

哈伯的天空

URI, etc.

----- 艾文 -----

自從有了網際網和有了Google這類超級搜索引擎以後，圖書資訊專業館員就一直耽心著讀者族群的流失和館藏的被冷落。其實，讀者族群的逐漸遠離圖書館和忽視圖書館館藏的真正原因，倒不是Google等搜索引擎，而是那幫它們作惡的「通用資源網址」URL (Uniform Resource Locators)¹。自從資訊網版化和有了URL以後，網版資訊資源(Internet-based information resources)才有了讓搜索引擎在網際網上搜尋的地址。否則，不要說Google，縱使如來菩薩的慧眼也不用想會找到它。不過，我們談URL，卻不能不知道「通用資源規格鑑定符號」URI (Uniform Resource Identifier)和「通用資源名稱」URN (Uniform Resource Name)這三者之間錯綜複雜的關係。現在不妨讓我們簡單地介紹一下。

什麼是「通用資源規格鑑定符號」URI?

在網際網上傳輸的任何資訊資源，都有它結構和內容上的特性。為了便於網訊資源的通訊、連線互動和傳輸，網際網設計專家早在 1990 年左右，便開始利用一組具有特殊意義的字母或文字來界定它們是那一類的資訊。這一組具有特殊意義的字母或文字的族群就統稱為 URI。根據報導，目前已經正式登記成為 URI 族群已達 65 個²，而尚未登記和各機構自己設計的 URI，還很多很多。譬如說，去年英國政府欲為各級政府發行的資料(data)，設計一套政府機構專用的「通用資源規格鑑定符號」，希望能為這些資料建立一個資料網，使網絡(Intranet)中的資料能夠連接、查詢和貯集。³ 假如設計成功，這些英國政府編製的 URI，就成了另一批區域性的「資源規格鑑定符號」。在已登記的「通用資源規格鑑定符號」族群中，比較常見的有：

<u>URI 類別</u>	<u>內容說明</u>
dict	dictionary service protocol
fax	fax
file	Host-specific file names
ftp	File Transfer Protocol
gopher	The Gopher Protocol
http	Hypertext Transfer Protocol
https	Hypertext Transfer Protocol Secure
imap	internet message access protocol
info	Public Namespaces

iris	Internet Registry Information Service
mailto	Electronic mail address
news	USENET news
nfs	network file system protocol
nntp	USENET news using NNTP access
rtsp	real time streaming protocol
shttp	Secure Hypertext Transfer Protocol
tel	telephone
telnet	Reference to interactive sessions
tftp	Trivial File Transfer Protocol
tv	TV Broadcasts
urn	Uniform Resource Names

其實，在這麼龐大的 URI 族群當中，真正隸屬於 URI 的只有 URL 和 URN 二類，其他都應納入適用於全球資訊網的「通用資源引用文獻」URC(Uniform Resource Citation)。這一類 URC，包括「詮釋資料」(Metadata) tags 和一些非“通用”的區域性的「資源規格鑑定符號」。在組織結構上，URL 和 URN 只有二種通用模式：“URI://”和“URI:”。不過，我們應該注意的是不同的 URI 代表不同的規格(Protocol)。

URL : URI://domain name/path/files

例如：<http://www.ietf.org/rfc/rfc2396.txt>
<ftp://ftp.is.co.za/rfc/rfc1808.txt>
<telnet://192.0.2.16:80/>

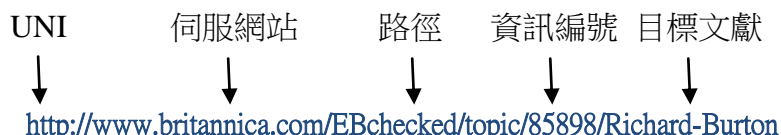
URN : URI:

