

第一章 緒論

第一節 研究動機

受到網際網路普及的影響，網路資源不只為圖書館提供了新的服務機會，也帶來了新的挑戰。圖書館界於網際網路發展初期即積極嘗試以編目、分類、與自動化技術來組織整理網路資源，但由於圖書館整理的資源向來是以實體為主，具有同質性高、定義嚴謹及穩定等特性，而網路資源相較於圖書館的資源，則具有大量、異質、分散、內容與位址經常改變、及成長快速等特性。因此若套用傳統上圖書館對印刷型式資源採用的人工編目或分類等方式，來組織與表徵網路資源，就可行性與經濟性來看，並不見得可行。

至於如何克服網路資源特性，以提昇對網路資源組織及知識表徵的成效，近年來電腦科學界普遍認為，解決之道是需對資源內容添入語意標記 (Semantic Markup)，並藉由「知識本體」(以下以 Ontology 表示)，以電腦能處理的方式來明確表徵資源的語意，協助機器自動處理與資源互通，進一步實現知識分享與複用的願景。

雖然圖書館的索引典與詮釋資料架構 (Schema) 亦屬 Ontology，但索引典面臨提煉的語意關係過於簡單，概念間的關係缺乏明確的語意，及結構可能不具一致性，以致難以協助電腦自動處理等問題；詮釋資料架構的設計通常以特定主題領域為主，標記語法亦不盡相同，同樣有資訊互通的問題。而人工智慧領域過去雖已發展一些知識表徵語言，例如 CycL、Loom、與 KIF 等，但這些語言一般相當複雜，並都有其自行採用的推論規則，亦不見得可用來處理網路環境之半結構式資源。

換言之，網路知識表徵需將網路資源的特性納入考量，重新思考與發展合適的處理方法。目前「全球資訊網協會」(World Wide Web Consortium，以下簡稱 W3C) 及「國際標準組織」(International

Organization for Standardization，以下簡稱 ISO）在 2000 年前後，分別提出「資源描述架構」(Resource Description Framework，以下簡稱 RDF) 與「主題地圖」(以下以 Topic Maps 表示)，作為可應用於網路資源組織與知識表徵方法。RDF 與 Topic Maps 雖然在概念上看似有許多相似之處，而且均是可用來表達資源間之關係，並可為 Ontology 提供語意互通的資料模型 (Data Model)，但事實上它們是兩個不同的語法家族。

最早的 RDF 規範是 W3C 於 1999 年制定和公布的，其可用於標示資源與資源間的關係，並作為邏輯推論 (Logical Inferencing) 的基礎；而 Topic Maps 則是為了提昇索引 (Index) 方法而發展的，並於 2000 年初期時成為 ISO 國際標準 (ISO/IEC 13250:2000)，TopicMaps.Org 組織更在 2001 年時將其應用於全球資訊網，制定公布了 XML Topic Maps (XTM) v1.0 規範；換言之，這兩種方法最初發展的目的並不相同，它們是為了解決不同問題而發展出來的不同方法。雖然 RDF 與 Topic Maps 皆是可用於網路資源組織及知識表徵的資料模型，但它們有各自的交換語法 (Interchange Syntax) 及語意限定語法 (Constraint)。(Garshol, 2003)如下圖所示。

TMCL	Constraints	OWL
		RDF Schema
Topic Maps	Data Models	RDF
XTM, HyTM, LTM	Syntaxes	RDF/XML, N3

圖 1-1 RDF 與 Topic Maps 之語法家族¹

資料來源：Garshol, L.M.(2003). Living with Topic Maps and RDF: Topic Maps, RDF, DAML, OIL, OWL, TMCL. Retrieved July 23, 2004, from http://www.idealliance.org/papers/dx_xmle03/papers/02-03-06/02-03-06.html

¹ XTM 是 XML Topic Maps 的縮寫；HyTM 是 Hypermedia/Time-based Structuring Language 的縮寫，最初是為 SGML 提供多媒體與超連結功能而發展，1992 年成為 ISO 10744:1992，1997 年更新為 ISO 10744:1997；LTM 是 Linear Topic Map Notation 的縮寫，是表示 Topic Maps 的一種語法；N3 是 Notation 3 的縮寫，是表示 RDF 的一種語法；TMCL 是 Topic Map Constraint Language 的縮寫，2005 年 2 月剛公布為 ISO/IEC JTC1/SC34 工作草案；OWL 是 W3C 提出的網路知識本體語言 (Web Ontology Language)。

