

中華大學資訊工程學系
101 學年度專題製作期末報告

網頁知識之智慧資產化系統

指導老師:曾秋蓉

組員：

B09802021 林剴業

B09802036 葉諺謙

B09802054 曾健祐

B09802206 謝孟宇

專題編號：PRJ2012-CSIE-10104

中華民國一百零二年六月二十四日

目錄

1.	摘要.....	2
2.	研製背景&開發目標.....	3
2.1	研製背景.....	3
2.2	開發目標.....	4
3.	開發環境.....	5
4.	時程規劃.....	6
5.	研究方法.....	8
5.1	Html Agility Pack.....	8
5.2	Regular Expression.....	9
6.	專題內容.....	11
6.1	系統架構.....	11
6.1.1	知識檢索模組.....	12
6.1.2	文字段落萃取模組.....	12
6.1.3	圖片萃取模組.....	12
6.1.4	表格萃取模組.....	12
6.1.5	圖/表與文字段落關聯模組.....	12
6.2	流程圖.....	13
6.3	資料庫 ER.....	15
7.	遇到困難與解決方案.....	19
8.	完成之工作項目及具體成果.....	20
9.	未來研究方向.....	24
10.	銘謝.....	24
11.	參考文獻.....	24
12.	安裝與教學.....	25

1. 摘要

近年來，知識管理活動廣泛的被一般公司推行與應用。知識管理的目的除了產生及分享知識外，協助公司同仁解決問題也是其中一個議題，藉由知識管理來提升公司內同仁解決問題的效率。當同仁在解決問題時，除了公司內部知識管理系統中的知識外，公司內部儲存的知識文件，如結案報告也都是寶貴的知識資產，可被用於解決同仁的問題。然而，公司中現存的知識素材並不一定足以協助同仁解決問題，或是可能有知識素材過舊不符合時效性的問題。由於網際網路的興起，網際網路中充滿了大量資訊，其中不乏有不同類型的網頁，這些網頁中包含的知識素材能夠作為同仁在解決問題時之參考依據。在過去，當公司內同仁利用網站的搜尋功能時，其搜尋結果常參差不齊，使用者需在大量搜尋結果中一一過濾，才能夠獲取所需之資訊，浪費了寶貴的時間，降低了行事效率。為了使使用者可以快速的從網際網路中找到有助於解決問題的知識素材，我們將運用自動知識萃取技術從網際網路中蒐集網頁資料並進一步分析內容，從中萃取出文字段落、圖片及表格。當使用者遇到問題需要解決時，便能夠更快找到與其問題相關的資料，進而提升問題解決之效率。

關鍵字：知識管理、知識萃取、圖片萃取、表格萃取。

2. 研製背景&開發目標

2.1 研製背景

在競爭激烈的環境裡，公司總會希望在短時間內獲得最大效益，然而公司在訓練新員工時需耗費許多時間與金錢，不論是公司規定、公司環境、硬體操作等等都需要有經驗的員工帶領或者等待員工歷經許多經驗從中學習，但在這期間往往會耗掉大量時間。此外，當員工碰到問題時，往往需到公司的知識管理系統中發問，被動的等待有經驗的同仁幫忙回答問題，在等待的過程中往往錯失了問題解決的時效性。另一方面，由於公司中儲存著大量過去累積的知識文件，員工可從過去的文件中找尋所需的資訊，但想從現有文件裡取得需要的資源相當的耗力費時，員工除了需花費時間找尋可能的知識文件外，還需花費大量時間於過濾知識文件以便從中獲取相關資訊。

由於上述問題，在傳統被動式的知管理模式中，當員工遇到緊急需求時無法有效率的得到支援。相較以往被動式的知識管理模式，過去的學長姐們的專題已解決了上述問題。他們所建置的知識管理系統採取主動式的問題解決模式，主動式的問題解決不但可自動萃取大量文件的圖、文、表資料，還可讓使用者在短時間內透過知識管理系統獲取需要的知識，並且從圖文並茂的資料中取得所需的資訊。不但省下員工尋找資料花費的時間，提供的知識也較準確且簡潔。

然而，過去學長姐們的專題雖能主動提供圖文並茂的資料供員工用於問題之解決，但所能取得的資源依然有限。除了知識管理系統中過去累積的經驗外，公司內的知識文件庫可能還是不足以提供所有員工所需的資訊，還可能會有資料較舊的問題。近年來由於資訊技術的發達，想從網際網路中獲取資訊已是一件非常容易的事情，但網際網路搜尋的結果往往龐大且需花費大量時間來整理，所以本組研製「網頁知識之智慧資產化系統」，從網頁中分析及萃取文字段落、圖片及表格，並將結果顯示出來，員工只要在我們建立的網路平台裡輸入關鍵字，就可透過系統從網際網路中快速的搜尋相關文字段落、圖、表，不但節省大量時間也提升解決問題的效率。

2.2 開發目標

本組專題將焦點放在網頁的自動知識萃取及知識檢索。透過文字、圖片及表格的分析，將網頁中的知識素材自動萃取出，並提供給使用者參考。當公司內部人員遇到問題時，可即時獲得系統的協助，找到解決方案。本專題開發目標如下：

(1) 研製知識萃取相關功能模組

使用 Microsoft Visual Studio 2010 研製網頁之知識萃取相關功能模組，萃取文件中之文字段落、圖片及表格並建立文字段落與圖表關聯，以提供圖文並茂的知識素材。

(2) 研製知識檢索相關功能模組

提供使用者輸入關鍵字檢索相關知識之功能，自動從網際網路中取得相關之文字段落、圖片與表格供使用者參考。

(3) 整合知識萃取與知識檢索

整合各模組，建構本專題研發之網頁知識之智慧資產化系統。

3. 開發環境

開發平台：

Windows 7

開發工具：

Microsoft Visual Studio 2010

Microsoft SQL Server 2008 R2

Internet Information Services (IIS)

程式語言：

Visual Basic

ASP.NET 4.0

4. 時程規劃

表 4.1 及圖 4.1 為本組之時程規劃，詳細說明如下：

表 4.1 時程規劃表

完成項目	開始日期	結束日期	總共使用時間
1. VB 語法的研讀	2012/03/01	2012/06/30	四個月
2. 資料蒐集	2012/05/01	2012/08/31	四個月
3. 系統架構規劃	2012/07/01	2012/09/31	三個月
4. 文字段落萃取模組開發	2012/09/01	2012/11/30	三個月
5. 表格萃取模組開發	2012/10/01	2012/12/31	三個月
6. 圖片萃取模組開發	2012/11/01	2013/01/31	三個月
7. 圖表標題及上下文關聯建立模組開發	2012/12/01	2013/02/28	三個月
8. 知識萃取模組測試及修正	2013/02/01	2013/03/31	兩個月
9. 知識檢索模組開發	2013/02/01	2013/04/30	三個月
10. 使用者平台建置	2013/05/01	2013/05/31	一個月
11. 檢索模組測試及修正	2013/05/01	2013/06/24	兩個月
12. 期末報告製作	2013/06/01	2013/06/24	一個月

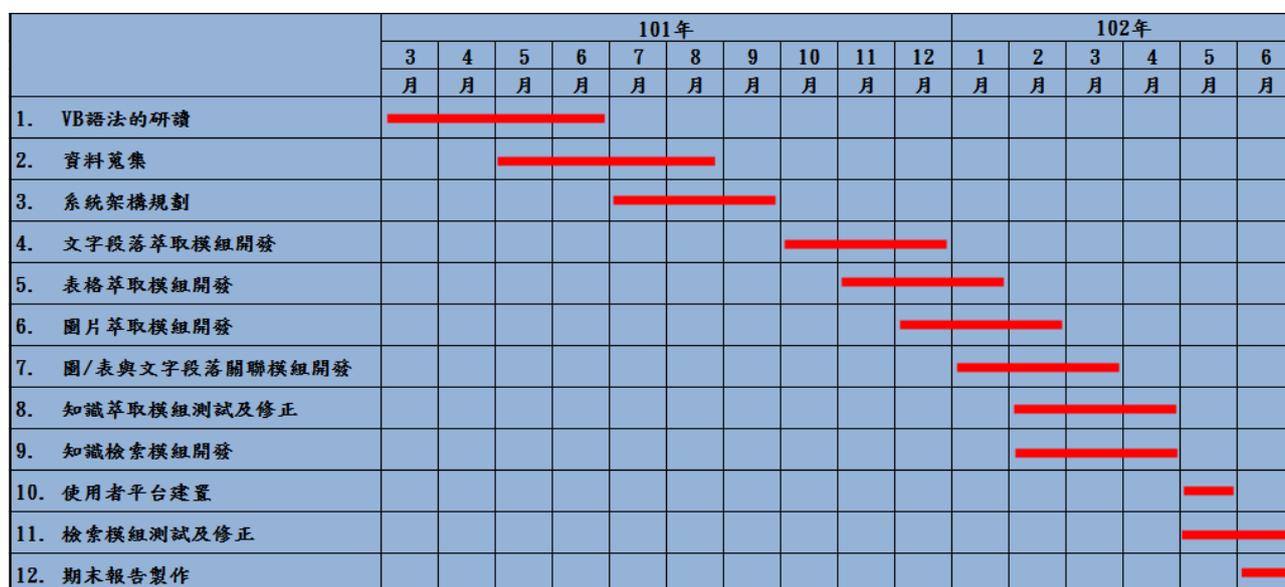


圖 4.1 時程規劃甘特圖

工作項目說明：

1. VB 語法研讀：本專題系統開發使用語言為 VB，因此需對語法之應用熟悉。
2. 資料蒐集：蒐集相關資料，包含如何分析網頁等。
3. 系統架構規劃：規劃系統整體架構，包含各系統模組之功能。
4. 文字段落萃取模組開發：由於 Html 網頁中有著許多不同的標籤，其中有些分類是專屬於文字使用的，所以需透過演算法將文字段落萃取出來。
5. 表格萃取模組開發：將網頁中使用表格的標籤及其內容透過演算法萃取。
6. 圖片萃取模組開發：將網頁中使用圖片的標籤及其內容透過演算法萃取。
7. 圖/表與文字段落關聯模組開發：將圖/表與文字段落建立關聯，以提供圖文並茂的知識素材。
8. 知識萃取模組測試及修正：針對知識萃取模組進行測試及修正。
9. 知識檢索模組開發：提供使用者輸入關鍵字檢索相關知識之功能，自動從網際網路中取得相關之文字段落、圖片與表格供使用者參考。
10. 使用者平台建置：利用 ASP.NET 建置使用者介面，供使用者進行知識檢索及知識素材瀏覽。
11. 知識檢索模組測試及修正：針對知識檢索模組進行測試以及修正。
12. 期末報告製作：整理專題的製作過程中蒐集之相關資料與紀錄，編撰成期末報告及簡報。

5. 研究方法

本專題製之相關研究方法說明如下：

5.1 Html Agility Pack

HTML Agility Pack 是由法國的一位軟體架構師 Simon Mourier 所發展，並且由 DarthObiwan 以及 Jessynoo 輔助開發出來的一個軟體工具，它可以讓剖析鬆散格式 HTML 的工作就像剖析 XML 一樣簡單，它也有類似於 System.Xml 命名空間中的 XML DOM 的許多類別，除了可以使用階層的方式存取 HTML 以外，它也支援使用 XPath 的方式來搜尋 HTML。相較於以往使用正規表示式(Regular Expression)過濾與擷取資訊這樣一個非常繁瑣的過程，Html Agility Pack 較以往使用文字比對或是 Regular Expression 的比對方式來得更精確。

