

運用智慧型多重代理人技術於數位典藏的資源搜尋與管理

林光龍，許永真，林士睿

國立台灣大學 資訊工程研究所

lkl@openfind.com.tw yjhsu@csie.ntu.edu.tw sjlin@csie.ntu.edu.tw

摘要

綜觀目前世上所進行的許多數位圖書館研究計畫，每個計畫的主題與方向，可說是上至天文，下至地理，無所不包。而典藏內容更是包羅萬象，多彩多姿，這些分散於不同平台與不同系統的典藏資料，如果能夠依據內容的主題與方向來建立彼此之間的關聯，充分地將它們整合在一起，想必是更能發揮數位圖書館在使用者服務功能上的價值。

本研究主要是提出運用智慧型多重代理人技術於數位典藏的資源搜尋與管理的概念，並嘗試採用「多重代理人社群模組」的技術，來整合不同數位典藏的資源，提供使用者能經由代理人社群模組的協同運作，建立文獻資源的搜尋與文獻資源的管理機制。在設計系統同時，我們亦對分散式網路運算架構加以考量，使得不同平台上的軟體間可以互相運用，以加強軟體系統的功能與彈性。

1. 緒論

1.1. 研究動機與目的

巨大的全球資訊網 (Web)，及其背後所隱藏之形形色色的資源是一個

佈滿商機與希望的奇幻世界。對現代人而言，全球資訊網正逐漸地融入每個人的日常生活，並將成為未來人們生活上不可或缺的一項重要資源。自從紙張與印刷術被發明以來，知識傳承的根基日漸穩固，而人類文明的發展更是一日千里。隨著資訊科技不斷地創新，如何應用新科技於知識傳承及知識創新的工作上，是一件刻不容緩的研究。

從 1994 年起，美國國家科學基金會 (NSF)、美國國防部高等研究計畫機構 (DARPA) 以及美國航空暨太空總署 (NASA)，宣佈合資贊助美國境內六所重點大學從事「數位圖書館先導研究計畫 (Digital Libraries Initiative)」。之後加拿大、歐盟各國、日本及俄羅斯等國家也紛紛提出國家型數位圖書館研究計畫。我國也於 1998 年，成立國科會數位博物館先導計畫。

綜觀目前世上所進行的許多數位圖書館研究計畫，每個計畫的主題與方向，可說是上至天文，下至地理，無所不包。而典藏內容更是包羅萬象，多彩多姿，這些分散於不同平台與不同系統的典藏資料，如果能夠依據內容的主題與方向來建立彼此之間

的關聯，充分地將它們整合在一起，想必是更能發揮數位圖書館在使用者服務功能上的價值。

此外，由於個人的時間與精力都是有限的，所以，如果能將智慧型多重代理人 (intelligent multi-agent) 技術，應用於數位典藏的資源搜尋與管理，將可以幫助使用者在個人化知識組織與管理方面，達到省力與省時的效果，更能有效提昇個人在知識學習上的量與質。所以，運用智慧型多重代理人技術於數位典藏的資源搜尋與管理，是當今知識經濟與全球化競爭時代裡一件刻不容緩的研究主題。

1.2. 研究方法

本研究主要是提出運用智慧型多重代理人技術於數位典藏的資源搜尋與管理的概念，並嘗試採用「多重代理人社群模組 (multi-agent community model)」的技術[10][11]，來整合不同數位典藏的資源，提供使用者一個自動化文獻資源搜尋與個人化文獻資源管理的機制。

使用多重代理人社群模組的好處是，我們可以針對不同的數位典藏資源，設計出各具特色的代理人，讓每一個代理人都能擁有它們的專業知識 (domain knowledge)，然後再搭配其它功能的代理人，例如：使用者界面代理人、程序監督代理人、文獻組織管理代理人、印表機代理人等，共同組織成多重代理人的供應鏈系統。所有代理人將採分工合作的方式，共同來解決使用者的需求。此外，由於代理人之間鬆散偶合的關係，使得整個多重代理人社群模組在執行時期，並

