

XML 應用案例介紹—國家圖書館與入口 網站合作的期刊論文索引搜尋服務

陳 嵩 榮

華藝數位藝術公司副總經理

【摘要】

本文介紹國家圖書館與入口網站合作的期刊論文索引搜尋服務，包含合作模式與系統架構、資料庫欄位、資料交換格式 等技術規格，做為後續類似合作案評估與建置的參考。

關 鍵 詞：國家圖書館、入口網站、新浪網、期刊論文索引、可延伸標示語言、可延伸樣式表語言、可延伸樣式表語言轉換

一、前 言

根據蕃薯藤在民國 90 年 12 月進行的「2001 年臺灣網路使用調查」結果，使用者在網際網路上最常進行的活動是：使用搜尋引擎搜尋資料。^[1] 目前在網路上，主要的搜尋服務提供者是入口網站 (Portal Site)，在國內如雅虎奇摩^[2]、蕃薯藤^[3]、PChome Online^[4]、新浪網^[5] 等。入口網站提供的搜尋服務以網站搜尋

[1] 蕃薯藤，「2001 年臺灣網路使用調查結果 - 在網際網路上最常進行的活動 (單選)」，網址：
http://survey.yam.com/survey2001/chart/a_29.html。

[2] 雅虎奇摩首頁，網址：<http://www.yahoo.com>。

[3] 蕃薯藤首頁，網址：<http://www.yam.com>。

[4] PChome Online 首頁，網址：<http://www.pchome.com.tw>。

[5] 新浪網首頁，網址：<http://www.sina.com.tw>。

與網頁搜尋為主，這兩項搜尋服務的訴求是「數量」與「便利」，至於「品質」則較為忽略。在網路上任何人可自由出版資訊，只要是網路上的資訊，都可能出現在網站和網頁搜尋的結果之中，品質自然是參差不齊。對使用者而言，要判斷查詢結果主題是否與需求相關較容易，但查詢結果內容是否正確、權威，使用者不一定有能力判斷。值得注意的是，許多使用者以入口網站搜尋引擎作為蒐集資訊的主要來源，卻沒有意識到搜尋結果品質的問題。

很多資料庫目前在網路上都提供了查詢介面，如國家圖書館的遠距圖書服務系統^[6]，整合了各種資料庫檢索服務，這些資料庫都是優質內容的集合。分析入口網站搜尋引擎每天的查詢記錄(Log)，可發現一些專業的關鍵字，能滿足需求的應是專業的資料庫，而非網站或網頁搜尋。但有很多使用者從不上圖書館網站，也不知道如何善用各種資料庫，幾乎以入口網站搜尋服務作為唯一的資訊蒐集來源。入口網站的搜尋服務是便利的代表，查詢界面與動線設計最為簡單易用，但查詢結果品質則有待加強。圖書館的資料庫是典型的優質內容，如果能讓使用者在入口網站也能查詢各種資料庫，並能很方便地與網站、網頁搜尋的結果做切換，除了可達到推廣優質內容的目的，也能逐漸讓入口網站的使用者在切換網站、網頁與資料庫的搜尋結果時，比較內容品質，感受資料庫的好處，並學習善用資料庫。

筆者任職於新浪網期間，參與推動新浪網與國家圖書館的合作，民國9年3月雙方合作在新浪網推出期刊論文索引搜尋服務，90年雙方再合作推出博碩士論文搜尋服務，目前蕃薯藤也和國家圖書館採行同樣的合作模式。本文介紹國家圖書館與新浪網合作的期刊論文索引搜尋服務，包含合作模式與系統架構、資料庫欄位、資料交換格式等技術規格，做為後續類似合作案評估與建置的參考。

二、國家圖書館與新浪網合作的期刊論文索引搜尋服務

(一)合作模式

雙方以推廣優質網路資源為共同理念，先從遠距圖書服務系統中最多人查詢的期刊論文索引資料庫開始合作。國家圖書館透過新浪網所匯集的網站人潮能提高所屬期刊論文索引資料庫服務的知名度及接觸到更多的讀者；新浪網透過該優

[6] 國家圖書館遠距圖書服務系統，網址：<http://readopac.ncl.edu.tw>。

上連結中英文摘要或電子全文，必須點選篇名瀏覽詳細書目資料才能知道。新浪網期刊論文索引搜尋則將這項識別往外拉一層，凡可線上連結中英文摘要或電子全文的文章，在查詢結果頁面序號欄位之前會出現一個小註記，讓使用者做為是否點選的參考。做這樣小小的改良，主要是遵循入口網站便利的精神，尤其對入口網站的使用者而言，搜尋後能直接在網路上取得摘要或全文資料是最佳的選擇。