利用 Html Agility Pack 分析 HTML 及擷取資料可分為四個步驟：

- (一) 利用其所提供之方法來擷取網頁。
- (二) 獲取網頁所有節點資料。
- (三) 設定所需之分析資料節點。以擷取表格節點資料為例(如圖 5.1)，可透過指定資料節點擷取其內容並儲存。

```
Private XPathOfTableTags As String = "//table | //TABLE"
```

圖 5.1 XPath 使用範例

- (四) 利用 GetAllNeededElement 方法來擷取所有所需節點資料，如圖 5.2

```
Dim htmlNodes As HtmlNodeCollection = GetAllNeededElement(htmlBody, XPath)
```

圖 5.2 使用 GetAllNeededElement 方法抓取所有節點

在擷取完所需之節點資料後就可以進行個別的使用了。利用 Html Agility Pack 來剖析網頁可以節省許多資料過濾的時間，且實作時非常具有彈性。

5.2 Regular Expression

Regular Expression 正規表示式，又稱正則表達式、正規表示法、常規表示法（英語：Regular Expression，在代碼中常簡寫為 regex、regexp 或 RE）。正規表示式使用單個字串來描述、匹配一系列符合某個句法規則的字串。在很多文字編輯器裡，正則運算式通常被用來檢索、替換那些符合某個模式的文字。簡單敘述而言，正規表示法就是處理字串的方法，是以行為單位來進行字串的處理行為。正規表示法透過一些特殊符號的輔助，可以讓使用者輕易的達到「搜尋/刪除/取代」某特定字串的處理程序。正規表示法基本上是一種「表示法」，只要工具程式支援這種表示法，那麼該工具程式就可以用來作為正規表示法的字串處理之用。

在編寫處理字符串的程序或網頁時，經常會有查找符合某些複雜規則的字符串的需要。正則表達式就是用於描述這些規則的工具。換句話說，正規表示式就是記錄文本規則的代碼。本專題製作最初是利用正規表示式來研製網頁中圖、文、表的內容萃取功能，後因使用 Html Agility Pack 省去了大量設計正規表示式的時間。儘管如此，有少部分僅需單純的設計功能還是以正規表示式進行資料擷取。以擷取 Google Search 所搜尋出的網頁為例，如圖 5.3。



The image shows a Google search result for 'HTC'. The search bar at the top contains 'HTC' and a magnifying glass icon. Below the search bar, there are navigation tabs for '網頁', '圖片', '地圖', '新聞', '更多', and '搜尋工具'. The search results show approximately 416,000,000 items found in 0.19 seconds. The first result is for '宏達電-HTC' with the URL 'www.htc.com/tw/'. Below this, there are several links: 'HTC', '支援服務', '客戶服務中心', '新 HTC One', '新品上市 HTC Desire U', and '產品'. To the right of the search results, there is a detailed information box for '宏達國際電子' (HTC Corporation), including its stock price (2498 TPE), market cap (NT\$150.00), and founding date (1997). The bottom of the image shows a section for '「HTC」的新聞搜尋結果' with several news snippets from various sources like '聯合新聞網' and 'NOWnews'.

圖 5.3 Google Search 回傳頁面

Google 的網頁格式非常的單純以及固定，檢視其原始程式碼後可發現，所搜尋到的網址都是存放在特定 Tag “<h3 class=...”裡(如圖 5.4)，因此就可以設計出一正規表示法來將搜尋結果的網頁網址進行擷取。

```
id="error_section" style="display:block;font-size:11px"></div></li></ul></li></ul></div></div>
</div></div><div id="appbar"><div id="extabar"><div id="topabar" style="position:relative"><div
class="ab_tnav_wrp" id="slim_appbar"><div id="sbfrm_1"><div id="resultStats">約有 416,000,000 項結
果<nobr> (搜尋時間: 0.19 秒)&nbsp;&nbsp;&nbsp;</nobr></div></div> </div></div><div id="botabar"
style="display:none"><div></div></div></div></div><div class="mw" id="ucs"></div><div
class="mw"><div id="rcnt" style="clear:both;position:relative;zoom:1"><div class="col"
style="width:0"><div id="leftnavc"></div></div><div class="col" style="width:0"><div
id="center_col"><div id="taw" style="margin-right:0"><div></div><div style="padding:0 8px"><div
class="med"></div></div></div></div><div class="med" id="res" role="main"><div id="topstuff"></div><div
id="search"><!--a--><h2 class="hd">搜尋結果</h2><div id="ires"><ol eid="Z94cUuDQDonyiAe_YFw"
id="rso"><li class="g" data-hveid="47"><div class="vk_c nmdc"><!--m--><link rel="prerender"
href="http://www.htc.com/tw/"><div class="rc" data-hveid="48"><span class="altcts"></span><h3
class="r"><a href="http://www.htc.com/tw/" onmousedown="return
rwt(this, '', '', '', '1', 'AFQjCNE5jmGmrF-
TAkjSHuGh2ekG41TpJw', 'e6RBHb1FhLmdrGtTeyhbUg', 'OCDEQFjAA', '', '', event)">宏達電- <em>HTC</em></a>
</h3><div class="s"><div><div class="f kv" style="white-space:nowrap"><cite>www.
<b>htc</b>.com/tw/</cite><div class="action-menu ab_ctl"><a href="#" class="clickable-dropdown-
arrow ab_button" id="am-b0" aria-label="搜尋結果的詳細資訊"
jsaction="ab.tdd;keydown:ab.hbke;keypress:ab.mskpe" role="button" aria-haspopup="true" aria-
expanded="false" data-ved="OCDIQ7B0wAA"><span class="mn-dwn-arw"></span></a><div class="action-
menu-panel ab_dropdown" jsaction="keydown:ab.hdke;mouseover:ab.hdhne;mouseout:ab.hdhue"
role="menu" tabindex="-1" data-ved="OCDMQgR8wAA"><ul><li class="action-menu-item ab_dropdownitem"
role="menuitem"><a href="http://weboache.googleusercontent.com/search?

```

圖 5.4 網頁原始碼截圖

正規表示式如下所示：

```
"<h3 class=+.+?r.+?><a.+?href=[\''"](?<href>.+?)[\''"]..+?></h3>"
```

應用於程式中範例如圖 5.5 所示：

```
Dim matches As System.Text.RegularExpressions.MatchCollection = _
System.Text.RegularExpressions.Regex.Matches(strContext, _
"<h3 class=+.+?r.+?><a.+?href=[\''"](?<href>.+?)[\''"]..+?></h3>)"
```

圖 5.5 正規表示式實際操作方法

6. 專題內容

6.1 系統架構

專題系統架構圖如圖 6.1 所示，包含知識檢索模組、表格萃取模組、圖片萃取模組、文字段落萃取模組及圖/表與文字段落關聯模組，各模組之詳細說明如下：

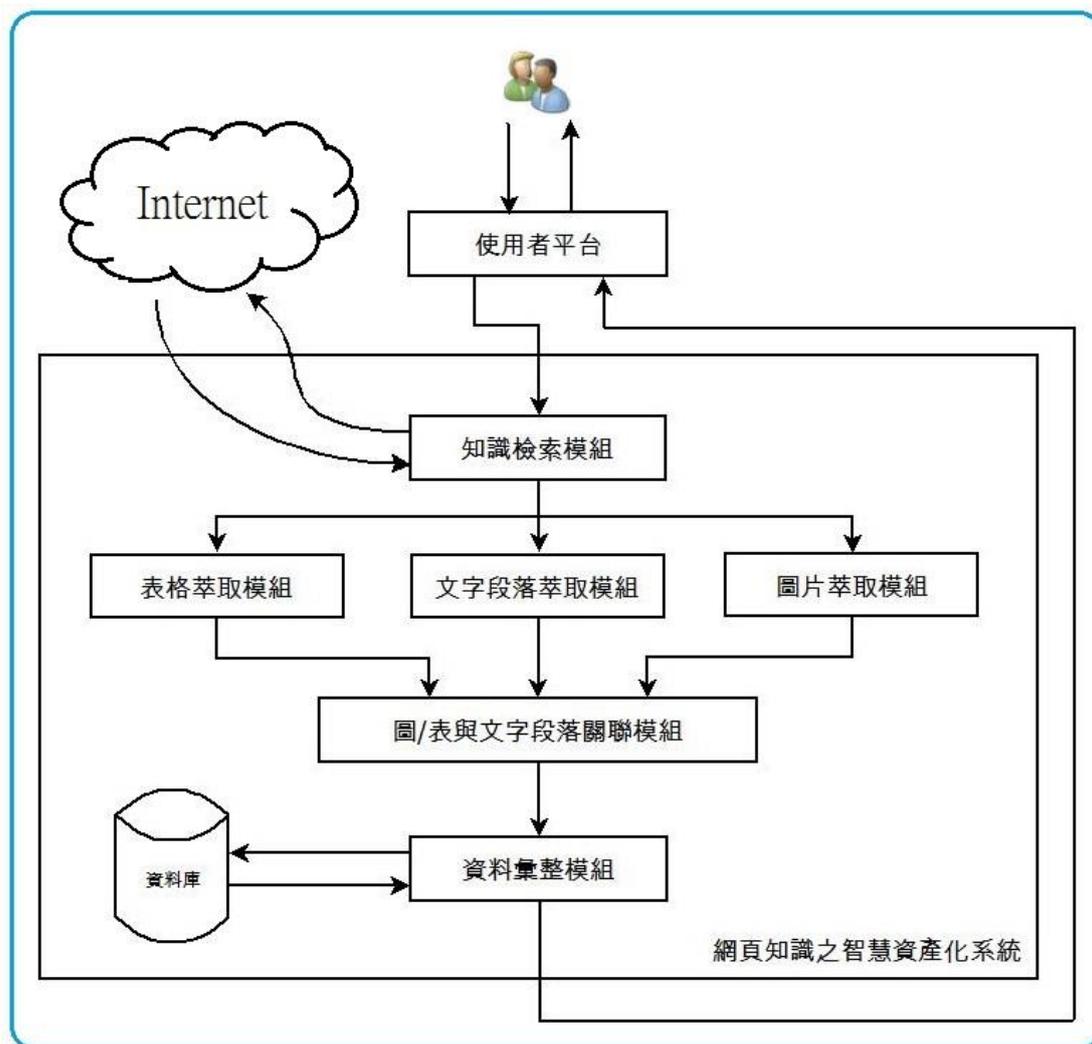


圖 6.1 系統架構圖

6.1.1 知識檢索模組

將使用者輸入之關鍵詞與網頁中包含之文字段落、圖片說明與表格說明進行比對，並以相似度排序，最後呈現相關知識素材提供給使用者參考。

6.1.2 文字段落萃取模組

從關鍵詞搜尋結果中包含的相關網頁中擷取出文字段落，並將文字段落內容暫存至資料庫，以供知識檢索模組進行知識素材比對，以及提供使用者瀏覽。

6.1.3 圖片萃取模組

從關鍵詞搜尋結果中包含的相關網頁中擷取出圖片及圖片說明，並將圖片及圖片說明暫存至資料庫，以供知識檢索模組進行知識素材比對，以及提供使用者瀏覽。

6.1.4 表格萃取模組

從關鍵詞搜尋結果中包含的相關網頁中擷取出表格及表格說明，並將表格及表格說明暫存至資料庫，以供知識檢索模組進行知識素材比對，以及提供使用者瀏覽。

6.1.5 圖/表與文字段落關聯模組

將文字段落與其引用之圖片及表格進行關聯建立。例如某文字段落中若提到「如圖 1...」，便將此文字段落與「圖 1」建立關聯；某段文字段落中若提到「如表 1...」，便將此文字段落與「表 1」建立關聯。當使用者瀏覽相關知識素材時，圖/表與文字段落關聯模組將呈現知識素材之關聯供使用者參考。