不會因為一個代理人產生故障，而影響到整個供應鏈系統的運作流程。

我們將實作出「文獻多重代理人社群 (bibliography multi-agent community)」。文獻多重代理人社群主要包括有多個文獻蒐集代理人所構成。文獻蒐集代理人的主要工作是「幫助使用者到指定的數位典藏單位，依據使用者需求來蒐集文獻資料」。此外，我們也提出文獻管理代理人的概念，其主要工作是「幫助使用者將蒐集到的文獻進行分析，之後再依據主題詞 (topic terms) 來管理這些文獻資料」。

在這當中我們將採用 RDF (Resource Description Framework) 的規範來詮釋 Web 所包含的資訊資源 [2]，藉以幫助代理人能理解這些資訊資源的內容。並運用 Topic Maps 來組織資訊資源[3][4]，以建構出個人化的知識地圖。此外，我們也將採用 Web Services 技術來解決多重代理人架構在分散式網路運算環境當中[1]，於異質性平台上互通性 (interoperability) 的問題。

1.3. 論文結構

本論文將分以下幾個部份來探討如何運用智慧型多重代理人技術於數位典藏的資源搜尋與管理，第二節將先對相關技術做一簡介，之後於第三節提出文獻多重代理人社群模組的系統架構，第四節則介紹代理人的角色與任務，第五節則為本論文之結論及未來展望。

2. 文獻探討

在這一節裡我們首先針對本研究將採用的相關技術分別探討。

2.1. Web Services

在實作多重代理人程式時，我們將遭遇到幾個問題：第一，所有代理人的平台可能不一致；第二，所有代理人所採用的程式語言可能不一致；第三，分散在各地的代理人可能不知道彼此的存在。因此，在目前開放式的網路環境裡，如何克服上述問題，而使得分散於不同平台或使用不同程式語言的代理人能夠順利的互相運作（interoperability），是一項重要的關鍵。採用 Web Services 技術來實作多重代理人程式，上述問題將可迎刃而解，因為 Web Services 提供一致性的資訊傳遞方式與資訊整合服務[1]。

Web Services 是一種基於開放式與標準式的新一代分散式運算技術。因為，它的基本原理是利用標準的 Web 通訊協定 HTTP 來傳送程序之間的呼叫訊息與回應訊息，以避免被企業網路（intranet）上的防火牆所阻擋；而所傳遞的訊息則是利用標準的資訊交換格式 XML 來定義，可避免系統的開發受限於某種平台或程式語言。所以，比起過去所制定太過複雜與獨特界面的分散式運算技術，例如 DCOM、CORBA、Java RMI 等，採用 Web Services 將可有效的減少開發的時間與降低整合的成本[5]。

我們可以簡單的把位於多重代理人社群裡的所有代理人，依據彼此之間的關係，區分為服務提供者（service provider）、服務消費者（service consumer）與服務仲介者（service

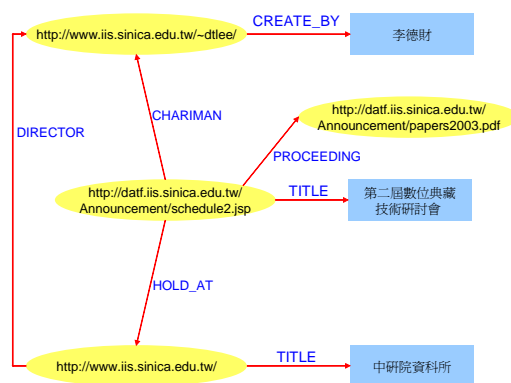
broker）三種角色。而存在於他們之間的溝通與整合問題，我們可分別採用 Web Services 的三項關鍵技術：SOAP（Simple Object Access Protocol）、WSDL（Web Services Description Language）及 UDDI（Universal Description, Discovery and Integration）來解決。以 SOAP 做為多重代理人社群的環境中，代理人彼此之間交換訊息的協定；以 WSDL 來描述服務提供者代理人所提供之功能的規格；以 UDDI 做為 Web Services 的註冊機制，使需要特定 Web Services 的服務消費者代理人，可以藉由 UDDI 的機制，找到所需要的服務。