圖二：國家圖書館期刊論文索引搜尋結果詳細書目資料

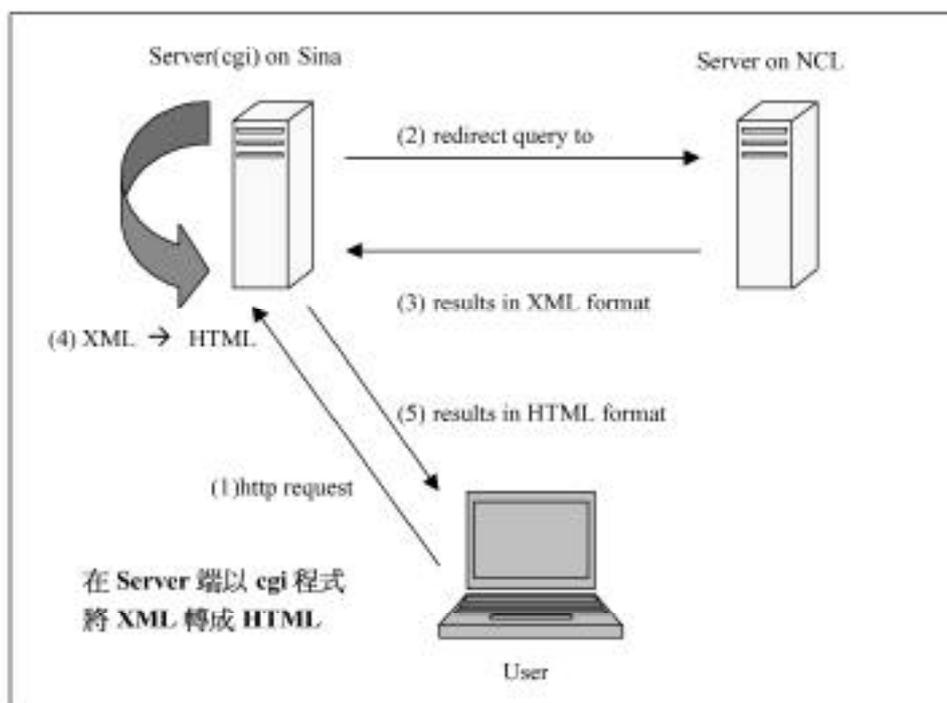


圖三：新浪網期刊論文索引搜尋結果（一）

(二)技術規格

圖四是新浪網期刊論文索引搜尋服務背後的系統架構，從使用者輸入關鍵字到顯示結果頁面經過以下幾道步驟：

1. 使用者輸入關鍵字，開始查詢。
2. 新浪網伺服器收到使用者的查詢需求，將查詢需求導向 (redirect query) 國家圖書館伺服器針對此合作案所設立的查詢介面。
3. 國家圖書館伺服器將查詢結果以XML 格式回傳。
4. 新浪網伺服器將查詢結果從XML 轉成 HTML。
5. 將轉換後的HTML 頁面回應使用者端，即使用者所看到的頁面。



圖四：新浪網期刊論文索引搜尋系統架構（一）

這樣的合作模式對雙方而言建置成本都很低，由於資料交換採用 XML 格式，因此不用管對方採用哪種平臺。國家圖書館只要提供一個專屬的查詢介面，接收從新浪網伺服器導向而來的查詢需求，並將查詢結果以XML 格式回傳；新浪網則準備一支 CGI 程式接受使用者端的查詢需求，並導向國家圖書館伺服器，並將國家圖書館伺服器回傳的XML 檔轉成新浪網風格的HTML 頁面，再回

應給使用者端。

根據以上的系統架構，建置時需先定義雙方資料交換的欄位與標籤。雖然雙方合作時，國家圖書館可能只釋出部分的查詢功能，或者入口網站只採納部分的查詢功能，例如只提供簡易查詢，不提供詳細查詢，但定義規格時儘可能涵蓋完整的服務範圍，這樣可保持規格的相對穩定性。以下是本合作案資料交換欄位與標籤的定義：