6.2 流程圖

本專題系統的流程圖如下圖 6.2 所示

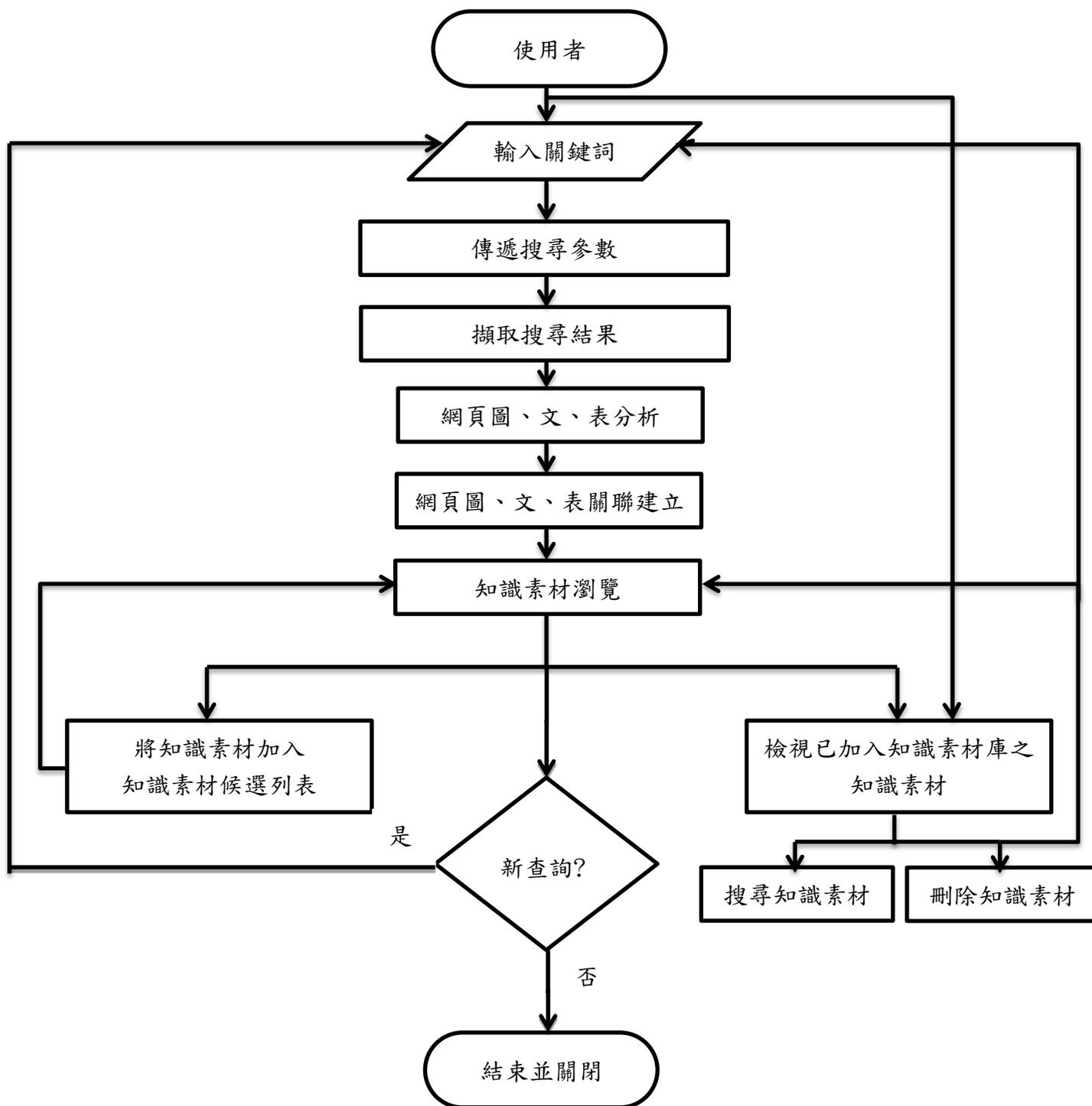


圖 6.2 系統流程圖

系統使用流程為以下七個步驟：

- (一) 系統登入，若無帳號可直接申請
- (二) 進行查詢時提供輸入欄位供使用者填入關鍵詞
- (三) 將關鍵詞紀錄並利用 Goolge 搜尋引擎進行搜尋
- (四) 將搜尋的網頁結果分別萃取出並分析圖、文、表
- (五) 分析後將資料的圖、文、表進行關聯的建立
- (六) 呈現結果給予使用者瀏覽，使用者可將較有價值的知識素材存入知識素材庫，使用者亦可檢視知識素材庫來看既有的知識素材
- (七) 使用者可繼續使用系統或是將網頁關閉來結束程式的運行