2.2. 資源描述架構（RDF）

想要在 Web 上自動處理任何事物是一件困難的事，因為儘管在 Web 上的任何事物都是電腦可以讀取的，但卻不表示電腦可以理解這些事物。可是如果要以人工的方式管理 Web 上所包含的大量資訊資源，似乎是力有未逮、事倍功半。所以也不可能完全以人工的方式來管理這些資訊資源。為了解決這樣的問題，我們首先使用詮釋資料（metadata）來描述 Web 包含的資訊資源[8]。詮釋資料就是"資料的資料（data about data）"，我們將採用 RDF 的規範來描述 Web 資源，以幫助代理人能夠理解這些 Web 資源，進而自動處理使用者的需求[2]。

一個 RDF 模型主要是由許多 RDF 陳述式（statements）所構成，其主要功能是用來表達 Web 上的資源。所謂的 RDF 陳述式是指一句資源描述式或一句知識表示法（knowledge representation）。而 RDF 陳述式則是由

資源 (resource)、特性 (property) 及特性值 (property value) 這三部份所構成。資源是指具有一合法「統一資源識別碼 (URI, Uniform Resource Indicator)」的任何事物，它可用「特性-特性值 (property-value pair)」這種架構來描述它所擁有的特性。因此，我們稱「{特性, 資源, 屬性值}」或「{屬性, 資源 1, 資源 2}」這種型式的有序對 (tuple) 為一句 RDF 陳述式。同時可被解讀成「<資源>的<屬性>是<屬性值>」或「<資源 1>是<屬性>於<資源 2>」這類型的資源描述式或知識表示法。例如：「第二屆數位典藏技術研討會的主席是李德財」，可用 RDF 的有序對{主席,[第二屆數位典藏技術研討會],[李德財]}來表示。而「第二屆數位典藏技術研討會是舉辦於中央研究院資訊科學研究所」，則可用 RDF 的有序對{舉辦,[第二屆數位典藏技術研討會],[中央研究院資訊科學研究所]}。此外，也可以用圖形的方式來表示 RDF 模型的設計概念，將有助於設計過程中的構思。例如圖一表示「第二屆數位典藏技術研討會」的 RDF 模型。我們將 Web 資源以 RDF 陳述式具體化 (reification) 地表達出來後，代理人即可經由理解這些 RDF 模型所隱含的語意 (semantic)，而自動去完成其所代理的工作。



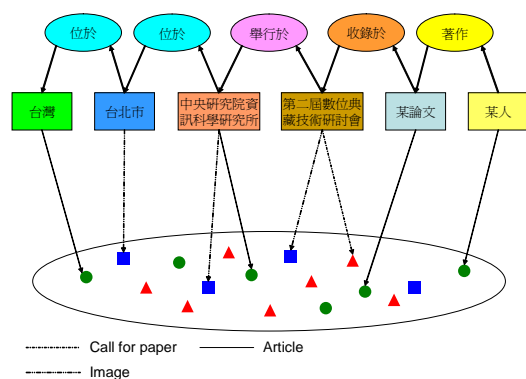
圖一、第二屆數位典藏技術研討會的 RDF 模型

2.3. 主題地圖 (Topic Maps)

隨著 Web 的飛速發展和知識經濟時代的來臨，個人及企業皆面臨排山倒海而來的各種資訊。此時個人及企業能否有效地累積與運用 Web 上豐富的資訊資源，而不至於因資訊過量而引發資訊恐懼或資訊焦慮的問題，落實知識管理將會是個人及企業面對這資訊洪流最好的策略。知識管理的首要工作便是要有一套簡易而完善的知識組織方法。過去的知識組織常用的方法不外乎是分類目錄的建立，可是當要管理 Web 上豐富且動態的資源時，分類目錄有如樹狀架構的特性，將無法及時反應類別的異動及類別之間在關係上的調整。因此，我們將採用主題式的管理方法，使用 Topic Maps 的規範[3][4]，建構一使用者友善且使用者導向的知識地圖，以因應 Web 資源的多變及快速成長。

Topic Maps 就好比是書本後面所附的關鍵詞索引。一份完備的關鍵詞索引，除了完整地列出所有出現於該書本內的重要詞彙之外，更重要的是組織這些詞彙之間的關聯，也就是詞彙之間「見 (See)」及「參見 (See Also)」