綜上所述，如何克服網路資源特性，以提昇對網路資源組織及知識表徵的成效，除了是圖書館界關心的課題外，同樣也是電腦科學界投入研究與發展的焦點。而 RDF 與 Topic Maps 皆是可應用於網路資源組織與知識表徵的方法，它們都在表徵人類的知識關聯，不過卻是來自不同的語法家族，這兩種語言在知識關聯程度的表徵上到底有何不同、孰強孰弱、適用範圍為何等，都是本研究想探討的課題。研究結果除將有助於瞭解其各自在知識表徵方法之異同外，亦期盼能帶給圖書館對網路資源組織及知識表徵，在方法上的不同觀點，以供參考。

第二節 研究目的

本研究之目的在探討網路知識表徵之發展趨勢，並分析比較 RDF 與 Topic Maps 在知識表徵方法之異同。具體而言，研究目的分述如下：

- 一、瞭解網路知識表徵之發展趨勢。
- 二、探究及比較 RDF 與 Topic Maps 在表達概念或主題關聯之語法、所能表達的語意、及應用情境等問題。
- 三、根據研究結果，除瞭解和辨明 RDF 與 Topic Maps 在語法、所能表達的語意、及應用情境等問題之異同外，亦期盼能帶給圖書館對網路資源組織及知識表徵，在方法上的不同觀點，以供參考。

第三節 研究問題

本研究主要在探討網路知識表徵之發展趨勢，並分析比較 RDF 與 Topic Maps 在知識表徵方法之異同。根據研究目的，將研究問題分述如下：

- 一、網路知識表徵之發展趨勢為何？
- 二、RDF 與 Topic Maps 在表達概念或主題關聯之語法、所能表達的語意、及應用情境為何？例如可否描述階層關係、可否指定類與實例、以及是否可協助資源整合或資源瀏覽等。
- 三、RDF 與 Topic Maps 在圖書館網路資源組織應用之可行性為何？例如資源組織方式是否更具彈性、是否有助於資源整合或分享等。

第四節 名詞解釋

本研究對文中談及之相關重要名詞，分述如下：

一、知識表徵 (Knowledge Representation)

知識表徵向來是人工智慧領域關心的重要課題，其主要在探討如何以正規的 (Formal) 方式在資訊系統中儲存及處理知識。目前已發展出許多知識表徵的語言 (例如 OWL、CycL 與 Loom 等) 和標記法 (例如 RDF 與 DATR 等)。這些語法一般都會以邏輯和數學為基礎，並具有易於電腦剖析的文法。(Wikipedia, 2004a)本研究欲探討的知識表徵是依其在電腦科學與圖書資訊學之定義，不涉及其在其他領域 (例如哲學與心理學等) 之討論。

二、資源描述架構 (Resource Description Framework, 簡稱 RDF)

RDF 是由 W3C 制定發展的規範，它是一種在網路環境中，可用於標示資源與資源間之關係，並可作為邏輯推論之基礎的資料模型。RDF 所表示的敘述 (Statement) 皆由「三元素」(Triples) 組成，分別為「主詞」(Subject)、「述語」(Predicate)、與「受詞」(Object)。所有以 RDF 來描述的事物皆可稱為「資源」，例如網頁、文獻、與人等。「述語」亦可稱為「屬性」(Property)，作用是標示主詞和受詞之關係。「受詞」亦可稱為「屬性值」。(Klyne & Carroll, 2004) RDF 也提供了 XML 的語法 RDF/XML 來標記及交換 RDF 圖形。本研究將以 W3C 於 2004 年 2 月 10 日公布的 RDF 規範為研究對象。本研究提及 RDF 時，RDF Schema 與 RDF/XML 亦在討論之內。

三、主題地圖 (Topic Maps)

Topic Maps 有三個基本核心概念，分別為「主題」(Topic)、「關聯」(Association)、與「資源指引」(Occurrence)。「關聯」的作用是能將多個主題關聯在一起；而「資源指引」，則負責連結和主題有關之資源。簡言之，Topic Maps 透過定義資源之主題、關聯、及資源指引，可作為網路知識表徵的方法。

表示 Topic Maps 的語法有二，一為 ISO/IEC 13250:2000，另一個則是 TopicMaps.Org 組織制定的 XML Topic Maps (以下簡稱 XTM)。XTM v1.0 是將 ISO/IEC 13250:2000 加以修改，應用於全球資訊網的語法規範。本研究將以 XTM v1.0 為研究對象。本研究提及 Topic Maps 時，XTM 語法亦在討論之內。

四、知識本體 (Ontology)