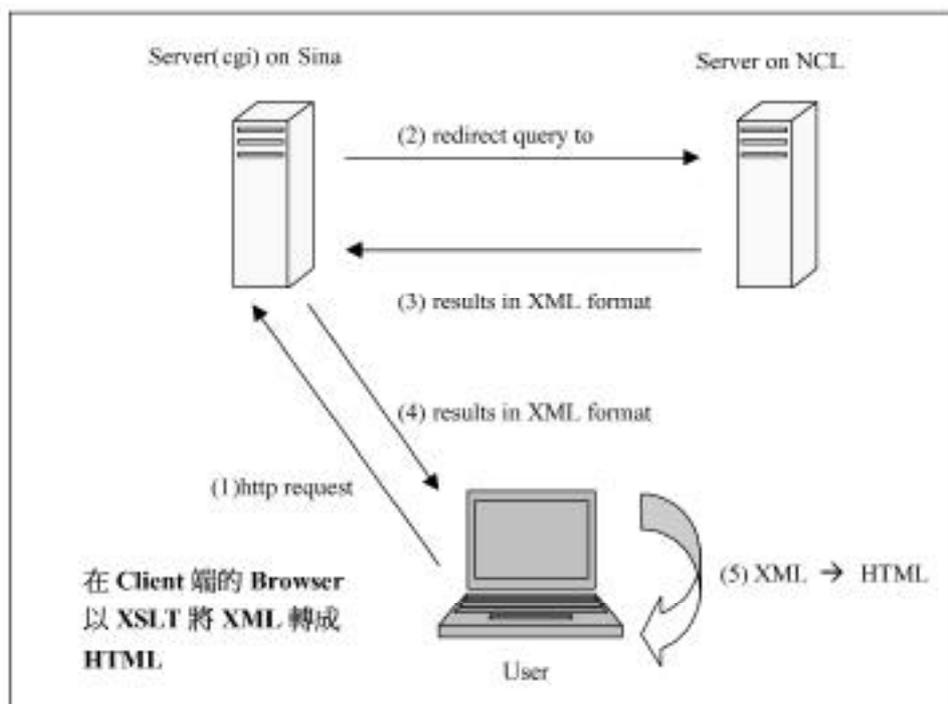
項 目	說 明
資料庫名稱	國家圖書館期刊論文索引資料庫 (http://readopac.ncl.edu.tw/)
資料庫欄位	<ul style="list-style-type: none"> * 系統識別號 (例：A9924261) * 篇名 (例：SGML、HTML 與 XML 之比較) * 並列篇名 (例：Comparison of SGML) * 作者中文名 (例：陳嵩榮) * 作者英文名 (例：Chen, Sung-jung) * 刊名 (例：大學圖書館) * 卷期 / 年月 (例：3:1 民 88.01 頁 89-103) * 資料語文 (例：中文) * 專輯 (例：圖書資訊組織) * 關鍵詞 (例：可擴展標示語言 XML) * 類號 (例：001.64) * 網路資源連結 (例：http://www.lib.ntu)
國家圖書館開放給新浪網伺服器 的查詢界面 (SINA--> NCL)	<p>CGI 格式：</p> <p>http://host_name/path_name/cgi_name?parameter1=value1&parameter2=value2&</p> <p>CGI 參數定義：</p> <ul style="list-style-type: none"> * input：指定關鍵字，例：input=新浪網 * mode：指定查詢模式，[精確 (1) 同音 (2) 模糊 (3) 羅馬拼音 (4) 漢語拼音 (5)]。例：mode=1 (預設值) * item_no：指定每頁顯示筆數，例：item_no=20 (預設值) * fields：指定檢索欄位，[全部欄位 (1) 篇名 (2) 關鍵詞 (3)

	<p>摘要 (4) 作者 (5)]。例 : fields=1 (預設值)</p> <p>* la : 指定資料語文 , [(不限 (nop) 中文 (chi) 英文 (eng) 法文 (fre) 德文 (ger) 日文 (jpn) 韓文 (kor) 葡萄牙文 (por) 俄文 (rus) 西班牙文 (spa)] , 例 : la=nop (預設值)</p> <p>* page : 針對查詢結果指定顯示的頁數號碼 , 例 : page=2</p> <p>* db : 指定資料庫 (預留未來擴充性) , 例 : db=1</p> <p>* from : 合作廠商 (預留未來擴充性及方便統計) , 例 : from=sina</p>
查詢結果格式、欄位與標籤定義	<p>回傳格式 : XML</p> <p>回傳欄位與標籤定義 :</p> <p>* 查詢關鍵字 : <keyword></p> <p>* 結果筆數 : <num></p> <p>(每筆記錄包含以下欄位)</p> <p>* 序號 : <seq></p> <p>* 篇名 : <title></p> <p>* 作者 : <creator></p> <p>* 刊名 : <journal></p> <p>* 卷期 : <vol></p> <p>* 出版年月 : <date></p> <p>* 全文影像掃描註記 : <bf></p> <p>* 授權標誌註記 : <lf></p> <p>* 書目資料頁面URL : <url></p> <p>* 是否有電子全文 : <fulltext></p> <p>* 是否有中文摘要 : <cabstract></p> <p>* 是否有英文摘要 : <eabstract></p> <p>如下例 : (NCL--> SINA)</p> <pre><?xml version="1.0" encoding="big5"?> <record-list> <keyword>xml</keyword></pre>