6.3 資料庫 ER

資料庫中包含網頁中圖、文、表的分析結果、知識素材庫中的圖、文、表素材、素材關連性紀錄和出處網址、使用者登入資訊等八個資料表。

1. 網頁分析後的圖片資訊

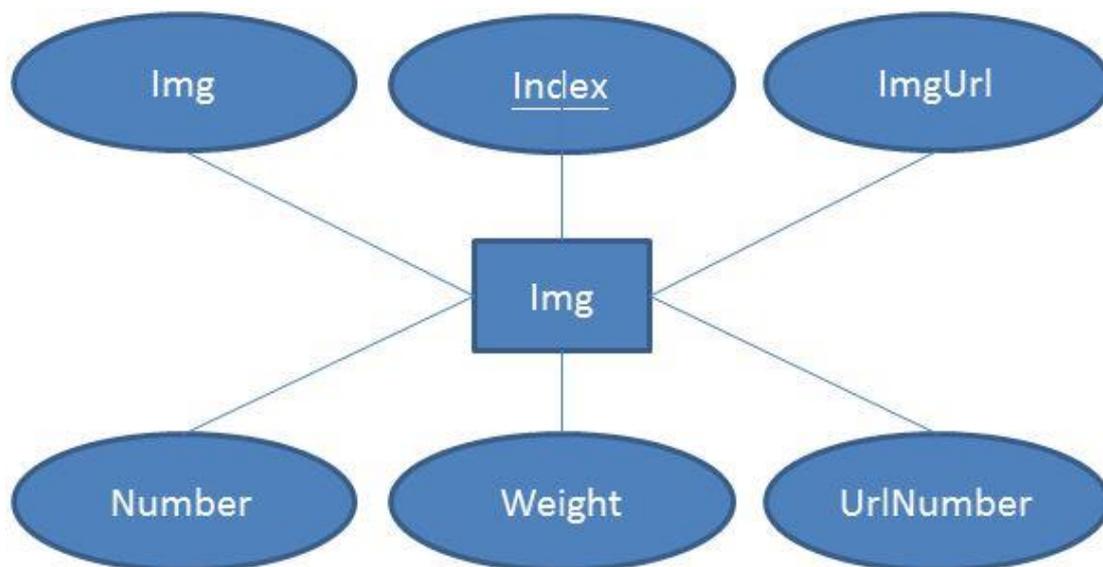


圖 6.3 網頁圖片分析 ER Model 圖

2. 網頁分析後的文字資訊

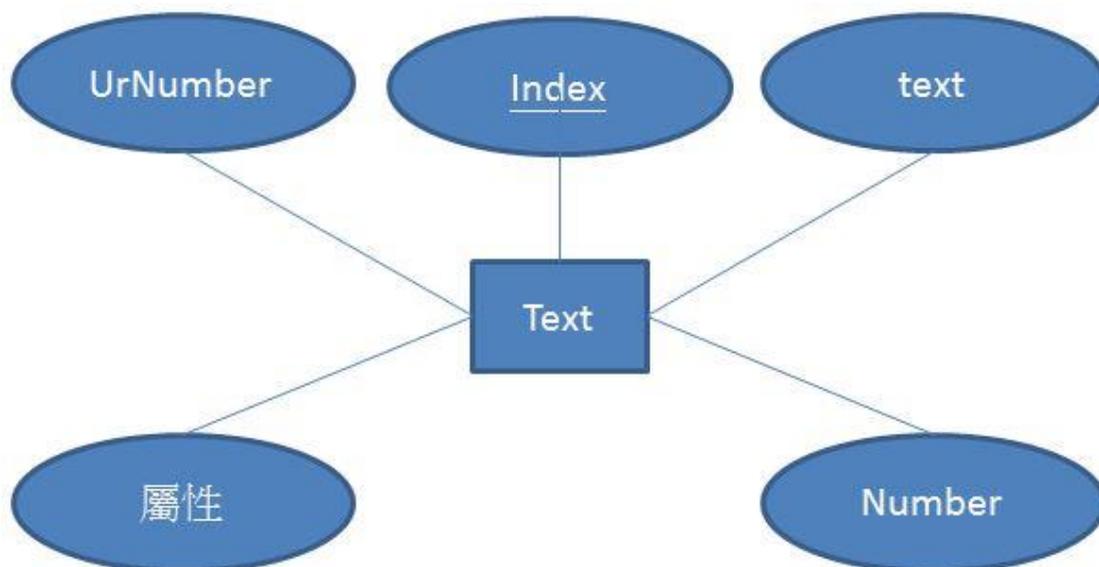


圖 6.5 網頁文字分析 ER Model 圖

3. 網頁分析後的表格資訊

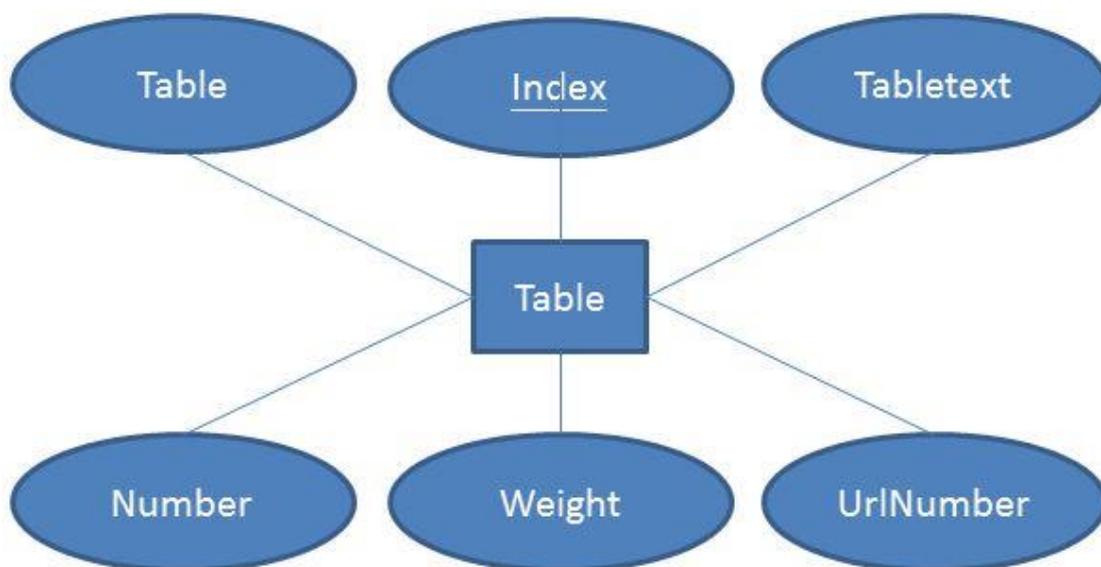


圖 6.6 網頁表格分析 ER Model 圖

4. 知識素材庫的圖片資訊

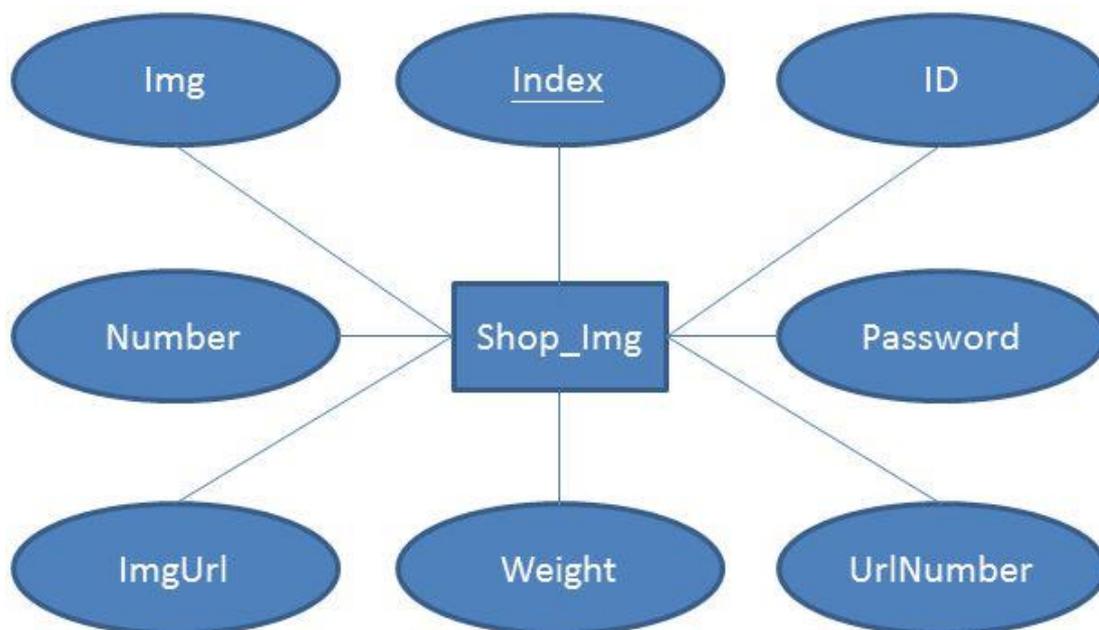


圖 6.7 知識素材庫圖片庫表 ER Model 圖

5. 知識素材庫的文字資訊

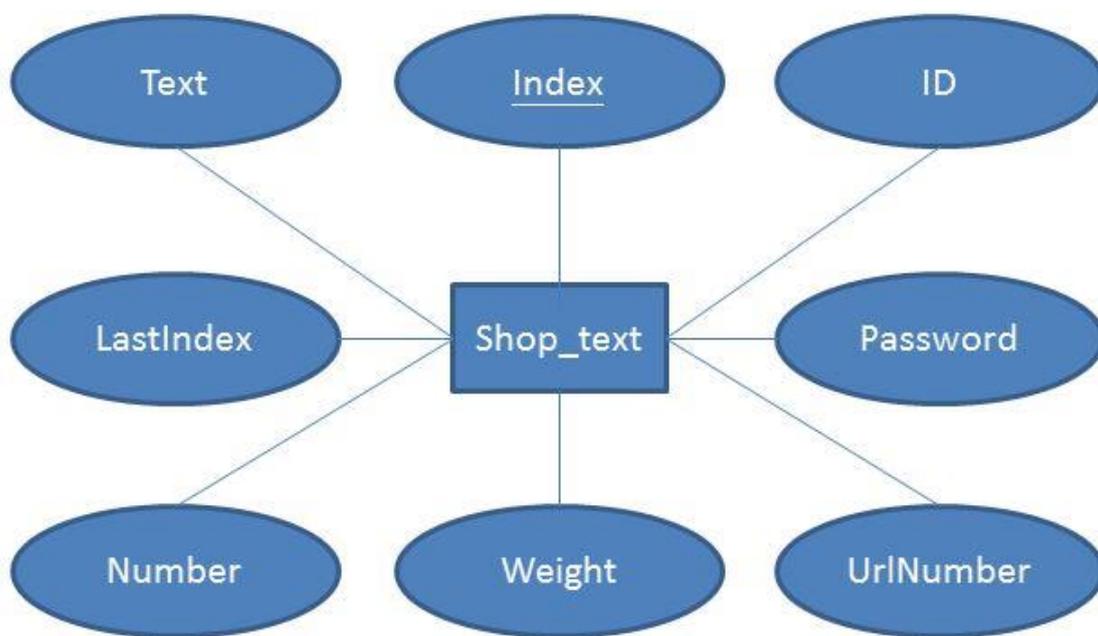


圖 6.8 知識素材庫文字庫表 ER Model 圖

6. 知識素材庫的表格資訊

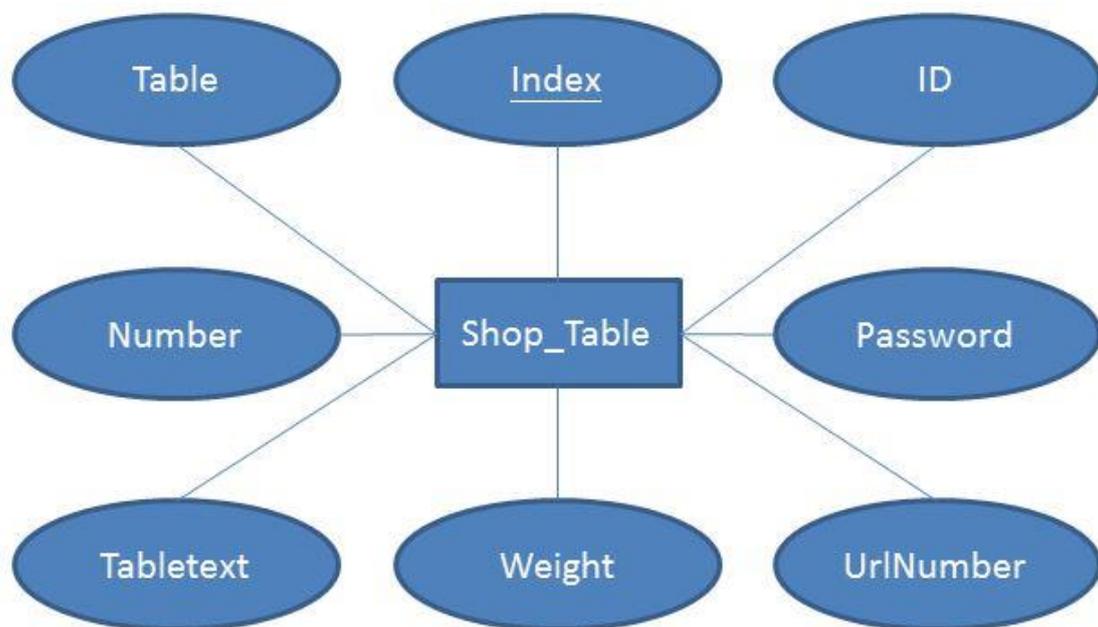


圖 6.9 知識素材庫表格庫表 ER Model 圖

7. 知識素材出處網址資訊

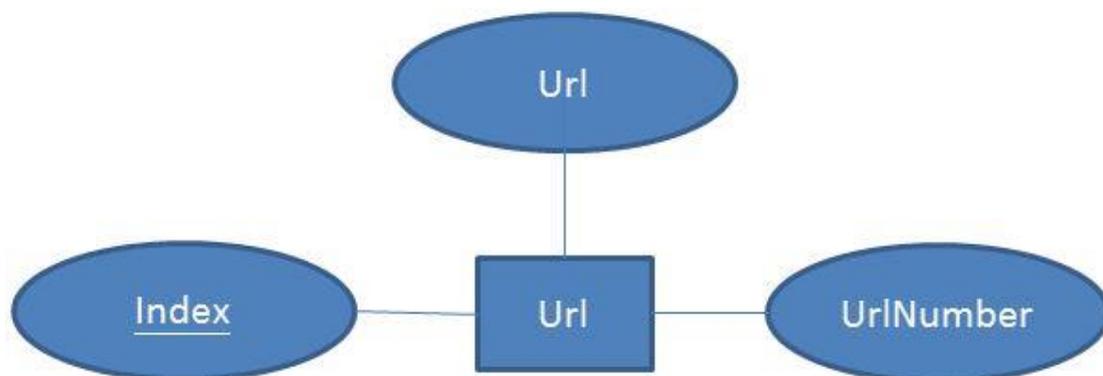


圖 6.10 知識素材出處表 ER Model 圖

8. 使用者帳號密碼管理

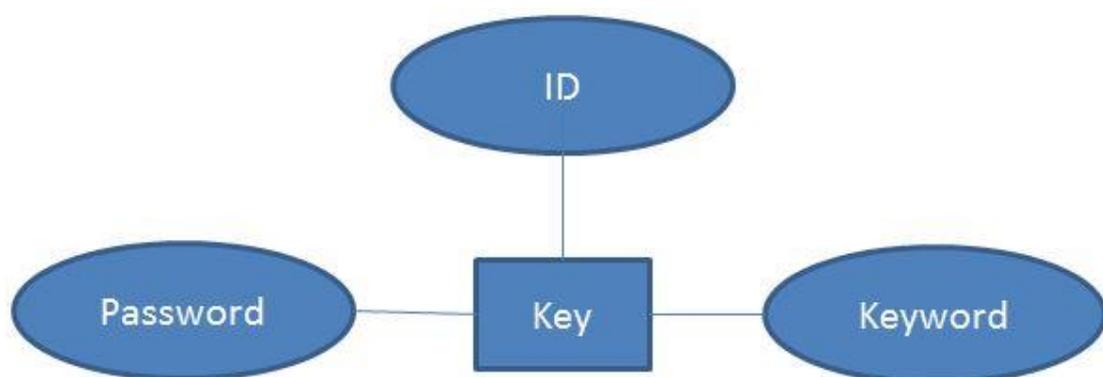


圖 6.11 使用者控管表 ER Model 圖

7. 遇到困難與解決方案

問題 1：

許多網頁的標籤裡不一定會附上<title>或是替代文字<alt>，因此無法將圖片與文字段落進行關聯分析。

解決方案：

搜尋圖片相鄰文字中，是否有文字內容類似「如上圖」、「如下圖」或「圖x」，藉以將圖片與文字段落進行關聯建立。

問題 2：

知識素材分析後呈現之表單分成「文字」、「圖片」、「表格」，其中，文字部分常因為字串長度過長，使系統無法正確地將知識素材呈現出來。

解決方案：

依據文字內容取前 50 個字元做為顯示的依據，並將文字嵌入超連結，讓使用者先瀏覽標題再決定是否繼續閱讀詳細內容。

8. 完成之工作項目及具體成果

工作項目：

- (一) 文字段落萃取模組之研製
- (二) 表格萃取模組之研製
- (三) 圖片萃取模組之研製
- (四) 圖/表與文字段落關聯模組之研製
- (五) 知識檢索模組之研製