例如：urn:isbn:0-395-36341-1

綜結起來說，假如我們將「統一資源規格鑑定符號」URI 視為一個集(set)，那麼它包括的子集(subsets)，並不只 URI、URN 二個，而是 URI、URN 和 URC 三個。若我們設定 URL={1, 2, 3}，URN={4, 5, 6}，URC={7,8,9}，那麼 URI=URL ∪ URN ∪ URC={1, 2, 3,4,5,6,7,8,9}，或 URL ∩ URN ∩ URC = 0。這也就是說 URL、URN 和 URC 之間，沒有共同的元素(Element)，為不交集(Disjoint sets)。對電腦科技專家來說，他(她)們都清楚三者間的區別，但對一般人來說，就可能將 URI 當作 URL(網址)，或者當作 URN(名稱)，或者看成 URC(資源索引)。現在就讓我們再來分別地概談一下這三個不同類別，不同名稱，不同性質的子集詞彙。

URL :

簡單地說，URL 是網版資訊資源在網路上的地址，也是提供一條找到這個網版資訊資源的路線和方法。它的重要實不可言喻。譬如說，Wikipedia 網站的網址為 <http://www.wikipedia.org/>。"http"說明這一個網頁是根據全球資訊網 http 規格編寫而成。一般典型的 HTTP 的結構，可以細分為 UNI，伺服網站，路徑，資訊編號，資訊文獻五個部份。前二部份是必需有的主幹，後面三部份是支幹，它們分別用來標明資料庫中目標文獻(Targeted documents)貯存的檔案路線。



搜尋網版資訊，URL 的功用非凡，它可說是不可或缺的 GPS(Global Positioning System)。可是，我們發現縱然有了 URL，在網上搜尋資訊的時候，仍舊少不了麻煩和困難，譬如說，高昂的在線查閱(Access)費用便是其中之一，然而，比金錢更重要的，則是網址的時時變更、消失和“深網”(Deep web)資訊資源的難以發掘等。為了解決這些錢財以外的搜索問題，軟體工程和網頁設計專家群中便有了所謂「靜態網址」(Static URL)和「動態網址」(Dynamic URL)的取捨考慮。⁴

1. 什麼是「靜態網址」?

簡單地說，「靜態網址」是指網址結構永遠不變。而且在結構中通常都沒有任何參數(Parameters)。以全球資訊網的 HTTP 為例，它的結構可能如下：

<http://www.domain.com/journal/january.htm>

問題是，像這一類的網址更新起來非常費人費時。假如是期刊，每出一期都必須跟著添改。以上例為例，二月的出版，就必須添加一條指明二月份的新網址，尤其是如今網訊數量和種類，每分每秒都不斷在增加，真正是加不勝加，改不勝改。因此很多 e-Commerce，包括線上商店(Online Stores)，甚至社交和部落格族群的網頁製作者，為了應付多元詮釋資料的原因，便想採用稍具彈性的「動態網址」。

2. 什麼是「動態網址」?。

「動態網址」是根據網版資訊供應者提供的參數，或資訊應用者如圖書館自我附加的參數所組織而成的網址。在這類網址的結構中，通常都可以發現一些特殊的符號，例如：&, \$, +, =, ?, % 等等。很多“深網”網訊大都採用這

樣的結構。下面的例式中就有一二個這樣的符號：

<http://www.example.com/article/bin/answer>
foo?language=en&answer=3&sid=8971&query=URL

說明：

language=en	文獻語文為英文
answer=3	3 號文獻
sid=8971	文獻 ID 為 8971
query=URL	文獻網址模式為 URL

上例 URL 可以簡化成：

www.example.com/article/bin/answer.foo?language=en&answer=3

2001 年以前，超級搜索引擎如 Google 的網絡搜索器(web crawler)都無法滲透多層次多參數的「動態網址」的文獻。直到現在，由於「動態網址」結構的逐漸規律化⁵，同時，許多文獻主頁的網址採用「靜態網址」結構的結果，Google 等搜索引擎和 Link Resolver 一類的服務都開始能夠滲透“深網”網訊和各重要大專院校圖書館的線上(Online)館藏。

