 3-11：文官制度資料庫系統應用流程圖

3.3 架構設計

本系統架構設計主要在於如何整合現有分散在各部會及異地有關文官制度資料庫，透過相關技術整合成為一個文官制度資料庫入口查詢。

整體設計上可分成三個系統架構：1. 搜尋整合系統架構 2. 資料情報蒐集系統架構 3. 權限管控系統架構 4. 資料分析擴充功能。以下針對各項系統架構作各項功能設計說明。

3.3.1 搜尋整合系統架構

搜尋整合系統架構主要在於文官制度資料庫中提供資料庫查詢及跨部會或異地資料庫整合，因此資料庫查詢的功能及輔助工具關係到整體資料庫在使用上的便利性及實用性。

本系統架構參考「**行政院施政整合檢索系統**」架構設計，並針對該系統所提供各項搜尋功能規劃說明及套用在文官制度資料庫建議方案。

■ 行政院施政整合檢索系統概述

「**行政院施政整合檢索系統**」主要整合行政院內部各式的資料庫、異質的資料種類，內容包括首長報告及輿情分析(文字檔)、院會議事(文字檔)、新聞知識庫(資料庫檔案)、全球資訊網(網頁資料)、公文系統(資料庫檔案)、立委質答詢系統(資料庫檔案)及員工入口網公告資料(資料庫檔案)，協助院內人員，在統一的平台，對多面向的政府內部資料庫，進行整合性的查詢，並建置多元化的輔助查詢工具，協助行政院同仁透過該系統以挖掘龐大政府組織內部的潛在資訊，並大量提升工作效率。以下就主要查詢功能及輔助工具做說明。

3.3.1.1 行政院施政整合檢索系統各項功能



圖 3-12：行政院施政整合檢索系統首頁畫面參考

- 動態分類提示詞

此功能是由系統自動計算而產生動態分類目錄的提示詞，不需以人工建置及維護詞庫。好處是可以提示使用者，關於查詢字串的次主題、加長詞，或狹義詞，而能快速縮小查詢範圍。如圖 3-13：

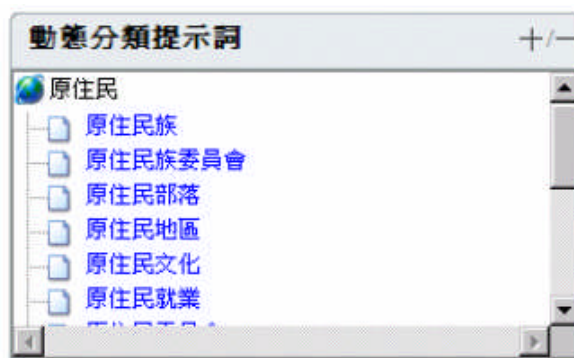


圖 3-13：動態分類提示詞

- 關聯提示詞

此功能也是系統自動計算而產生與查詢主題相關的聯想主題詞彙。和動態分類提示詞相同，亦不需由人工建置及維護詞庫。其呈現的

方式，是以查詢主題為中心，提供使用者「字串不必相近、但主題相近」的類似詞彙，意即字面上或許看不出關聯性，但兩詞彙之間常在同一段落或句子中出現，表示它們在字義上有某種層度的相關性。此關聯地圖功能還可進階輻射展開其他相關主題詞彙。如圖 3-14：

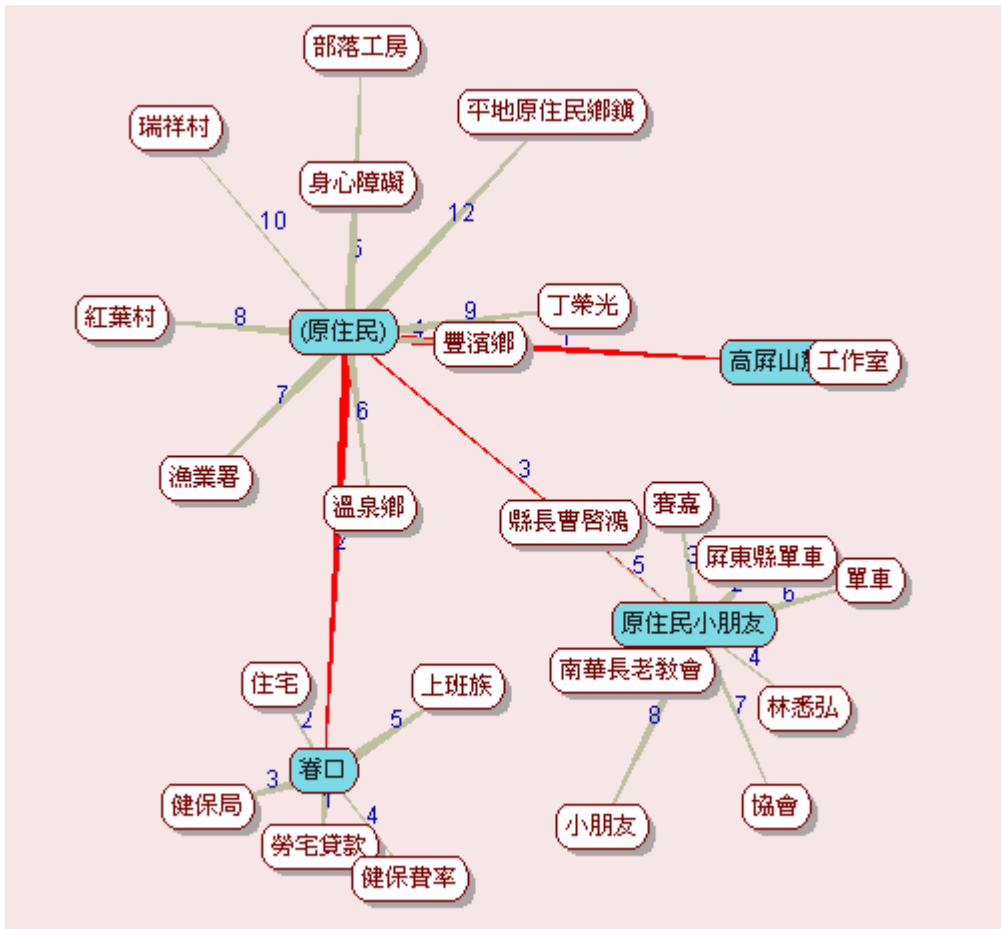


圖 3-14：關聯性知識地圖

- 時間序列分析

系統會自動統計搜尋結果之日期，以年、月、日的方式排列，以幫助使用者分析主題事件發生頻率與趨勢。此為互動式統計長條圖，可直接點選長條圖，以閱讀該時間區間內，符合搜尋主題的文件。如圖 3-15：