的參照關聯，方便讀者可按圖索驥地，輕鬆找到需要的正確資訊。在 ISO/IEC 13250 中對 Topic Maps 的定義為「Topic Maps 是一套用來組織資訊的方法，使用這個方法可以提供最佳的資訊導航」。一個 Topic Maps 的模型主要是由主題 (topics)、連繫各主題與相關資訊資源的資源鏈結 (occurrences) 及存在於主題之間的關聯 (associations) 這三種元素所組成。不過太多的主題也會造成使用者不必要的困擾與無所適從。因此，我們採用資訊遮罩 (information mask) 的技巧，讓使用者可暫時性的將不需要的主題、資源鏈結或關聯隱藏起來。圖二是描述一個會議期刊的 Topic Maps 模型。



圖二、會議期刊的 Topic Maps 模型。

3. 文獻多重代理人社群

3.1. 系統架構

文獻多重代理人社群主要是由使用者界面代理人 (user interface agent)、資源包裹代理人 (resource wrapper)、資源仲介代理人 (resource broker)、文獻蒐集代理人 (bibliography search agent) 等所組成。

當使用者界面代理人收到使用者的查詢請求後，它將經由資源仲介代理人的仲介，委託文獻蒐集代理人來達成文獻蒐集的工作。文獻蒐集代理人可漸接的透過資源包裹代理人來與文獻典藏單位取得溝通。最後，文獻蒐集代理人將蒐集到的文獻資訊，回傳給使用者界面代理人做彙整後，呈現給使用者瀏覽。各代理人的角色與任務，以及整個社群的詳細工作流程，將在下一節中作介紹。

3.2. 通訊協定

使用者界面代理人與使用者之間的溝通主要是透過 BibTex。也就是說，使用者的查詢條件將使用 BibTex 的格式來描述，而使用者界面代理人則會將查詢結果依據 BibTex 的格式呈現給使用者瀏覽。所包含的欄位整理於表一。表二則是一筆 BibTex 的範例。

title	文獻標題	month	出版月份
author	文獻作者	year	出版年份
journal	期刊名稱	publisher	出版單位
volume	期刊編號	institution	贊助單位
number	期刊序號	key	關鍵字
pages	頁數	type	文獻型態

表一、BibTex 欄位

```
@article{avery:98,
  author = {Avery, Christopher},
  year = 1998,
  title = {Strategic Jump Bidding in English Auctions},
  journal = {The Review of Economic Studies},
  volume = 65,
  number = 2,
  month = {April},
  pages = {185-210}
}
```

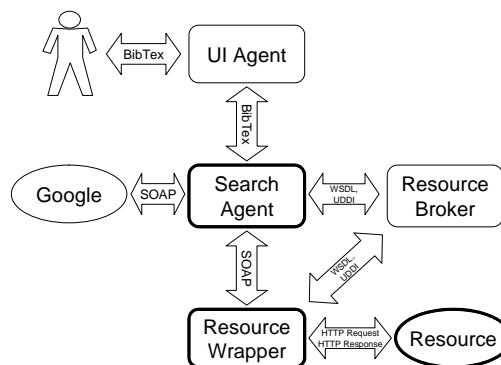
表二、BibTex 範例

由於整個系統將採用 Web Services 的技術來實作。因此，文獻多重代理人社群裡的所有元件在實作時，可依據服務的供需關係，分別使用 Web Services 伺服器端 (server side) 與 Web Services 客戶端 (client side) 的技術來建構。例如：資源包裹代理人的主要功能是提供透通模式 (transparent mode) 服務，讓文獻蒐集代理人與文獻典藏單位彼此間的通訊，能達到開放式與標準式的目的。所以，我們可將資源包裹代理人視為服務提供者，將文獻蒐集代理人視為服務消費者。

另一方面，為了有效的完成使用者的需求，在文獻蒐集代理人與資源包裹代理人之間我們採用集中式的通訊管理機制。也就是說，使用文獻蒐集代理人與所有資源包裹代理人之間的通訊將透過資源仲介代理人來負則協調，所以我們可將資源仲介代理人視為服務仲介者。

這些元件可以是分散於網路上的不同節點，也可以是位於同一節點。它們雙方之間的通訊將遵照 Web Services 的架構來進行。所以，他們彼此之間都是以 XML 格式為基底的訊息來互相溝通，這將有別於傳統的程序呼叫方式，使得代理人之間不論是在功能上的互相運作，或者在服務上的整合，都更為彈性與靈活。除此之外，使用者與使用者界面代理人之間則採用互動式網頁程式的通訊協定來溝通。資源包裹代理人與數位典藏單位之間的通訊方式則必須配合數位典藏單位所提供的通訊協定來進行。Google 是目前少數有提供 Web