	<pre> <num>3</num> <record> <seq>1</seq> <title>XML及RDF技術介紹</title> <creator>梁高榮</creator> <journal>機械工業</journal> <vol>220</vol> <date>90.07</date> <bf>N</bf> <lf>N</lf> <url>http://www2.read.com.tw/cgi/ncl3/ncl3detail?732c632c51756572 793d41303131393536302c776f464b666b73464c74585656727977414a 2c32302c58534c010158534c0158534c010158534c01</url> <fulltext>N</fulltext> <cabstract>N</cabstract> <eabstract>N</eabstract> </record> <record> </record> </record-list> </pre>
查不到資料的回傳結果 (NCL--> SINA)	<pre> <?xml version="1.0" encoding="big5"> <record-list> <num>0</num> </record-list> </pre>
伺服器發生錯誤的回傳結果 (NCL--> SINA)	<pre> <?xml version="1.0" encoding="big5"> <record-list> <num>0</num> <error>Y</error> </record-list> </pre>

以上的後端規格是在新浪網伺服器把XML 轉成 HTML，另一種做法是運用 XSLT (XSL Transformation) 在使用者端將 XML 轉成 HTML，如圖五。從使用者輸入關鍵字到看到結果頁面經過以下幾道步驟：

1. 使用者輸入關鍵字，開始查詢。
2. 新浪網伺服器收到使用者的查詢需求，將查詢需求導往 (redirect query) 國家圖書館伺服器針對此合作案所設立的查詢介面。
3. 國家圖書館伺服器將查詢結果以XML 格式回傳。
4. 新浪網伺服器將XML 檔案附加上樣式表路徑回應給用戶端。
5. 在用戶端依指定的樣式表，將XML 轉成 HTML 顯示。



圖五：新浪網期刊論文索引搜尋系統架構（二）

目前大部分的瀏覽器都已支援 XSLT，在瀏覽器端將 XML 轉成 HTML，這樣在伺服器端就可以省略轉換的動作。入口網站準備的 CGI 程式更簡單，接受使用者端的查詢需求，導向國家圖書館伺服器，並將國家圖書館伺服器回傳的 XML 附加上樣式表路徑，再回應給使用者端。以下是使用者端收到的 XML 檔案範例，第二行是樣式表宣告，指定了樣式表路徑。

```
<?xml version="1.0" encoding="big5"?>
<?xml-stylesheet href="http://hostname/pathname/filename.xml" type="text/xsl" ?>
<record-list>
<keyword>xml</keyword>
<num>3</num>
<record>
<seq>1</seq>
<title>XML及RDF 技術介紹 </title>
<creator> 梁高榮 </creator>
<journal> 機械工業 </journal>
<vol>220</vol>
<date>90.07</date>
<bf>N</bf>
<lf>N</lf>
<url>http://www2.read.com.tw/cgi/ncl3/ncl3detail?732c632c51756572793d4130313
1393536302c776f464b666b73464c74585656727977414a2c32302c58534c01015853
4c0158534c010158534c01</url>
<fulltext>N</fulltext>
<cabstract>N</cabstract>
<eabstract>N</eabstract>
</record>
<record>
.....
</record>
.....
</record-list>
```