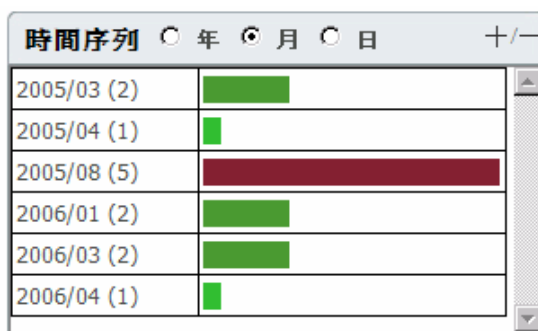


圖 3-15：時間序列分析

● 互動式查詢詞建議

使用者選擇典範文件後，利用中文之斷字斷詞功能，讓使用者勾選文件中具代表性之重要詞彙，以強化檢索詞。使用者以單篇文章中的關鍵字作為查詢條件，再做進階檢索。此功能能夠協助使用者更精確的查詢。

如圖 3-16：

[流水號]： 2
 [文件型別]：
 [文件標題]： 原住民族基本法過關 保障原住民族自治權
 [文件出處]： 新聞
 [CAKE位置]： CAKE->內政
 [檢索詞出現次數]： 44
 [內容摘要]：原住民族基本法過關保障原住民族自治權...法院會今天三讀通過「原住民族基本法」，本案攸關原住民族基本權利.....
 [查詢相關文章 (以文找文)] [關鍵字]

<input type="checkbox"/> 人員	<input type="checkbox"/> 土石	<input type="checkbox"/> 土地	<input type="checkbox"/> 司法	<input type="checkbox"/> 台北	<input type="checkbox"/> 生命	<input type="checkbox"/> 行為
<input type="checkbox"/> 法案	<input type="checkbox"/> 政府	<input type="checkbox"/> 政策	<input type="checkbox"/> 重點	<input type="checkbox"/> 席次	<input type="checkbox"/> 特考	<input type="checkbox"/> 能力
<input type="checkbox"/> 測驗	<input type="checkbox"/> 程序	<input type="checkbox"/> 菌類	<input type="checkbox"/> 業務	<input type="checkbox"/> 福利	<input type="checkbox"/> 礦物	<input type="checkbox"/> 權利
<input type="checkbox"/> 立法院	<input type="checkbox"/> 自治 權	<input type="checkbox"/> 行政 院	<input type="checkbox"/> 委員 會	<input type="checkbox"/> 原住民	<input type="checkbox"/> 基本法	<input type="checkbox"/> 應考人
<input type="checkbox"/> 事業機 構	<input type="checkbox"/> 協商 會議	<input type="checkbox"/> 原住 民族	<input type="checkbox"/> 原住 民語	<input type="checkbox"/> 國家公 園	<input type="checkbox"/> 基本權 利	<input type="checkbox"/> 野生動 物
<input type="checkbox"/> 原住民 地區	<input type="checkbox"/> 原住 民事務	<input type="checkbox"/> 原住 民保留地	<input type="checkbox"/> 原住 民族爭	<input type="checkbox"/> 原住 民教育	<input type="checkbox"/> 生物多 樣性知識	<input type="checkbox"/> 陳舒棠
<input type="checkbox"/> 原住民 傳統領域	<input type="checkbox"/> 原住 民土地	<input type="checkbox"/> 原住 民語言	<input type="checkbox"/> 人口 比例	<input checked="" type="checkbox"/> 原住 民 綜合發展基 金	<input type="checkbox"/> 原住 民 族文化	<input type="checkbox"/> 國家級 風景特定區
<input type="checkbox"/> 原住 民 政策	<input type="checkbox"/> 山川 傳統名稱	<input type="checkbox"/> 原住 民自治權	<input type="checkbox"/> 全國 醫療網	<input type="checkbox"/> 原住 民 傳統領域土 地	<input type="checkbox"/> 原住 民 公務人員	

加入新查詢詞 關閉視窗

圖 3-16：自動關鍵詞示意圖

● 相關文章推薦

使用者可以對某一文件，進行『以文找文』。此功能是以自動歸類的技術，在資料源的文件中，推薦與之相關的其他文件。並且，再推薦的文章中，使用者亦可依照資料來源做進一步篩選。此功能快速協助使用者找到更多相關資訊。如圖 3-17：