具體成果：

本組研製「網頁知識之智慧資產化系統」，從網頁中分析及萃取文字段落、圖片及表格，並將結果顯示於使用者介面上，使用者只要在我們建立的網路平台裡輸入關鍵字，就可透過系統從網際網路中檢索相關文字段落、圖、表知識素材，不但節省大量時間並有效率的解決問題。相對於傳統被動的知識管理模式，本專題研發之系統提供主動式的知識管理模式，省去員工過濾大量網頁以及尋找相關知識的時間，並提供更精確及簡潔的知識檢索功能協助員工快速解決遭遇之問題。

本系統的使用者介面如下：

1. 登入畫面：提供系統使用者登入，如圖 8.1，若無帳號則可註冊新帳號，如圖 8.2。

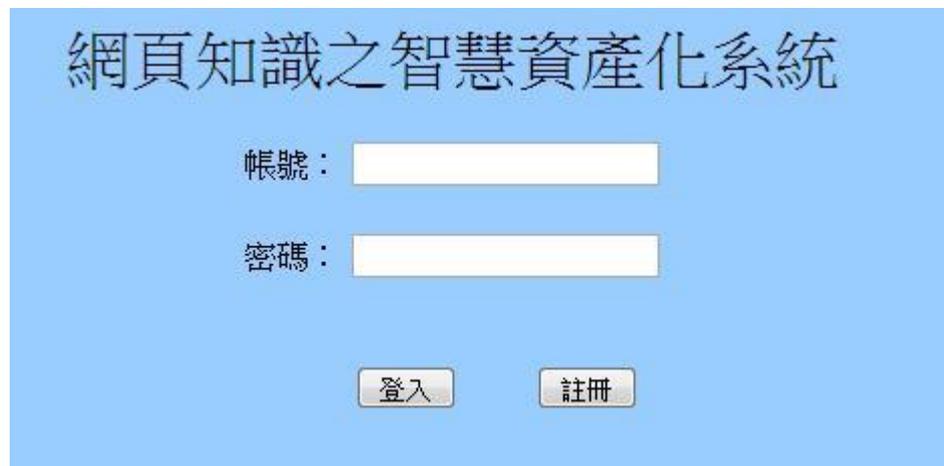


圖 8.1 系統登入畫面

圖 8.2 新辦帳號畫面

2. 使用者介面：提供系統使用者輸入關鍵詞開始搜尋及分析，如圖 8.3。搜尋後結果將呈現所有經過分析後的知識素材，內容包含(1)文字；(2)圖片；(3)表格。使用者可選擇顯示內容，如圖 8.4。使用者可點選「相關文字段落之摘要」、「圖片說明」、「表格說明」，來瀏覽資料，如圖 8.5，若使用者覺得某知識素材具有參考價值，可以將該關鍵詞加入至知識素材候選列表並檢視，如圖 8.6。在知識素材候選列表中可以做篩選的動作來避免使用者加入大量且不同類型的知識素材後導致資料雜亂，如圖 8.7。

圖 8.3 使用者介面

歡迎 jeffate 使用 登出

輸入關鍵字:

顯示資料類型:

編號	相關文字段落之摘要
<input type="button" value="加入知識素材候選列表"/> 55681	當微軟準備開發一種新的編程工具的時候，第一決定就是利用最搶手的 VB6 來進行修改，或者就是重新組
<input type="button" value="加入知識素材候選列表"/> 55659	1995年8月，VB4.0發行了32位元版本和16位元的版本。其中包含了對類別的支援。從VB4開始
<input type="button" value="加入知識素材候選列表"/> 55703	VB是一種充滿了爭議的語言；很多程式設計師對VB程式的品質有強烈的反感。很多人認為VB不配他們使用
<input type="button" value="加入知識素材候選列表"/> 55704	一些批評家認為VB的簡單特性使得其在未來具有傷害性。很多人自學了VB，但是並沒有學到好的編程習慣。
<input type="button" value="加入知識素材候選列表"/> 55707	但是很多VB的支持者解釋說VB的易用性就是它最大的優勢，可以讓經驗豐富的VB程式設計師或是剛剛懂得
<input type="button" value="加入知識素材候選列表"/> 55766	本站於二零零零年創辦，以分享VB6及API程式寫作的範例、技巧為宗旨&
<input type="button" value="加入知識素材候選列表"/> 55683	VB的中心思想就是要便於程式設計師使用，無論是新手或者專家。VB使用了可以簡單建立應用程式的GUI
<input type="button" value="加入知識素材候選列表"/> 55695	很多程式編寫者也會留戀VB6。VB6仍保留了一些script語言的特性，例如：不用事先宣告變量類型
<input type="button" value="加入知識素材候選列表"/> 55699	這樣大的變革一度被認為是微軟為了延長VB的生命力而做出的，結果把VB改得幾乎面目全非，成了一個類別
<input type="button" value="加入知識素材候選列表"/> 55700	2005年，微軟宣布將不會再對非.NET版本的VB進行支援。VB社群立即作出反應表示關心這個訊息。

1 2 3 4 5

編號	圖片說明
<input type="button" value="加入知識素材候選列表"/> 4576	270個VB實例學API
<input type="button" value="加入知識素材候選列表"/> 4577	你是一位剛接觸程式撰寫的學生嗎? 或是你是一位自學的程式撰寫新手? 想要在家自己寫程式，卻對於Vi

編號	表格說明
<input type="button" value="加入知識素材候選列表"/> 1639	Visual Basic (VB) 是由微軟公司開發的包含協助開發環境的事件驅動程式語言。它源自於BA

圖 8.4 搜尋結果

此文字段落來源為:
https://zh.wikipedia.org/zh-tw/Visual_Basic

當微軟準備開發一種新的編程工具的時候，第一決定就是利用最搶手的 VB6 來進行修改，或者就是重新組建工程開發新工具。微軟後來開發了 VB的繼任者Visual Basic .NET，同時也是.NET平台的一部分。VB.NET程式語言是一種真正的物件導向程式語言（VB4-VB6 只能稱為基於物件而不是物件導向，最明顯的特徵就是 VB4-VB6 都不支援類別的繼承），和傳統 Visual Basic 並不完全相容。

[瀏覽此文字段落中引用之圖片/表格](#)

圖 8.5 瀏覽相關文字之摘要

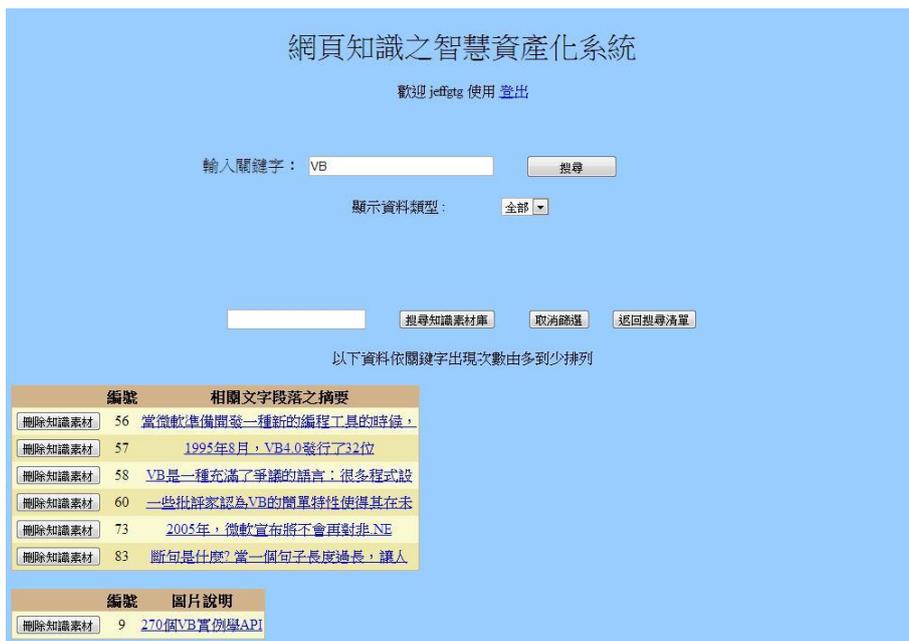


圖 8.6 檢視已儲存之知識素材

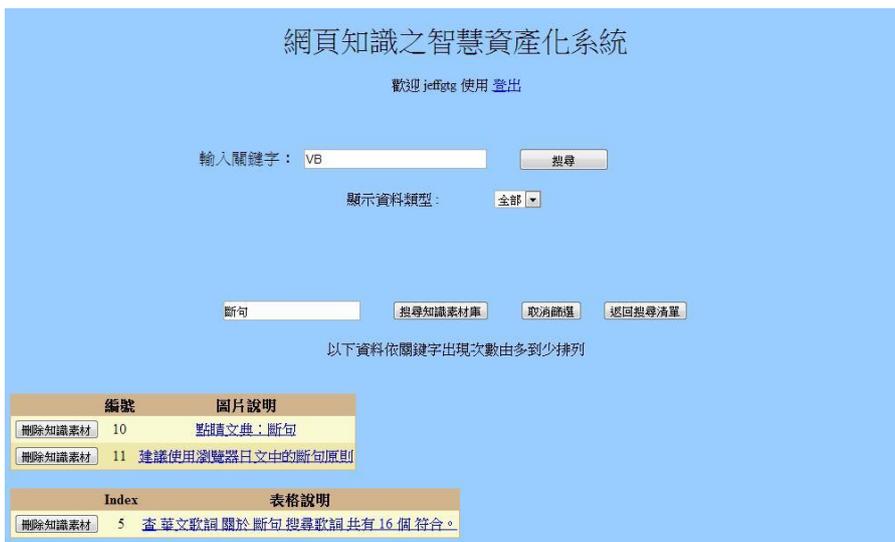


圖 8.7 知識素材候選列表篩選功能

9. 未來研究方向

雖然目前系統已依規劃之功能及時程完成建置，但仍有可進一步優化的地方，未來我們將進一步進行系統優化：

1. 優化圖/表與文字段落關聯模組精準度：

進一步提升資料分析及關聯建立的精準度，使分析後之知識素材更貼近使用者需求。

2. 優化系統執行效率

目前系統執行所耗費時間會因輸入關鍵詞的不同而有差異，未來希望能夠進一步提升系統執行效能。

3. 更精準相似度權重計算

目前僅以較簡單之權重計算方式來分析網頁中知識素材的權重，未來將以更為準確的方式來計算權重。

10. 銘謝

感謝曾秋蓉教授與吳智瑋學長的幫助及指導，在製作專題的過程中給予幫助及建議，在碰到困難時，熱心與耐心兼備的教導我們，幫助我們解決問題。謝謝老師以及學長。

11. 參考文獻

- [1] ISBN: 9789862760024 - ASP.NET 4.0 使用 VB 完美入門 奚江華 碁峯
- [2] ISBN: 9789862764695 - ASP.NET 4.0 專家技術手冊 I 奚江華 碁峯
- [3] ISBN: 9789867309662 - ASP.NET 專題與實務 I 4.0 周棟祥、吳進魯 松崗
- [4] ISBN: 9789572238530 - ASP.NET 專題與實務 II 4.0 周棟祥、吳進魯 松崗
- [5] ISBN: 9787302263746 - ASP.NET 安全編成入門經典 Barry Dorrans 清華大學出版社(譯)
- [6] ISBN: 9861258965 -.NET 網路與 I/O 技術手冊 康廷 數位工坊 松崗