樣式表是一個 .xsl 檔，目前坊間已有許多介紹 XSL 語法的書籍。以下是筆者摸索出來的流程，用於設計將特定XML 格式轉成 HTML 網頁的 .xsl 檔：

1. 先準備好原始的 XML 檔（如：a.xml）與目的 HTML 檔（如：b.html），以

本案為例，查詢結果頁面為動態產生的網頁，因此先以任一個關鍵字（如：xsl）的查詢結果做為目的HTML檔。

2. 將 XML 檔附加上樣式表路徑，如 href="b.xsl"，附加後的 XML 檔命名為 b.xml。
3. 將 b.html 最外層包上 XSL 根元素 <xsl:template>，命名為 b.xsl，以瀏覽器開啟 b.xml，看到的結果應與開啟 b.html 相同，但此時 b.xsl 仍不具有動態轉換的功能。
4. 針對 b.xsl 中 HTML 動態變化的部分以 XSL 語法取代，邊寫可邊以瀏覽器開啟 b.xml，查看顯示是否正常。
5. 所有動態變化的部分都以 XSL 語法取代後，b.xsl 就是所要的 .xsl 檔。

以下是以此方法寫作的樣式表範例：

```
<?xml version="1.0" encoding="big5"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl">
<xsl:template match="/">
<html>
<head>
<title> 新浪搜尋 - 期刊論文索引搜尋</title>
</head>
<body>

<xsl:for-each select="record-list/record">
<tr>
<xsl:if test=".[fulltext='Y' or cabstract='Y' or eabstract='Y']">
<td align="center" bgcolor="#FFFFFF">
</td>
</xsl:if>
<xsl:if test=".[fulltext='N' and cabstract='N' and eabstract='N']">
<td align="center" bgcolor="#FFFFFF"></td>
</xsl:if>
```

```
<xsl:apply-templates />
</tr>
</xsl:for-each>
.....
</body>

</html>
</xsl:template>

<xsl:template match="*/seq">
<td align="center" bgcolor="#FFFFFF"><font face="Verdana" size="2">
<xsl:value-of />
</font></td>
</xsl:template>

<xsl:template match="*/title">
<td bgcolor="#FFFFFF"><font size="2" color="#02621A">
<a><xsl:attribute name="href">
<xsl:value-of select="/record-list/*/url"/>
</xsl:attribute>
<xsl:value-of />
</a></font></td>
</xsl:template>

<xsl:template match="*/creator">
<td bgcolor="#FFFFFF"><font size="2">
<xsl:value-of />
</font></td>
</xsl:template>
```

```

<xsl:template match="*/journal">
<td bgcolor="#FFFFFF"><font size="2">
<xsl:value-of />
</font></td>
</xsl:template>

<xsl:template match="*/date">
<td align="center" bgcolor="#FFFFFF">
<font size="2" face="Verdana" color="#02621A">
<xsl:value-of />
</font></td>
</xsl:template>
</xsl:stylesheet>

```

圖六的新浪網期刊論文索引搜尋結果頁面，是透過 XSLT 將 XML 轉成 HTML，與圖三比較，只少了上方的橫幅廣告，其餘效果都相同。在圖三的 HTML 原始檔中，橫幅廣告是以 Javascript 來控制，夾在 HTML 註釋標籤 `<!-- -->` 之間，在圖六 XSLT 進行轉換時，會略過 .xsl 檔裡頭的註釋資訊，因此橫幅廣告就不見了。



圖六：新浪網期刊論文索引搜尋結果（二）

三、結 語

國家圖書館與新浪網合作的期刊論文索引搜尋服務，建置工程不大，為資料庫服務提供者與入口網站合作的可行模式。對國家圖書館而言，藉由與入口網站合作，可有效推廣所屬資料庫服務，尤其對那些以入口網站為主要資訊蒐集來源的使用者，讓他們除了熟悉網站、網頁搜尋服務外，在很自然的情況下，體驗到資料庫搜尋服務的好處，認知不同的資訊需求應該採用不同的資訊服務。對入口網站而言，既有的搜尋服務在品質方面，進步的空間還很大，若能與各類專門資料庫檢索服務整合可提升服務的深度，也可吸引消費力較高的專業族群到訪，利於分眾行銷。期待未來看到更多類似的合作案，讓使用者更便利地接觸各類優質網路資源。

A Case Study of XML Applications: Searching Service of Periodical Index Cooperated by the National Central Library and Portal Site

Sung-Jung Chen

Director, Airiti Coporation

【Abstract】

This article is to introduce a case study on the searching service of periodical index cooperated by the National Central Library and portal sites, including cooperative models and technical specifications such as system architecture, database schema and data exchange format. This can be a reference for evaluating and implementing the related projects in the future.

Keywords: NCL (National Central Library), Portal Site, SINA, Periodical Index, XML, XSL, XSLT