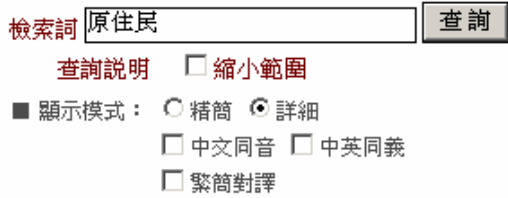

圖 3-17：以文找文示意圖



圖 3-18：資料探勘查詢結果示意圖


3.3.1.2 搜尋系統架構之需求與功能建議

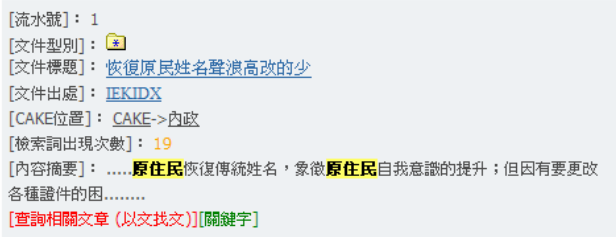

需求	功能特性
(1)搜尋資料範圍	<p>可針對Oracle資料庫、檔案目錄系統、網站網頁進行搜尋資料範圍設定</p> <p>本系統特別可針對公文之特性(一個公文對映多份簽呈)，進行關聯性檢索</p>
(2)搜尋資料型態	<p>支援常見之檔案格式如txt、html、xml、Office、eml、PDF、Dyna Doc、mht之檔案</p>
(3)搜尋功能需求	
關鍵詞簡易全文檢索	<p>搜尋標的不限時間、資料源與檔案格式</p>
進階查詢模式	<p>進階選項再行縮小查詢範圍</p> <p>檢索詞 <input type="text" value="原住民"/> <input type="button" value="查詢"/></p> <p>查詢說明 <input type="checkbox"/> 縮小範圍</p>
容錯模糊、自然語言與精確查詢	<p>容錯模糊搜尋:中文模糊查詢係為針對中文使用習慣上，可能產生的盲點，提供輔助的功能</p> <p>「中研院」/「中央研究院」等以簡稱合併查詢全名等，提高搜尋的完整性。</p> <p>檢索詞 <input type="text" value="原住民"/> <input type="button" value="查詢"/></p> <p>查詢說明 <input type="checkbox"/> 縮小範圍</p> <p>■ 查詢模式：<input checked="" type="radio"/> 精確 <input type="radio"/> 模糊 <input type="radio"/> 主題式</p> <p>自然語言: 真正讓使用者輕鬆查詢，可輸入較</p>

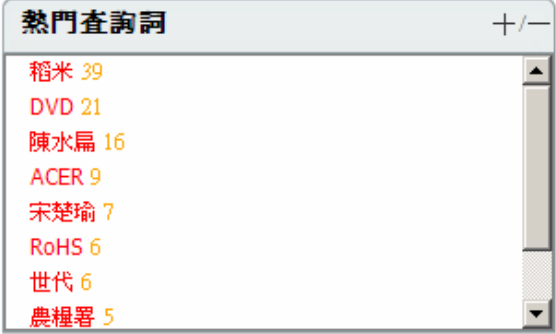

需求	功能特性
	長且口語化的句子(topic-base sentence)，針對口語化查詢進行語意分析
布林邏輯查詢	包含複合式之AND、OR及NOT運算
中文同音查詢	針對中文使用習慣上，提供中文同音、中英文同義輔助功能。例如：「台灣」/「臺灣」的同音異形字。
中英文同義詞查詢	 <p>可將其他中、英文同義詞一併納入檢索。</p>
欄位式檢索功能	可針對關聯式資料庫挑選單一或多個索引欄位進行合併查詢，可指定欄位間布林邏輯關係等。
指定資料源檢索	<p>可指定多個資料源索引檔進行檢索</p> 
檔案資料類型搜尋篩選	提供常見檔案型篩選功能

需求	功能特性														
檔案時間區間檢索	<p>提供時間區間內之檢索</p> <thead> <tr> <th>日期</th> <th>數量</th> </tr> </thead> <tbody> <tr> <td>2005/03 (2)</td> <td>2</td> </tr> <tr> <td>2005/04 (1)</td> <td>1</td> </tr> <tr> <td>2005/08 (5)</td> <td>5</td> </tr> <tr> <td>2006/01 (2)</td> <td>2</td> </tr> <tr> <td>2006/03 (2)</td> <td>2</td> </tr> <tr> <td>2006/04 (1)</td> <td>1</td> </tr> </tbody>	日期	數量	2005/03 (2)	2	2005/04 (1)	1	2005/08 (5)	5	2006/01 (2)	2	2006/03 (2)	2	2006/04 (1)	1
日期	數量														
2005/03 (2)	2														
2005/04 (1)	1														
2005/08 (5)	5														
2006/01 (2)	2														
2006/03 (2)	2														
2006/04 (1)	1														

| 以文找文檢索 | 提供以文找文之檢索 |
| 關聯圖檢索 | 查詢關鍵詞與查詢結果，動態產生視覺化之概念關聯圖，提供使用者依據概念關聯圖進一步檢索資料。 |

需求	功能特性
	
<p>整合知識分類架構</p>	<p>可結合CAKE分類架構，在特定分類節點下檢 索全文。</p> <p>檢索詞 <input type="text" value="原住民"/> <input type="button" value="查詢"/></p> <p>查詢說明 <input type="checkbox"/> 縮小範圍</p> <p>■ 尋找在：</p> <ul style="list-style-type: none"> CAKE分類樹 CAKE分類樹 ...國家發展類 ...內政類 ...交通 ...國防 ...財政金融 ...經濟 ...教育文化 ...法務 ...大陸工作 ...衛生
<p>查詢與結果中再查詢功能</p>	<p>提供Search within(縮小範圍)功能</p> <p>檢索詞 <input type="text" value="原住民"/> <input type="button" value="查詢"/></p> <p>查詢說明 <input type="checkbox"/> 縮小範圍</p>
<p>(4) 查詢結果功能需求</p>	
<p>簡目與詳細輸出</p>	<p>使用者可自行設定輸出模式</p>
<p>顯示總筆數、所需時間、</p>	<p>顯示查詢總筆數、查詢所需時間、排序方式、</p>

需求	功能特性
排序方式、資料來源以及每頁顯示筆數	資料來源以及每頁顯示筆數。
依相關性、時間及指定欄位順序，可遞增或遞減排序	本系統可指定相關性,時間及欄位排序
顯示文件標題、日期、型態、自動摘要、資料來源與關鍵字位置	標題、出處及文件所在CAKE的位置 同義詞標示反白顏色 本系統提供動態摘要、重點摘要及首行摘要三
同義字標示功能	種(三選一)
顯示關鍵字的自動摘要	 <p>[流水號]: 1 [文件型別]:  [文件標題]: 恢復原民姓名聲浪高改的少 [文件出處]: IEKIDX [CAKE位置]: CAKE->內政 [檢索詞出現次數]: 19 [內容摘要]:原住民恢復傳統姓名，象徵原住民自我意識的提升；但因有要更改各種證件的困..... [查詢相關文章 (以文找文)][關鍵字]</p>
進一步連結顯示詳細資料	詳細之連結資料，配合院之呈現界面，進行客製
熱門文章熱門關鍵字推薦	主動提供前十名熱門關鍵字