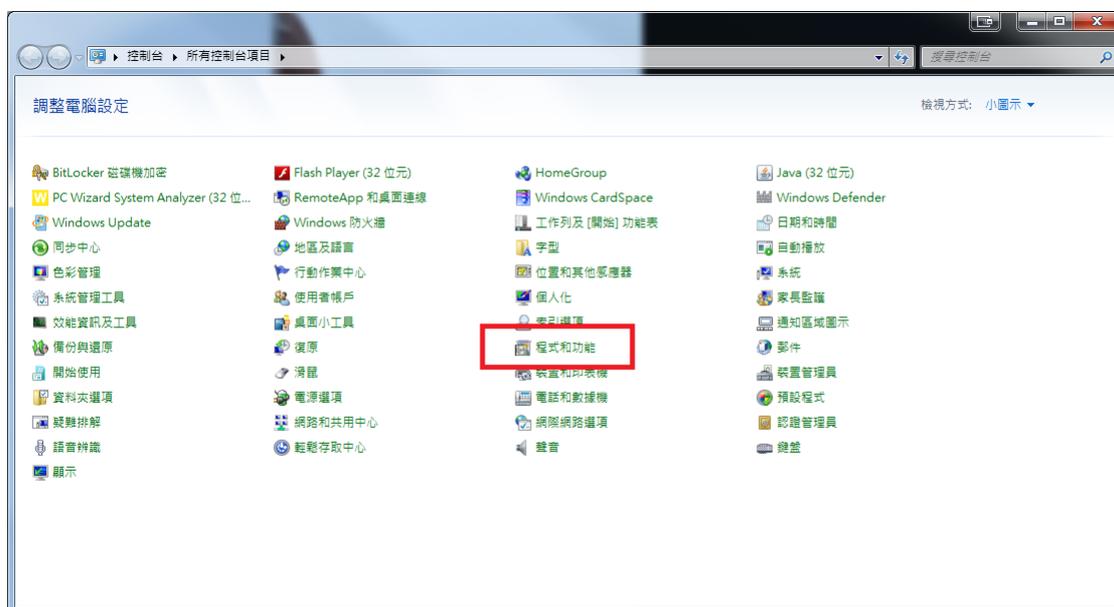
12. 安裝與教學

1. 安裝需求

- (1) Windows XP 以上的作業系統
- (2) IIS
- (3) Microsoft Visual Studio 2010
- (4) SQL Server Management Studio 2008 R2

2. 安裝教學(安裝示範之作業系統為 Windows 7)

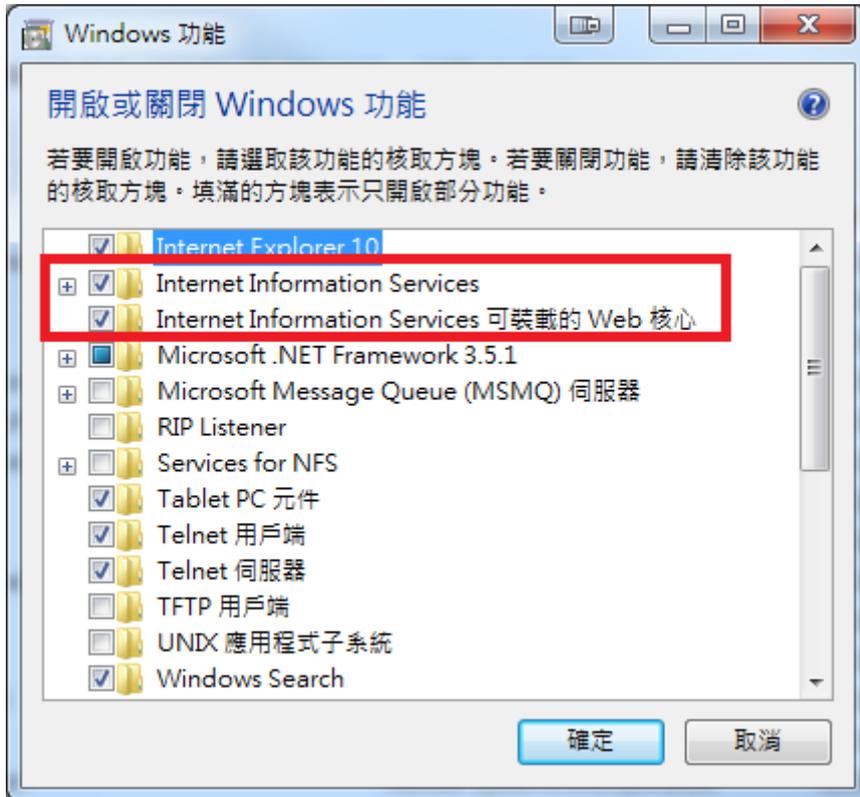
- (1) 首先，先安裝 IIS，打開控制台(右上角檢視方式切換為:小圖示)，
點擊「程式與功能」



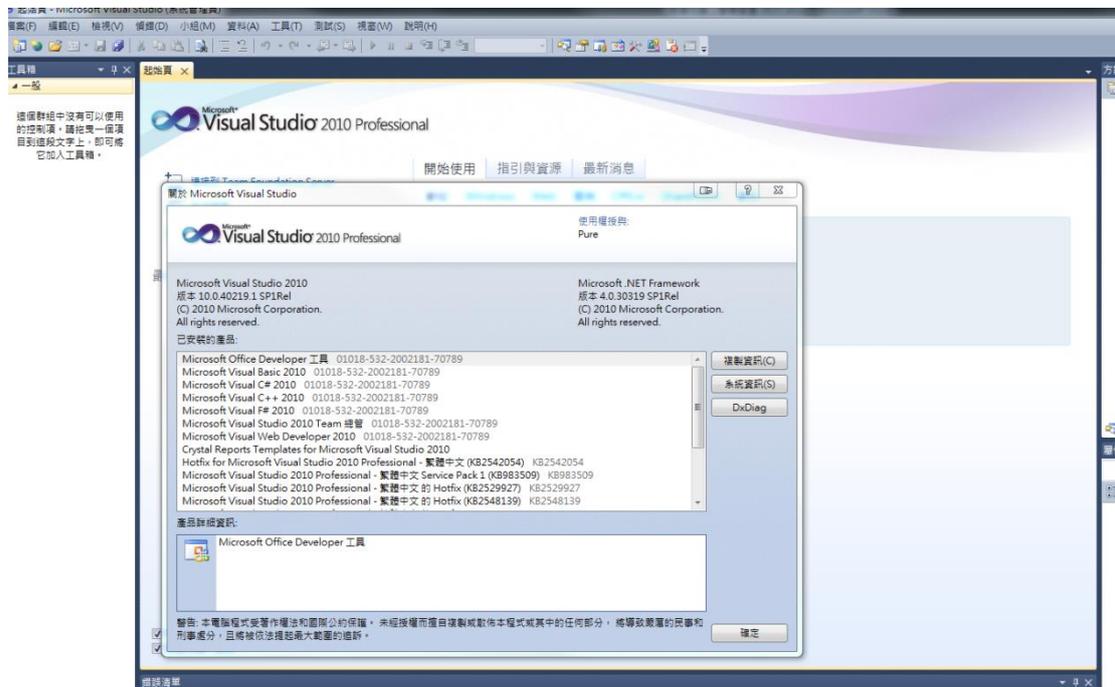
- (2) 進入後點擊左上角「開啟或關閉 Windows 功能」



- (3) 將 Internet Information Services 以及 Internet Information Services 可裝載的 Web 核心 可裝載的 Web 核心勾選後按下確認鍵進行安裝即完成 IIS 的安裝



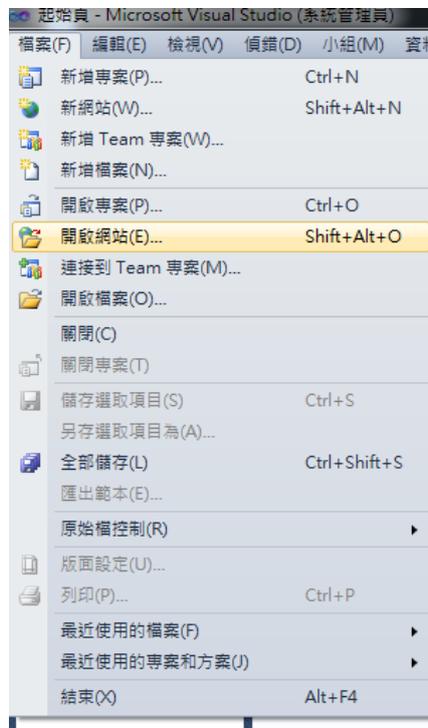
- (4) 安裝 Microsoft Visual Studio 2010



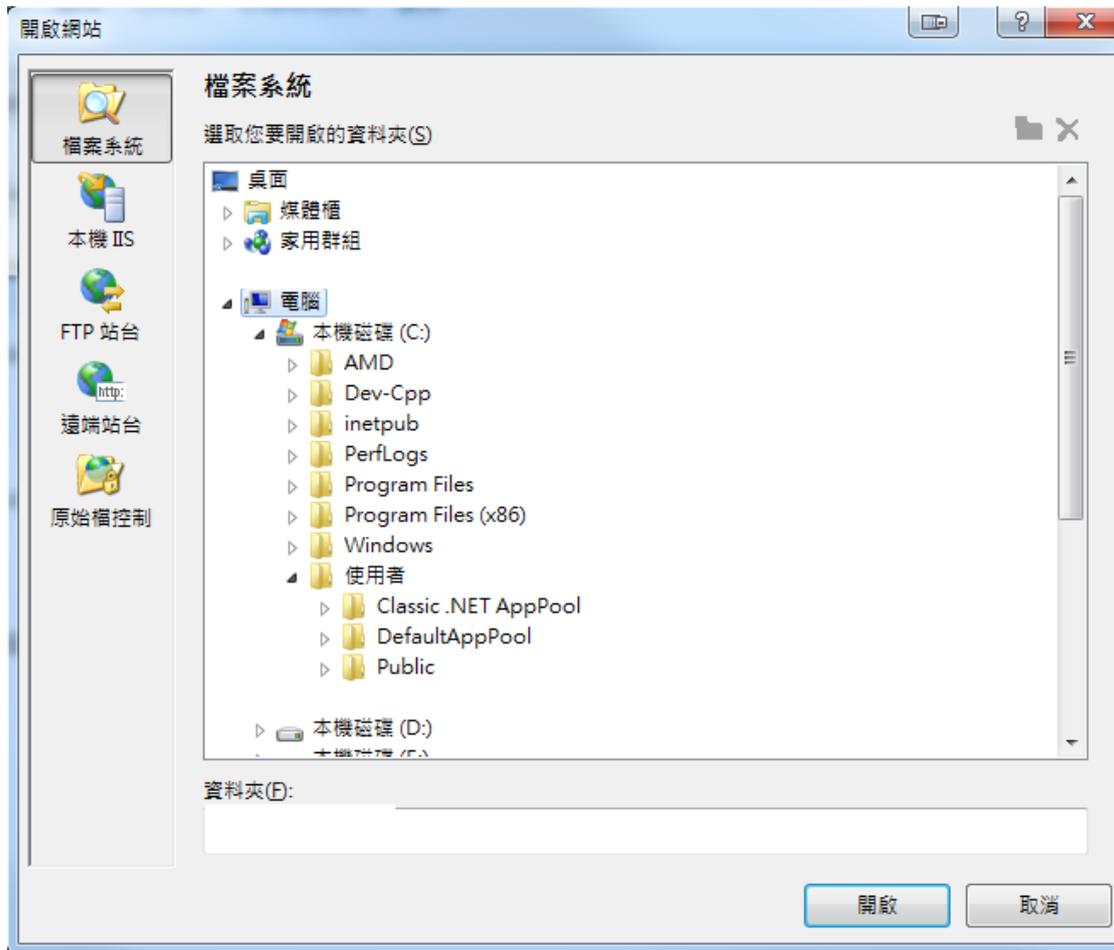
(5) 安裝 SQL Server Management Studio 2008 R2