3. 「靜態網址」vs. 「動態網址」：

老實地說，網版資訊搜索者並不在乎它是「靜態網址」或是「動態網址」，他(她)們重視的只是檢索結果。近年來，網絡搜索器的功能越來越進步，它們已能滲透並搜索結構非常複雜的網頁。而網版資訊搜索者透過搜索引擎搜得的網訊的相關度也逐漸增高，品質也相對地較為優良，同時，主題含蓋面也日漸週全。至於網址究竟應該如何結構，網版資訊搜索者不必去費心，但是卻不能不關心。

URN：

URN 和 URL 之間的唯一且最重要的區別是 URN 只說明資訊資源是什麼，而沒有說明它來自何處和如何在網際網上去找到它。譬如海明威(Ernest Hemingway, 1899-1961)於 1951 年撰成，1952 年出版的《老人與海》(The Old Man And The Sea)。這本書，到今天為止，複印無數，各種版本都有，且出版商也不相同。但是無論是那一家或那一種版本，它們都有一個專門的「國際標準書號」ISBN(International Standard Book Number)。我們且以 1952 年由美國紐約市 Scribner Book Company 出版的精裝本為例，它的專用的標準書號為 0-684-15363-7。當我們在圖書館中採

用 ISBN 檢索方法，鍵入該號碼，只要館藏中有 1952 年版本的《老人與海》，我們便可以找到它。因此，這個書號就像一個人的姓名，不同版本的書號，代表不同的版本，就如同一個人有別名，字，號，筆名等等。而 URN 的結構模式，也顯現出這樣的不同，例如：

Urn:isbn:0-684-15363-7 (精裝版)

Urn:isbn:0-684-10215-5 (平裝版)

Urn:isbn:0-743-56436-7 (音響版)

根據上面這些例子，我們只知道《老人與海》這本書有不同的版本，但是它們並沒有告訴我們從何處去獲得這些書。⁶ 對圖書館和它的讀者族群來說，URN 的功用比起 URL 來，要差很多。

URC：

「通用資源引用文獻」URC (Uniform Resource Citation) 是一組尚未正式登記或認可的句法結構。它們多根據全球資訊網規格，利用文字和符號來描述一個資訊目標。最好的例子，就是以各種詮釋資料(Metadata) 標準規格編寫的句子。這些句子組織的特點是每一組成元素都有名稱和有特定的意義。而每種規格多少都概括一特定文獻的內容、出處、型格，如文字或影像，地圖或模型等等；版本格式，如紙本位資訊或電子本位資訊資源。同時，為了便利線上檢索，詮釋資料元素還包括著作者、出版公司、題名和主題。再如 Dublin Core 等，還允許加上 .asp .html .js .php 等等用來指示文件類型的「擴展檔名」(extension)。總之，URC 並不符合一般正規的句法結構，它們只是根據規格自由編製的欄位(fields)和參數值(value)。

綜上所述，我們知道 URI 並不等於 URL，它代表的是一個數目相當大，功能特殊的族群。其中包括 URL，URN 和其他 63 種不同類別的鑑定符號。當然，到目前為止，對圖書資訊服務的影響最大的只有配合全球資訊網網訊的 URL 和被編入 http 格式中的 URN。假如沒有它們的存在，檢索網版資訊資源，可能還不會成為事實。或許傳統式的圖書館服務還可以高枕無憂地再過上三十年！

spacehubble@msn.com

02/20/2010

¹ 此為筆者所釋，由於中釋名稱各異，建議讀者應盡量習用原文。

² 參考 <http://www.iana.org/assignments/uri-schemes.html>

³ “Designing URI Sets for the UK Public Sector”，

<http://writetoreply.org/ukgovuriset/introduction/#ixzz0UIvHjdSq>

⁴ 有關「靜態網址」vs.「動態網址」的討論，請參考“Dynamic URLs vs. static URLs”，Google Webmaster Central Blog，November 11，2009

<http://googlewebmastercentral.blogspot.com/2008/09/dynamic-urls-vs-static-urls.html> 和相關網訊。

⁵ “URIs, URLs, and URNs: Clarifications and Recommendations 1.0”，

<http://www.w3.org/TR/uri-clarification>

⁶ 查詢書目紀錄，可以 <http://www.worldcat.org/isbn/0684153637> 和 WorldCat 連接。