需求	功能特性
	
(5) 個人化查詢知識工具	
提供查詢結果自訂主題分類樹	<p data-bbox="639 763 1050 801">專案客製個人化查詢工具</p> 
自動顯示該主題最新文件	
分享個人化主題	
個人主題分類樹管理介面	
(6) 使用介面功能需求	<p data-bbox="600 1346 1347 1480">符合IE6版(含)以上版本及Firefox 1.0版(含)以上功能。</p> <p data-bbox="600 1536 963 1574">配合考試院設計風格。</p> <p data-bbox="600 1630 1150 1668">提供使用手冊等線上文件及FAQ。</p> <p data-bbox="600 1724 1347 1859">配合考試院內部網路資訊系統，提供介接功能模組。</p>
(7) 系統索引(含索引管理)	

需求	功能特性																																										
中文與英文(Big5, Unicode)	本系統支援Big5, Unicode																																										
索引全文、資料庫欄位及 metadata	可針對全文、及指定欄位進行索引																																										
索引建立狀態	<p>提供紀錄檔及管理功能，包含建置每個索引時間、每個索引檔案之筆數及索引失敗等，資料來源索引建立狀態等，方便管理與分析</p> <table border="1"> <thead> <tr> <th>索引名稱</th> <th>資料來源</th> <th>文件總數</th> <th>上次索引時間</th> <th>索引所費時間</th> <th>索引失敗文件</th> </tr> </thead> <tbody> <tr> <td>MTKNE...</td> <td>資料庫</td> <td>13577筆</td> <td>2006/11/27 10:50...</td> <td>0分7秒</td> <td>1筆</td> </tr> <tr> <td>IEKIDX</td> <td>資料庫</td> <td>134278筆</td> <td>2006/11/22 17:30...</td> <td>31分43秒</td> <td>無</td> </tr> <tr> <td>Notes</td> <td>Lotus Notes資料</td> <td>0筆</td> <td>1900/1/1 00:00:00</td> <td>0分0秒</td> <td>無</td> </tr> <tr> <td>農資</td> <td>目錄檔案</td> <td>1073筆</td> <td>2006/12/11 12:30...</td> <td>19分2秒</td> <td>無</td> </tr> <tr> <td>pdata</td> <td>資料庫</td> <td>263筆</td> <td>2006/11/22 11:20...</td> <td>1分14秒</td> <td>無</td> </tr> <tr> <td>fox</td> <td>資料庫</td> <td>2筆</td> <td>2006/11/28 18:48...</td> <td>0分24秒</td> <td>無</td> </tr> </tbody> </table>	索引名稱	資料來源	文件總數	上次索引時間	索引所費時間	索引失敗文件	MTKNE...	資料庫	13577筆	2006/11/27 10:50...	0分7秒	1筆	IEKIDX	資料庫	134278筆	2006/11/22 17:30...	31分43秒	無	Notes	Lotus Notes資料	0筆	1900/1/1 00:00:00	0分0秒	無	農資	目錄檔案	1073筆	2006/12/11 12:30...	19分2秒	無	pdata	資料庫	263筆	2006/11/22 11:20...	1分14秒	無	fox	資料庫	2筆	2006/11/28 18:48...	0分24秒	無
索引名稱	資料來源	文件總數	上次索引時間	索引所費時間	索引失敗文件																																						
MTKNE...	資料庫	13577筆	2006/11/27 10:50...	0分7秒	1筆																																						
IEKIDX	資料庫	134278筆	2006/11/22 17:30...	31分43秒	無																																						
Notes	Lotus Notes資料	0筆	1900/1/1 00:00:00	0分0秒	無																																						
農資	目錄檔案	1073筆	2006/12/11 12:30...	19分2秒	無																																						
pdata	資料庫	263筆	2006/11/22 11:20...	1分14秒	無																																						
fox	資料庫	2筆	2006/11/28 18:48...	0分24秒	無																																						
同義詞庫及停用詞建置與維護功能	系統提供友善之同義詞及停用詞維護管理界面																																										
索引排程功能	系統可指定各資料來源定期（每小時、每日、循環週期等）自動更新索引功能，並提供時間差排程模式																																										
漸進式索引	系統可配合索引排程功能，僅針對有異動資料進行更新索引作業，以加快索引更新時間																																										
手動建置/中止索引功能	提供排程暫停、恢復、刪除等設定管理功能																																										
(8) 分散索引/分散查詢	將建立索引及檢索的工作負載，平均分散於各																																										

需求	功能特性
	<p>伺服器中，有效分散查詢流量增進系統效能，降低當機風險提高系統可用程度，預留擴充機制升級系統便利迅速，以達成系統穩定性、擴充性、可用性、管理性。</p>
(9) 系統管理功能需求	<p>系統提供使用訊息，並提供查詢統計與分析資料，包含區段使用量、熱門關鍵字統計、熱門文章與熱門文件排行分析、熱門索引庫分析（以調校索引時間區間）及查不到關鍵字等統計報表分析</p>
(10)系統安全/權限功能	<p>提供系統備份與還原功能</p> <p>系統應與資料源系統權限設定一致</p> <p>防止SQL Injection等功能。</p> <p>提供「自動登出」功能</p>

一般知識工作者面對大量資料時，傳統搜尋引擎的搜尋結果只是相關文件的排序清單，呈現檢索詞彙與儲存的文件間的關係，但對於這些文件間的關聯性，則很少或沒有提供。

相對知識工作者仍須花費大量時間瀏覽這些文件，整理後才能知道它們彼此間的關係，以及判斷每一份文件的重要性，方能整理出脈絡。

因此，針對上述之問題，主動提供查詢線索(動態分類詞、關聯提示圖、時

間序列及內容摘要)，讓每份文件從孤立的資訊變成互相關聯的知識網絡，提升對於資訊檢索的效率。

3.3.2 資料情報蒐集系統架構

資料情報蒐集系統架構功能可用於外部資料庫或是網際網路等，具有研究價值資料蒐集，例如：新聞媒體資料簡報蒐集、國外文官制度網站資料蒐集、國際論文或是學術機構網站資料蒐集等。