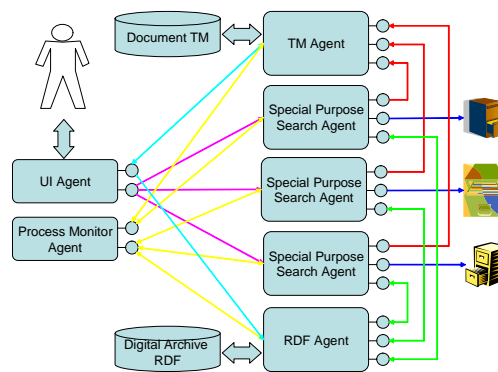
Services 機制的資源，因此我們的文獻蒐集代理人可直接以 Web Services 的架構來與其溝通。圖三為文獻多重代理人社群裡各個元件之間的通訊協定。



圖三、通訊協定

4. 代理人的角色與任務

我們的文獻代理人社群系統其主要工作是用來幫助使用者於 Web 上自動尋找書籍、期刊、論文等資料[9]。為了達成此目標，除了使用者界面代理人、資源包裹代理人、資源仲介代理人、文獻蒐集代理人之外，我們還需要有 Topic Maps 代理人、RDF 代理人及程序監視代理人的輔助，圖四為這些代理人的簡圖。



圖四、代理人的簡圖

4.1. 使用者界面代理人 (UI Agent)

使用者界面代理人的工作主要是與使用者互動，其工作流程首先是了解使用者的需求，然後與其它資源搜尋代理人聯絡，並提出搜尋請求，最後將資源搜尋代理人回傳的結果彙整輸出給使用者。

使用者界面代理人可以和 Topic Maps 代理人合作，依據使用者的查詢條件來建立個人化文獻資源的 Topic Maps，而當使用者作下一次查詢時，Topic Maps 代理人可以將 Topic Maps 裡的知識告訴使用者界面代理人，讓使用者界面代理人有更多的資訊，可以作更精確完整的查詢。

4.2 Topic Maps 代理人 (TM Agent)

Topic Maps 代理人的主要作用有二項。第一項是依據 Topic Maps 對使用者搜尋回來的個人文獻作管理。第二項是當使用者作查詢時，根據 Topic Maps 提供使用者界面代理人更多的資訊，以作更精確完整的查詢。

Topic Maps 的內容是依據使用者的查詢記錄累積而成的。Topic Maps 代理人將分析使用者的查詢條件，建立查詢條件之間的關聯；另一方面，Topic Maps 代理人將從專用搜尋代理人回傳的結果與使用者的反應意見來建立此一 Topic Maps 的資源鏈結 (occurrence) [6]。

4.2. 專用搜尋代理人 (special purpose agent)

由於 Web 上的許多數位典藏單位，不論是在資料儲存的格式或是檢索機制的運作都各有異同，我們很難

只由一個資源搜尋代理人來負責與所有的數位典藏單位溝通。因此，我們依據數位典藏單位的相似性，設計有四類專用搜尋代理人。分別是：第一類、圖書館資源專用搜尋代理人 (library agent)；第二類、電子期刊資源專用搜尋代理人 (E-journal agent)；第三類、電子商務資源專用搜尋代理人 (E-commerce agent)；第四類、Web 資源專用搜尋代理人 (Web agent)。每一個專用搜尋代理人主要是由一組資源包裹代理人與文獻蒐集代理人所組成。

圖書館資源專用搜尋代理人其主要任務是負責代理一般大學圖書館及政府圖書館的線上書籍查詢作業；電子期刊資源專用搜尋代理人其主要任務是負責代理各學術領域裡，主要電子期刊 (例如：ACM、IEEE) 的線上論文查詢作業；電子商務資源專用搜尋代理人其主要任務是負責代理各電子商務網站 (例如：Amazon、博客來等網路書店) 所提供的線上產品查詢作業；而 Web 資源專用搜尋代理人其主要任務是負責代理各網頁搜尋網站 (例如：Google、Openfind 等 Web 資源檢索網站) 所提供的網頁查詢作業。