本研究談論的 Ontology 是依其在電腦科學界之定義，不涉及其在哲學領域之討論，因此將 Ontology 譯為「知識本體」而非「本體論」，以茲區別。

Ontology 是對概念的詳盡敘述，(Gruber, 1993)就電腦科學界而言，是用於描述及表達特定領域知識的一組概念或術語，其可用來標示特定領域知識的類 (Class)、類的屬性，以及類和類間的關係，協助達成以概念和語意為本的資訊檢索；(Berners-Lee, 2000)換言之，Ontology 是嘗試為某特定領域訂定詳盡且精確的概念架構 (Conceptual Schema)，而這個概念架構通常是由包含該領域之相關實體 (Entity)、關係、及規則 (Rule) 結合而成。Ontology 在人工智慧及知識表徵的應用較為常見。透過電腦程式，可將 Ontology 作多種用途，例如歸納式推論、分類、以及協助不同系統間之資訊溝通和分享。目前有關 Ontology 的語法，包括「網路知識本體語言」(Web Ontology Language, 簡稱 OWL)、CycL、及「知識交換格式」(Knowledge Interchange Format, 簡稱 KIF) 等。(Wikipedia, 2004b)

五、語意網 (Semantic Web)

「語意網」主要是由全球資訊網發明人 Tim Berners-Lee 倡導的概念與願景，他強調「語意網」並非要取代目前的全球資訊網，而是目前全球資訊網的延伸，願景在使全球資訊網上之資訊資源皆具有定義明確的語意，以促進電腦的自動處理及電腦和人類的合作。(Berners-Lee, 2000)W3C 更於 2001 年 2 月啟動「語意網活動」(Semantic Web Activity)，目標在發展能支援資訊發掘、資源整合、及自動化處理的標準與技術。

六、名稱空間 (Namespace)

名稱空間 (以下均以 Namespace 表示) 在 XML 中有兩個目的：一是為了區分不同 XML 應用中，使用同樣名稱的元素和屬性；二是為了在單一 XML 應用中歸類相關的元素和屬性，讓軟體能容易地識別它們。Namespace 在實作上是將每個元素和屬性加上一個前置詞 (Prefix)，每個前置詞都對應到一個 URI。附屬於同一個 URI 的元素和屬性，會屬於同一個 Namespace，以消除同名所產生的模稜兩可。具有相同名稱卻有不同 URI 的兩個元素，視為不同；反之，名稱和 URI 都相同者，才是相同的元素。Namespace URI 並非一定要指到一

份實際存在的文件或網頁，它只是識別字而已。(Harold & Means, 2001)

例如欲區分某份 XML 文件中使用的 RDF 之 Description 元素不是都柏林核心集 (Dublin Core) 的 description 元素，那麼就宣告 Namespace 如 `xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"` ，藉由 `xmlns:rdf` 屬性為 `rdf:Description` 元素和它的子元素們宣告了前置詞 `rdf`，那麼 RDF 處理器可以認出 `rdf:Description` 是 RDF 元素，因為其具有置詞 `rdf`，該前置詞連結到 RDF 規範中定義的特定 URI `"http://www.w3.org/1999/02/22-rdf-syntax-ns#"` 。在冒號之前的所有內容，稱作前置詞；在冒號之後的所有內容，稱作本地部分 (Local Part)。包含冒號在內的完整名稱，稱作全稱名稱 (Qualified Name 或 QName) 或未處理的名稱 (Raw Name)。(Harold & Means, 2001)

七、主題 (Subject or Topic)

在 XTM v1.0 規範中 “Subject” 和 “Topic” 的意義有所不同，在規範中稱以 “Topic” 表徵 “Subject” 於 Topic Maps 文件的過程為「具體化」(Reification)；換言之，“Subject” 是指任何可被人們表達或想像的事物，而 “Topic” 是指撰寫 Topic Maps 文件之作者，企圖在 Topic Maps 文件中將 “Subject” 予以表徵之事物。但在中文的環境中，通常 “Topic” 和 “Subject” 都是譯為「主題」，因此，本文在翻譯上並不特別區分這兩個單字，皆譯為「主題」，但若遇原文為 “Subject” 時，會在中文後頭特別括號將英文標示出來。

八、屬性 (Property or Attribute)

在中文環境中 “Property” 與 “Attribute” 通常皆譯「屬性」，但在 RDF/XML 中其意義有所不同，“Property” 是指 RDF 三元素中的屬性，而 “Attribute” 則是指 XML 語法所指的屬性。本文在翻譯上並不特別區分這兩個單字，皆譯為「屬性」，但若遇原文為 “Attribute” 時，會在中文後頭特別括號將英文標示出來。