資料情報蒐集系統的價值在於增加文官制度資料庫的外部知識價值，以疾病管制局「國際疫情擷取系統」為參考範例說明，過去曾經發生過的 SARS 疫情，由於缺少充足的資料情報，因此當政府決策部門遇到疫情發生時，確無法立即提出防範解決方案，由於資料的缺乏因此需要耗費相當大的人力去研究，往往失去第一時間的關鍵決策，造成疫情擴大，因此疾病管制局在 SARS 爆發後經過深思檢討建立了「國際疫情擷取系統」，在二十一世紀的時代，知識的擴充與來源必須跟隨全球性的同步才能提升競爭力，當今年爆發「禽流感」疫情，疾病管制局從「國際疫情擷取系統」所蒐集的全球性有關「禽流感」資訊及過去 SARS 所蒐集整理的解決方案作為參考，使得決策部門可在第一時間提出 SOP 處理方案與預防措施，因此控制「禽流感」的疫情擴大。

由此可知，資料情報蒐集系統對於文官制度資料庫的知識擴充價值具有整合建立指標意義。

以下參考疾病管制局「國際疫情擷取系統」，並針對該系統功能概述作為未來建置資料情報蒐集系統參考及建議。

■ 疾病管制局「國際疫情擷取系統」概述

疾病管制局
Centers for Disease Control, R.O.C. (Taiwan)

回首頁 | 與我們連絡 | English | 操作手冊 | 網站導覽 | RSS | 字彙對照

疫情報導 網站

依照關鍵字 go 進階檢索

會員登入 加入會員

2009/08/14
目前已瀏覽人數 5709684

各期內容總覽 | 留言與回覆 | 投稿需知

本刊物之簡介
編輯委員會
訂閱方式
紙本訂閱
免費訂閱

疫情報導
 疫情報導文章
 合約實驗室病毒監測
 國內重要疫情
 國際疫情

最新訊息

線上測驗開始了！ 2009/8/1
親愛的會員請注意：疫情報導線上測驗開始了！本局「疫情報導」線上測驗正與「傳染病數位學習網」進行會員整合，以提供會員更優質與高效率的學習環境與功能，目前「疫情報導」線上測驗會連線到「傳染病數位學習網」網頁中進行測驗。欲參加者請先加入會員，再登錄即可。測驗成績達80分，可取得公務人員數位學習時數1小時，護理人員可取得通訊課程2點。

疫情報導文章

2004-2008年高高屏登革熱病媒蚊密度調查分析 2009年7月
本研究分析高高屏地區71個鄉鎮市區1,364村里2004-2008年的登革熱病媒蚊調查資料，並分析每週登革熱個案數與病媒蚊各項指數之相關性。根據調查，近5年在高雄縣市方面僅高雄縣茂林鄉未發現埃及斑蚊，而屏東縣95%鄉鎮市有埃及斑蚊分布，但高高屏均有白線斑蚊分布，雖然整個高高屏地區白線斑蚊分布廣於埃及斑蚊，但是在人口密度較高的高雄市，埃及斑蚊分布則廣於白線斑蚊，且高雄市僅查獲埃及斑蚊的里數有增加趨勢而高雄縣及屏東縣均無此情形。孳生容器種類統計以水溝、水塔、帆布、排水管等其他類最多，次之為水桶、花瓶、底盤、水缸陶甕、輪胎、保利龍箱蓋、塑膠盤、罐子、杯子、冰箱底盤及地下室，由於其他類比率高達33%，顯示調查種類已有重新設定必要。近5年高高屏各村里每年登革熱發生率與埃及斑蚊成蟲指數具相關性，且相關係數較白線斑蚊成蟲指數高，另每週病例數與每週埃及斑蚊成蟲指數亦具相關性，但與白線斑蚊成蟲指數則無相關性。分析高高屏近5年密度調查執行情形如調查里次方面以高雄市最多累計42,497里次，調查戶數方面以屏東縣最多累計3,067,830戶，積水容器方面以屏東縣調查最多，累計2,835,610件，陽性容器方面亦以屏東縣清除最多，累計162,826件。另外在戶內外積水容器比值方面以高雄市最高約4:6，另高高屏在戶內外積水容器比值介於1:9至2:8間。關鍵字：登革熱、埃及斑蚊、白線斑蚊、孳生源、病媒蚊分布。(全文)

圖 3-19：疾管局疫情報導網站

資料來源：<http://teb.cdc.gov.tw/main/main.aspx>

疾病管制局「國際疫情擷取系統」主要可分成以下幾個功能特性：

1. 外部資訊擷取：自動擷取外部網站，協助人員有效地蒐集最新訊息。
2. 資訊自動分類及分類關鍵自動提示，協助標定細分類內容：將所擷取蒐集的情報資訊依照分類架構自動歸入所屬類別，以節省人工閱讀、分類的時間，讓權責單位可以快速掌握應取得的訊息。除此之外，提供關鍵詞自動提示功能，使用者選取具代表性的文件，系統即透過關鍵特徵擷取機制，將文件中重要關鍵詞彙提示作為類別建議詞彙。
3. 資訊個人化/派送機制：建立個人化機制，人員可依照需求訂閱分類架構的某些節點，或使用者自訂關鍵詞組合。

4. 資訊檢索：在單一平台上整合疾病管制局過去整理過的資訊加上本專案擷取的外部資訊，透過文字探勘的運算，提供文件檢索、分析、關聯運算、事件時間序列分析等功能，使人員在單一的操作界面上，取得相同主題概念之資訊，以加速研究效率。如圖 3-20：