為了讓專用搜尋代理人具有自我調適 (self-adaptive) 搜尋方式的能力，我們使用 RDF 來描述每一個數位典藏單位本身所具備的搜尋機制。這些專用搜尋代理人將透過 RDF 代理人的輔助來調整搜尋方式。如此，專用搜尋代理人的能力就不會因數位典藏單位所提供的搜尋機制改變了而無法正常運作。

5. 結論及未來展望

隨著知識經濟與全球化競爭時代的來臨，預期數位典藏的重要性將從以往只是負責儲藏歷史紀錄與知識傳承的功能，更進一步地往知識整合與知識創新邁進。為了能夠整合所有數位典藏單位的典藏內容，我們設計了一個應用於數位典藏的資源搜尋與管理的代理人社群模式，我們將完成界面代理人、文獻蒐集代理人、程序監督代理人及個人文獻管理代理人。當使用者經由界面代理人來設定搜尋的條件之後，位於代理人社群裡的專業文獻蒐集代理人，將於 Web 上各司其職，進行文獻的蒐集。之後再交由文獻管理代理人，負責個人化的文獻組織與管理工作。整個過程當中程序監督代理人將負責監督所有交易伙伴的交易順序與情形。

本計畫的未來展望是結合印表機代理人，它的功能是幫助使用者輕鬆的完成各種文獻類型的列印工作。在列印的過程中，它必須能夠辨識文獻的儲存格式，判斷是否要先進行格式的轉換，接下來它必須負責找到一台符合使用者需求且待命中 (ready) 的印表機進行列印，最後它還必須保證印表工作的成功。在未來的研究裡，我們將繼續完成此部份。

此外在 Web Services 技術的應用與研究中，我們發現安全性問題與服務品質的保證是使用 Web Services 技術的最大隱憂。因此，如何能在兼具開放式與標準式的環境底下，提供代理人社群模式一個安全與可靠的協同運作管理技術，以及如何加強代理人的強韌性，以使得使用者在運用多重

代理人技術時，提供更好的服務品質，進而更有效率地獲得豐富且完整與正確的訊息。這兩點亦是未來值得繼續研究的方向。

參考文獻

1. Venu Vasudevan, "A Web Services Primer.", In: XML.com, April 04, 2001, URL: <http://webservices.xml.com/pub/a/ws/2001/04/04/webservices/index.html>
2. Tim Bray, "What is RDF?", In: XML.com, January 24, 2001 URL: <http://www.xml.com/pub/a/2001/01/24/rdf.html>
3. Lars Marius Garshol, "What Are Topic Maps?", In: XML.com, September 11, 2002, URL: <http://www.xml.com/pub/a/2002/09/11/topicmaps.html>
4. Members of the TopicMaps.Org Authoring Group. "XML Topic Maps (XTM) 1.0 -- TopicMaps.Org Specification.", Aug 2001.
5. R.V. Guha and Rob McCool, "A System for integrating Web Services into a Global Knowledge Base.", In: Proceedings of Semantic Web Workshop 2002, May 2002.
6. R.V.Guha and Rob McCool, "TAP: A Semantic Web Platform.", In: XML 2002 - Conference Proceedings, December 2002.
7. Shankar R. Ponnekanti, Armando Fox, "SWORD: A Developer Toolkit for Web Service Composition.", In: Proceedings of the WWW2002 - Eleventh International World Wide Web Conference, May 2002.
8. Heiner Stuckenschmidt and Frank van Harmelen, "Knowledge-Based Validation, Aggregation and Visualization of Meta-data: Analyzing a Web-Based Information System.", In: Proceedings of The First Asia-Pacific Conference on Web Intelligence (WI'2001), October 2001, Maebashi City, Japan.
9. In-Young Ko, Robert Neches and Ke-Thia Yao, "Semantically-Based Active Document Collection Templates for Web Information Management Systems", In: Proceedings of the ECDL 2000 Workshop on the Semantic Web, September 2000, Lisbon, Portugal.
10. 李步健(民 87)，區域網路之代理型程式與服務共享架構。國立台灣大學資訊工程研究所碩士論文。
11. 駱克明(民 88)，建構於分散式物件之上的多重助理程式合作架構。國立台灣大學資訊工程研究所碩士論文。