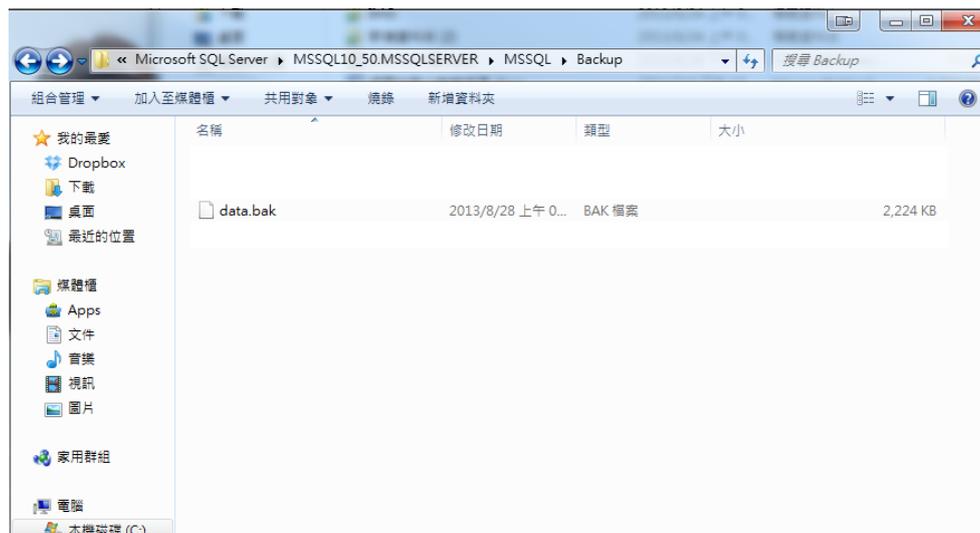
(5) 將光碟中附檔 01 中的壓縮檔案，解壓縮後放置電腦中並利用 Microsoft Visual Studio 開啟



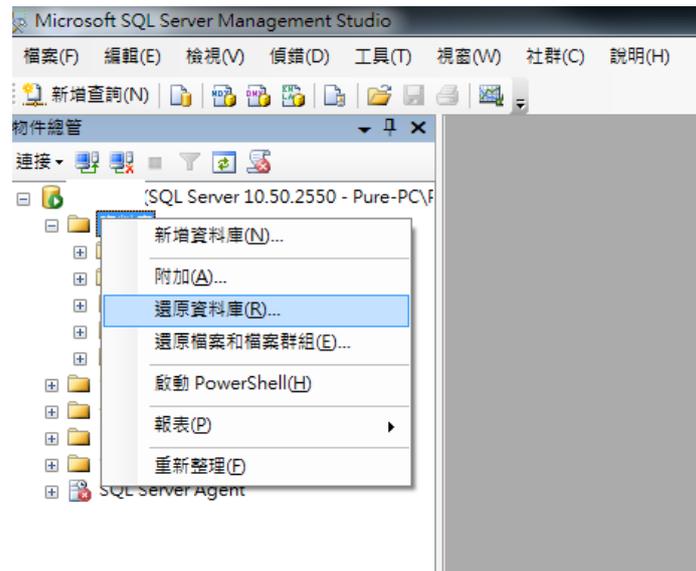
(6) 選擇資料夾位置後按開啟



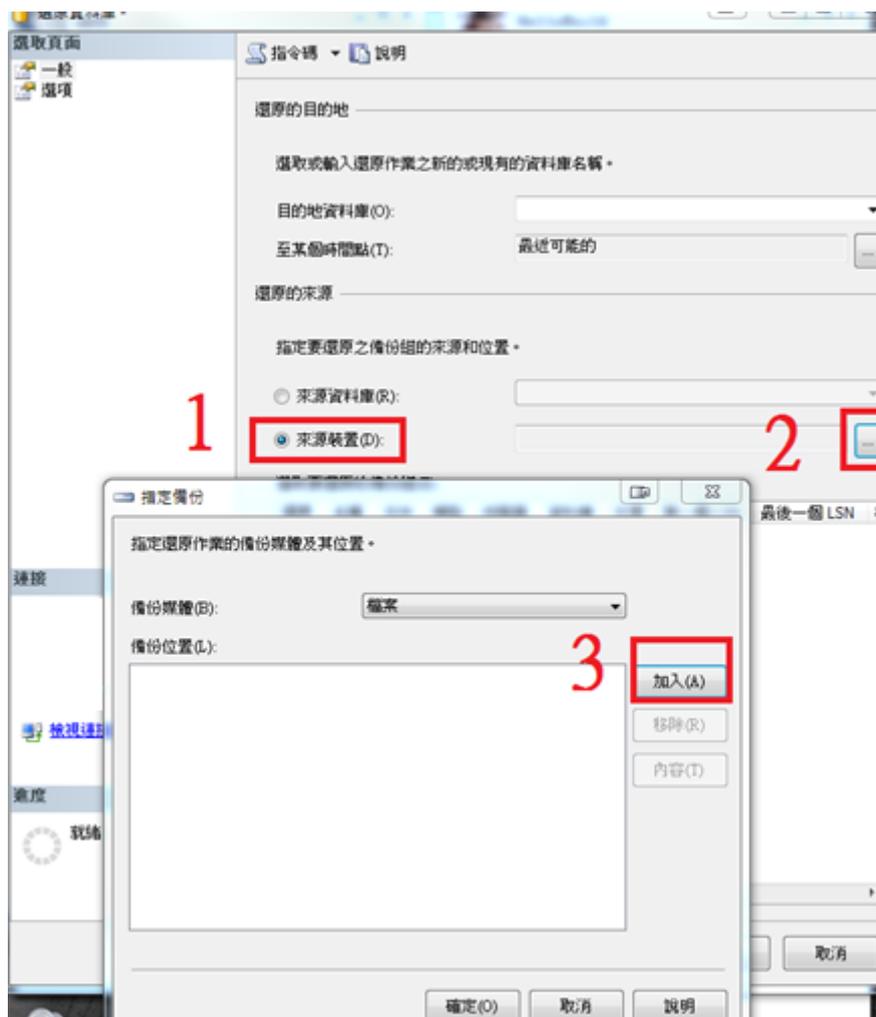
(7) 備份資料庫，將副檔 02 中的檔案放置下列路徑「C:\Program Files\Microsoft SQL Server\MSSQL10_50.MSSQLSERVER\MSSQL\Backup」



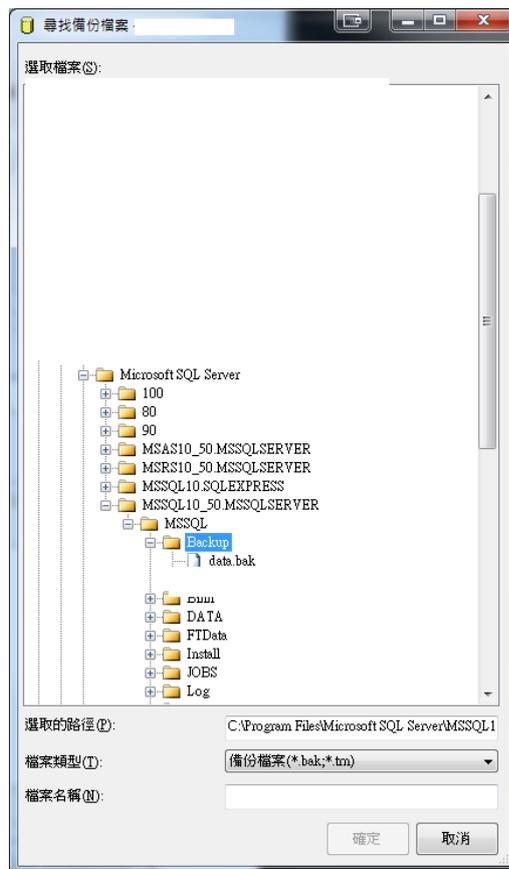
- (8) 接著打開安裝好的 Microsoft SQL Server Management Studio R2，對「資料庫」按下右鍵，並選取「還原資料庫」