圖 3-20：疾管局國際疫情擷取系統

3.3.2.1 資料情報蒐集系統架構之需求與功能建議

3.3.2.1.1 外部網站資料擷取

目前網路上充斥大量而且非結構性的網頁資料，使用者在搜尋、利用以及儲存上均面臨很大的困難，知識擷取系統即是在此前提之下，提供文官制度資料庫指定資訊來源網站，由系統進行網頁切割模式，過濾網頁中不要擷取的資訊，再將切割擷取的文字自動轉化成為資料庫，並藉由可事先自由設定的排程

功能，自動擷取網路上之資料，以節省資訊蒐集時所需大量耗費之人力成本與時間，同時依管理者事先定義完成之格式，將非結構化之網頁資訊，重新切割並存入結構化關聯資料庫中相對應之欄位，以提供更進一步之利用。如圖 3-21：



圖 3-21：頁面切割設定參考

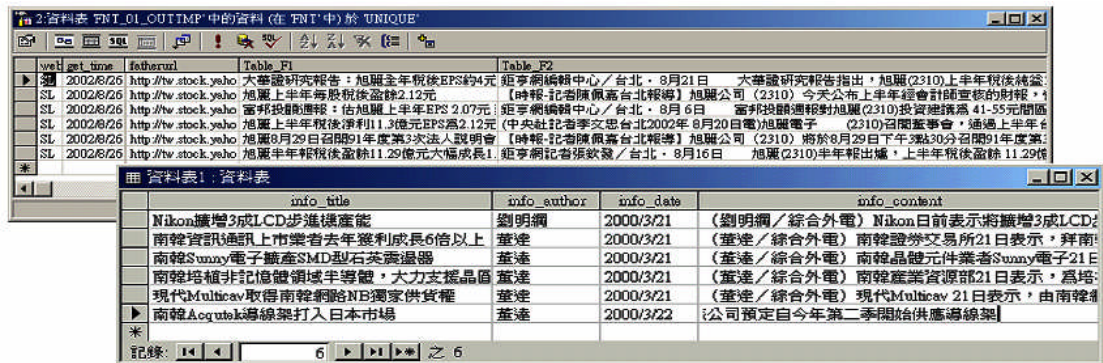


圖 3-22：擷取後資料庫產生畫面參考圖

3.3.2.1.2 知識分類(自動分類導入)

每日之外部訊息經過知識擷取系統帶回內部之後，應經過專家訓練將所有資訊加以分類設定，經過訓練之後未來擷取下載資訊即可達到自動分類功能，並可依據文官制度資料庫所規劃知識分類自動導入，以簡化人力分類，並且可依照類別提供使用者每日訂閱電子報或 RSS，並可再進一步提供使用者個人化的服務。如圖 3-23：

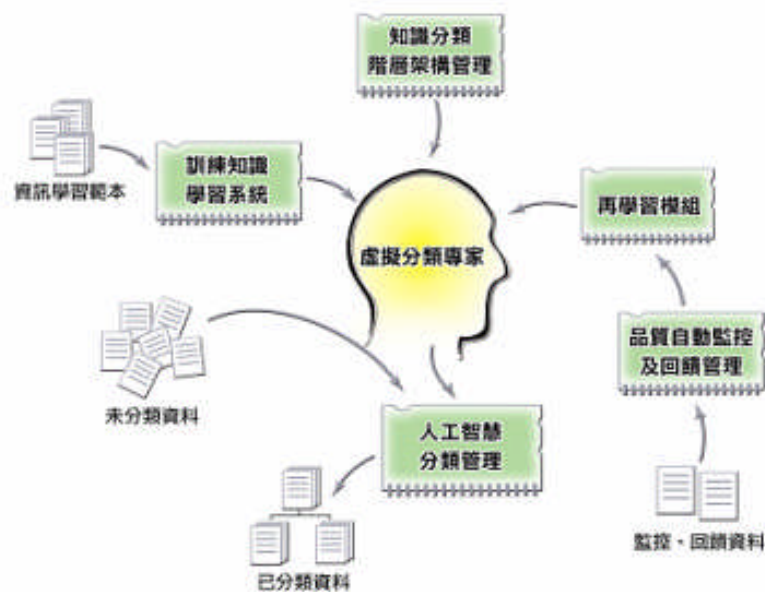


圖 3-23：知識分類訓練流程圖

3.3.3 權限管控系統架構

文官制度資料庫在整合上，可將資料區分，依照權限管控給予瀏覽設定，例如有些資訊並不對外公開，則必須透過申請帳號密碼，透過會員登入方式才可進行瀏覽。

由於文官制度資料庫所提供服務對象跨及各部會，本系統也規劃一個專屬個人化主題蒐集與分類區，使用者可透過帳號密碼登入，自行規劃設定自己的專屬分類，可將經常性閱讀或是個人蒐集的資料透過維護管理介面刊登放至個人化主題蒐集與分類區，可自行設定提供是否開放權限給予其他人瀏覽或分

享，達到知識共用及知識分享目的。如圖 3-24：



圖 3-24：個人化主題蒐集專區瀏覽參考

3.3.4 資料分析擴充功能

知識文件經常在知識內容中與時間及地理位置有密切關係，透過文字探勘技術可將這些知識文件中的時間及地理位置找出關聯性，對於一位決策者分析者來說，可透過時間序列及地理位置清楚掌握事件的發展過程及地理位置的關聯。

3.3.4.1 地理事件分析

在資料的蒐集分析上，【事件】通常與【地理】有一定的關係。因此，

本系統建議在進行資料庫整合後，分析事件與地理的關係，將會是一個重要的課題。

以文官制度為例說明，一位決策者經常需要掌握國際文官制度發展及改革變化，當然包括國際上許多國家的總統或是內閣改組等，此外，也可以掌握台灣地區各縣市首長包括人事異動或是組織改選，這些都可以透過地理事件快速查詢及交叉分析事件變化過程，迅速掌握相關資訊。

主要的功能規劃建議如下：

提供【事件分類樹】及【地區分類樹】，以進行【事件】與【地區】之交叉分析比對。

【事件分類樹】可由文官制度知識分類架構或使用者自訂分類架構中擷取出來。

【地區分類樹】提供以世界地圖及台灣地圖為主，包括【23 縣市】的地區分類樹，並可考量是否將中國大陸地圖各省畫分分類作為情資蒐集分析。

提供國際及台灣地區新聞總覽：使用者可點選世界各國以及台灣各縣市，以綜觀各地區之事件變化。

提供事件分類瀏覽：依據各使用者在業務上所需要之分類節點，再根據國家、台灣各縣市，以進行交叉分析比對。如圖 3-25：



圖 3-25：台灣地區地理事件分析圖參考

3.3.4.2 時間事件分析

在資料的蒐集分析上，【時間】因素也是一個重點的面向，主要可依據主題事件及時間將未來可能發生的事件，進行系統化管理，並可針對不同的主題，進行時間序列上的比較，以利決策支援。

此外，從【時間】方向來分析事件，我們可以充分掌握事件的發展歷史與過程，例如選情的變化、事件處理的變化、民意調查的變化以及人民對於政府決策的滿意變化等。如圖 3-26：

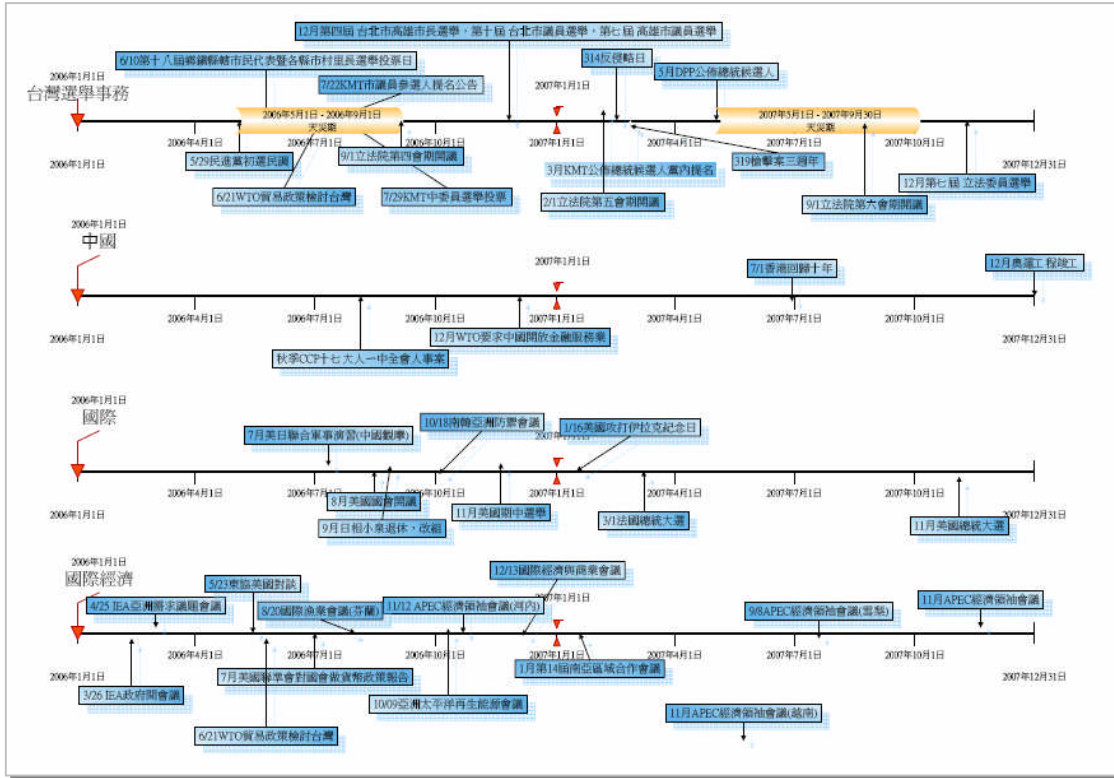


圖 3-26：時間與事件發展關係過程參考圖

3.4 方案可行性分析

依據考試院「文官制度興革規劃方案」報告，從政治可行性、行政可行性、時程可行性、經濟可行性、法律可行性、技術可行性，做為所提之方案可行性探討建置該資料庫平台是否可行。

3.4.1 政治可行性

本系統資料庫建置在於蒐集整合文官體制相關資料庫，提供具有價值資料，將來可作為提供跨部門機關查詢使用，協助公務人員面臨重大事項發生時，作為決策分析參考，可有效提升行政部門競爭力並擴大知識效益，應可獲得多數機關認同及支持。

3.4.2 行政可行性

本系統資料庫建置經過評估，主要整合現有機關已建置完成資料庫，並公開可透過網路查詢、取得、交換等，部份內部網路資料庫可透過技術交換方式整合，並採不重新建置及修改現有跨部門機關資料庫平台及架構方向規劃，因此不會增加耗費太多行政部門人力投入。

3.4.3 時程可行性

如果研究報告經討論及評估方向正確，且跨部門機關又可以全力配合提供開放資料庫整合，未來建置時程預估可在六個月左右完成系統建置及上線。

3.4.4 經濟可行性

經評估未來系統建置，主要導入透過搜尋引擎技術整合相關資料庫查詢並提供資料分析工具，其他可透過連結方式連結至各獨立資料庫網站，因此可降低絕大多數資料庫建置經費。

3.4.5 法律可行性

本資料庫來源主要整合現有跨部門機關相關資料庫網站形成知識庫，因此絕大多數屬於公部門自己建立資料，並無涉及法律規定，未來如要蒐集其他非公部門或是其他國家資料擷取，必須考慮是否涉及法律保護及智慧財產，除可透過發文獲得許可之外，並可考慮是否需要付費以取得有價值之資料整合。

3.4.6 技術可行性

未來除行政部門需配合相關資料庫開放取得授權，如內部未公開資料庫則必須考慮協商如何透過通訊協定方式取得，其他系統建置所使用技術皆為目前現行所使用技術皆可完成建置，可請資訊室與廠商協助洽詢。