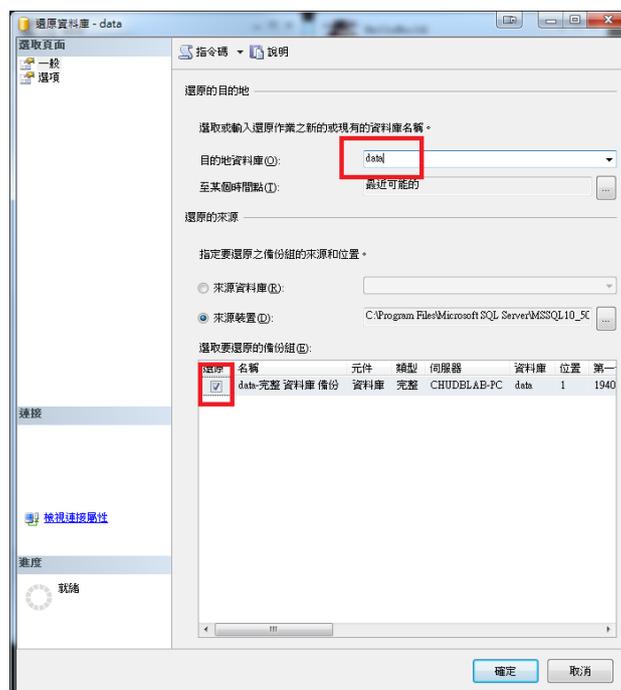
- (9) 跟著 3 步驟來選擇要還原的資料庫



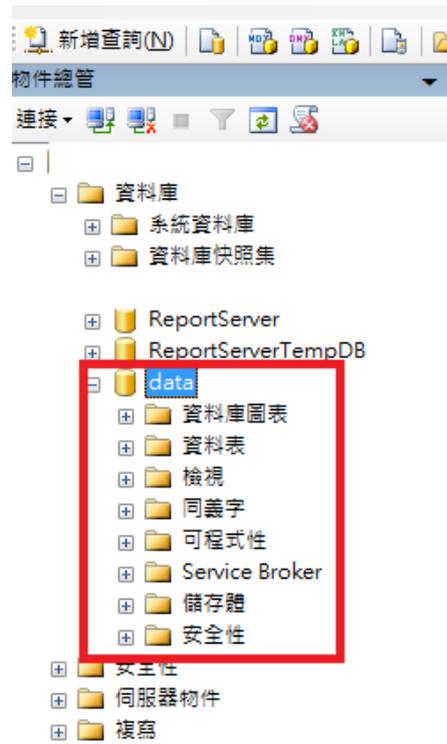
(10) 選擇先前步驟中所放入的資料庫備份檔，選擇完後按下確定鍵



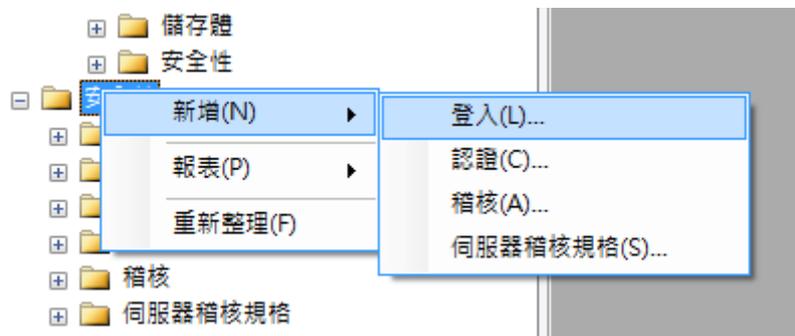
(11) 目的資料庫名稱輸入「data」以及將所選擇還原的資料庫勾選後按下確定鍵並等待資料庫還原即可



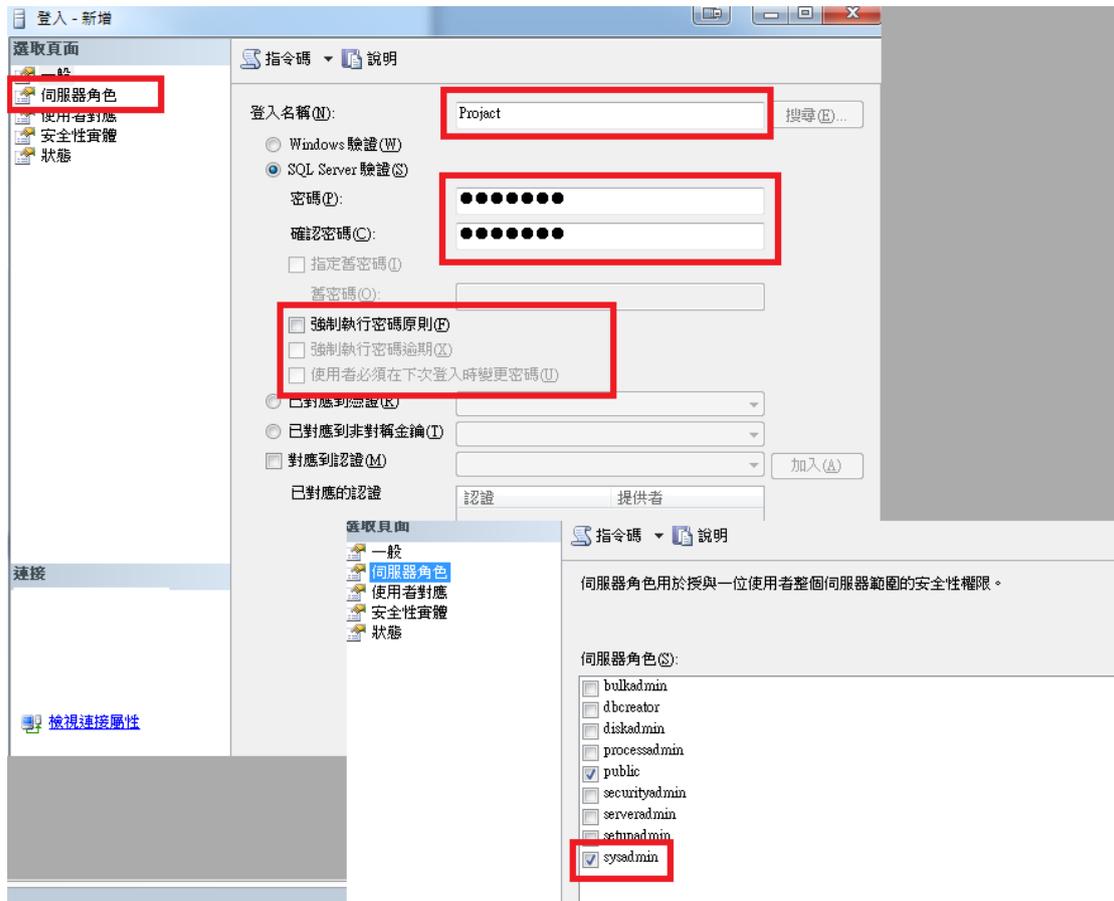
(12) 資料庫還原完成



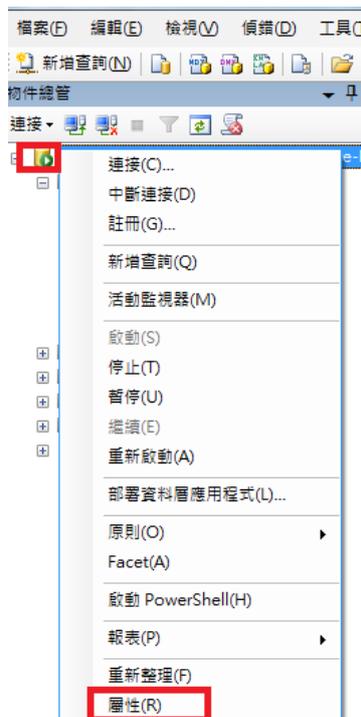
(13) 設定使用者，對「安全性」按下右鍵並選擇「新增」內的「登入」



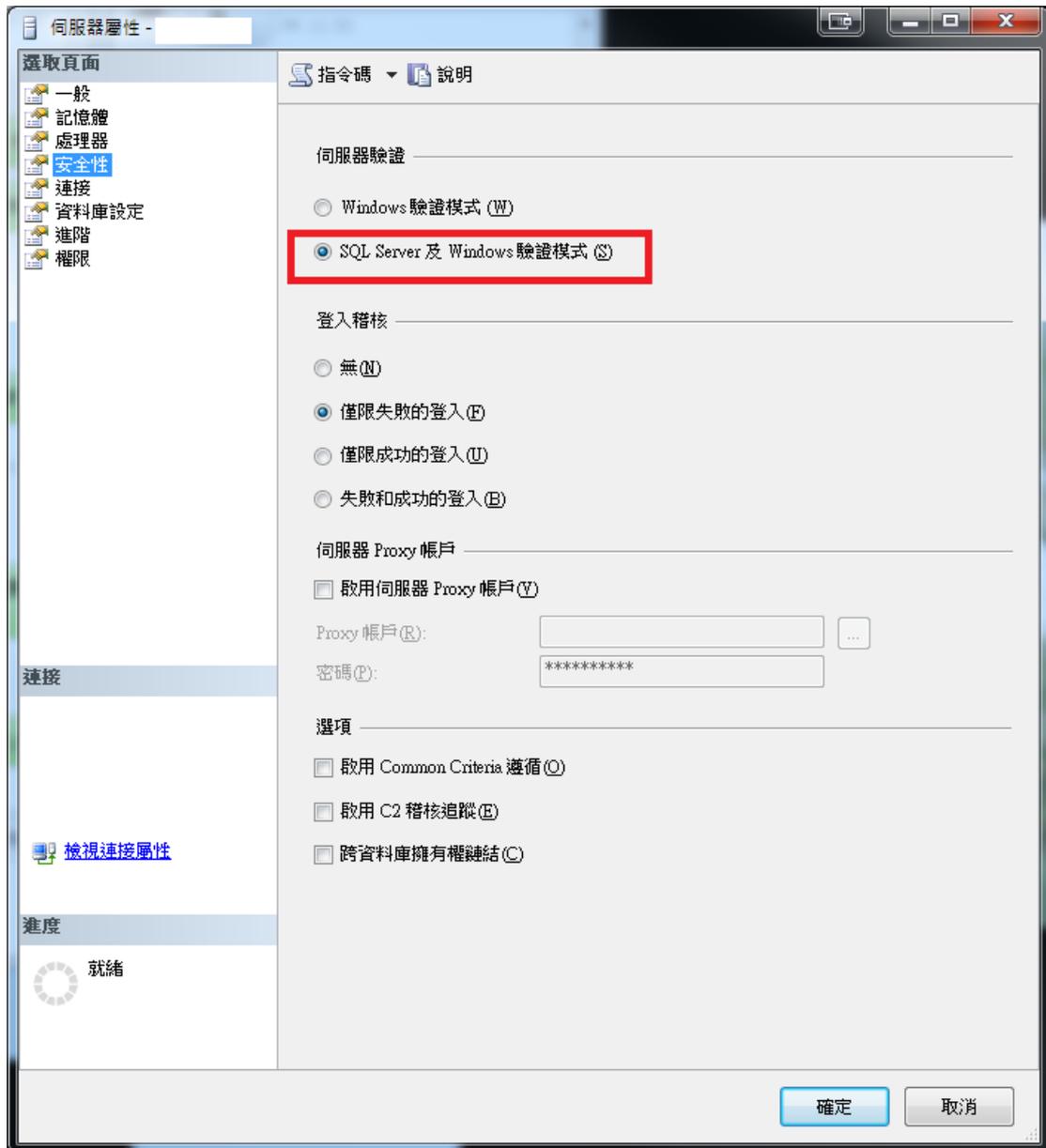
(14) 登入名稱和密碼皆為「Project」，請將「強制執行密碼原則」取消，再來點選左上角伺服器腳色候選下「sysadmin」，使用者即設定完成



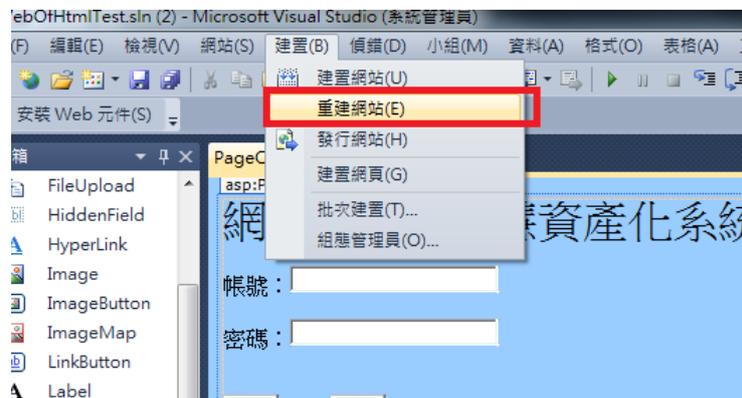
(15) 點選資料庫根目錄後選擇屬性



(16) 選擇安全性並將驗證選項選為 SQL Server 及 Windows 驗證模式後按下確定，資料庫設定即完成



(17) 回到 Microsoft Visual Studio 2010 中將網站重新建置



(18) 執行網站