肆、預期效益

- (一) 運用文字探勘(Text Mining)技術，從大量的資料中發掘資訊特徵及關聯性等有用資訊，以協助分析人員在知識萃取、知識管理與決策輔助等相關需求提供研究與協助。並在有限的人力與時間下，快速的掌握大量資料背後所蘊含的意義與趨勢徵兆，以協助政策的研擬與制定而達成其預期效益。
- (二) 善用資訊科技，強化搜尋功能機制，運用輔助工具包括：關鍵字、動態提示詞、關聯性以文找文、關聯性地圖、地理事件分析、時間事件分析等，可簡化快速搜尋及提高事件分析準確性，達成資料庫完整性功能效益。
- (三) 整合現有跨部會文官制度相關資料庫，成為知識累積與擴散之平台，以知識分享與知識傳承為原則，提升公務人員素質及績效，使其成為文官制度知識入口整合之標竿，節省資料查詢所需耗費時間。此外，以整合目前已建置資料庫為主要規劃，可有效節省經費，充份發揮最大經濟效益。

伍、研究結果分析

本研究主要從文官制度探討，分析資料庫建立的價值與可行性，以下從制度面與需求面作為研究結果分析。

我國文官制度資料庫主要分散在考試院暨所屬機關，如：考選部、銓敘部、公務人員保障暨培訓委員會、公務人員退休撫卹基金監理委員會，及受監督機關：行政院人事行政局。在權責劃分之下所建立的資料庫也是各別獨立，因此往往需要透過不同的機關網站才能查詢所要的資料。

知識除了要經過創造、分類、儲存、整合及分享外，還要不斷更新，才能產生價值，此外知識的價值在於如何從眾多的資料中經過整理分析並供應關鍵性資料，成為一個決策者的輔助分析來源，進而創造最大的競爭優勢。

依照目前各機關所建立的資料庫，如果不能整合就失去知識的價值，因為文官制度必須包括考試院暨所屬機關才算完整，相對文官制度資料庫的完整性必須涵蓋所有分散在各機關的資料庫，本研究報告首重在於規劃整合分散在各機關的異地資料庫，透過資訊技術的應用完成研究目的。

經透過文獻研究以及參考其他機關已建置的相關資料庫發現，「文字探勘」(Text Mining)技術在學術界已有深入研究的基礎(曾元顯，文字知識探勘與自動化資訊組織研究成果)，更在政府部門(行政院施政整合檢索系統)有廣泛的應用。

由於文官制度的資料庫需求涵括相當大量的政策資訊，包括人事、法規、期刊、論文(研究報告)、新聞、公報、研究計畫、各國政策訊息與政策白皮書等，如能在文官制度資料庫引進文字探勘技術，可在大量的文字資料中挖掘出

有價值的知識，可協助決策人員、政策研究人員或幕僚人員更快速地掌握全球性文官制度趨勢。

此外，多元化的搜尋工具，有助於在資料庫查詢中提升精準的資料查詢，並縮短大量文字比對時間，提升整體資料庫效能，此為資料庫建置的必要條件。

透過資料蒐集系統的建置，如能針對國外文官制度相關資料蒐集，可強化擴大資料庫的知識價值，則文官制度資料庫更具完整性。

陸、結論與建議

6.1 結論

近年來面對民眾積極要求政府提高行政效率的強烈期待，在全球化競爭壓力的推波助瀾下，唯有積極落實文官制度之興革，才能真正提昇競爭優勢，透過文官制度資料庫系統的建置，從而快速提昇行政效率、國家競爭力，以符合民意的期待。

經由上述各項研究說明，基於資料庫的價值轉換成為知識價值，文官制度資料庫的整合，在現行的制度中實有長遠規劃與建置的必要性。本研究初步完成第一階段評估報告，至於各項功能詳細規劃與效能評估，後續，請資訊單位更為深入探討，並洽詢相關廠商協助規劃與建置。

6.2 建議

- (一) 文官制度資料庫的建置，必須透過跨部會機關協力合作才能建置完成，未來研擬規劃完善措施，有賴各行政部門充分合作。
- (二) 基於資料庫的安全考量及管理維護，請資訊單位研擬規劃完善的安全機制。
- (三) 考量整體系統效能，目前考試院設備主機是否足夠因應協助完成建置，是否需要擴充主機或是提升硬體規格。
- (四) 文官制度資料庫建置完成，需要研擬一套使用及操作規範。

參考書目

中文部分

1. 考試院 (2009)。文官制度興革規劃方案。
2. 桂思強 (2003)。資料庫理論與 Access 範例教本，學慣行銷股份有限公司。
3. 張真誠、蔡文輝、林敏惠 (2003)。挑戰資料庫管理系統，旗標出版股份有限公司。
4. 黃三益 (2004)。資料庫的核心理論與實務，台灣東華書局股份有限公司。
5. 鄧承根 (2006)。專案工程監造作業資料庫建構及管理之研究--以軍事工程為例。國立中央大學土木工程研究所碩士論文。
6. 田子龍 (2006)。組織人事技能資料庫之設計-以 HR 部門為例，國立中央大學人力資源管理研究所碩士論文。
7. 曾元顯 (2002)。「文件主題自動分類成效因素探討」。中國圖書館學會會報，第 68 期，頁 62-83。
8. 曾元顯 (2002)。「數位文件之資訊組織與主題分析自動化之技術與應用」。台北市立圖書館館訊，第 20 卷，第 2 期，頁 23-35。
9. 曾元顯 (2004)。「文字知識探勘與自動化資訊組織研究成果」。
10. 曾元顯 (1997)。「關鍵詞自動擷取技術與相關詞回饋」。中國圖書館學會會報，第 59 期，頁 59-64。

英文部分

1. Christopher J.M., Philip K.C., Gregory (1993), "system for Knowledge Discovery in Database" , IEEE Transactions on Knowledge and Data Engineering; Vo1.5, No.6 December.
2. ED Title (2005) , "Managing Data with Database Tools" , Certification Managing, August.