

以主題地圖建構索引典之語意網路模型(註 1)

Constructing a Semantic Network Model of Thesaurus: Topic Maps Approach

林信成 Sinn-Cheng Lin
淡江大學資訊與圖書館學系副教授
Associate Professor
Dept. of Information and Library Science
Tamkang University
E-mail: sclin@mail.tku.edu.tw

歐陽慧 Hui Ou-Yang
淡江大學資訊與圖書館學系研究生
Graduate Student
Dept. of Information and Library Science
Tamkang University
E-mail: 691070014@s91.tku.edu.tw

歐陽崇榮 James C. Ouyang
淡江大學資訊與圖書館學系助理教授
Assistant Professor
Dept. of Information and Library Science
Tamkang University
E-mail: cjouyang@mail.tku.edu.tw

【摘要 Abstract】

主題地圖是一種分散式知識表示法，可以有效的整合網路上的知識庫，從而建構出一個分散式知識管理系統。本文首先從主題地圖之發展切入，介紹其相關標準及其內涵；接著探討知識組織、書後索引、術語彙編與索引典等相關概念；然後實際以教育論文線上資料庫的中文教育類詞庫中，選擇十九個主題詞及其替代詞作為實驗用素材，歸納出其詞間關係並建構語意網路模型，再將此語意網路依據國際標準組織的主題地圖規範進行實作，並在伺服器端建置主題地圖處理器，成功的完成一個實驗性質的線上索引典；最後，我們從主題地圖之相關應用議題與未來發展探討其潛力與限制。

Topic map is a distributed method of knowledge representation; it can integrate the knowledge on the network and build a distributed knowledge management system. This paper begins with a survey of the standards and content of topic maps and follows by concept definitions of knowledge organization, index, glossary and thesaurus. Then, an experiment is run to construct a semantic network model by selecting 19 descriptors and their non-descriptors from the thesaurus. An experimental on-line thesaurus is completed with topic map processor successfully installed on the server. The paper ends with a discussion on the potentiality and limitation of topic maps.

關鍵詞 Keyword

主題地圖 索引典 語意網路 知識組織 人工智慧 知識表達

Topic maps ; Thesaurus ; Semantic networks ; Knowledge organization ; Artificial intelligence ; Knowledge representation



壹、前言

知識 (Knowledge) 是智慧的結晶, 更是文明的重要資產, 人類經由認知、學習、思考等心智活動, 以獲取、創造、累積知識, 並利用各種媒體, 儲存、傳播、共享知識。長久以來各個領域的研究者無不投注心力, 從不同的角度、不同的觀點來研究知識的產生與運用。知識工程 (Knowledge Engineering) 是人工智慧 (Artificial Intelligence, 簡稱 AI) 領域中的一個分支, 主要是一門研究如何擷取、表達、處理知識的科學方法與技術。而在知識工程中最基本的研究主題之一是知識表示法 (Knowledge Representation, 簡稱 KR), 乃是研究如何將知識轉化為數理模型的基本方法。目前, 已被提出的眾多知識表示法中, 較重要的有: 規則式 (Production Rules) 表示法、邏輯與集合 (Logic and Sets) 表示法、語意網路 (Semantic Nets) 表示法、框架 (Frames) 表示法、劇本 (Scripts) 表示法、物件導向表示法 (Object-Oriented Representation) 等。(註 2) 其中, 語意網路是由概念 (Concept) 與概念關係 (Conceptual relations) 組成概念圖 (Conceptual graphs), 可有效表示人類專家知識, 常被用來建構專家系統 (Expert System) 的知識庫 (Knowledge base), 以支援推論引擎 (Inference engine) 的推理與運作。另一方面, 在圖書資訊界中, 為了有效的將資訊適當的分析、拆解、封裝與組織, 讓使用者能快速、正確的找尋到所需資料, 再加以選擇、利用、吸收, 內化成自己的知識, 進而滿足其資訊需求, 也發展出一個極為重要的研究領域: 知識組織 (Knowledge Organization, 簡稱 KO)。在圖書館學中, 索引 (Indexes)、術語彙編 (Glossaries) 與索引典 (Thesaurus) 是幾種較重要的知識組織法。

本文所要探討的主題地圖 (Topic Maps) 是一種類似語意網路的知識表示法, 十分適用於管理資

訊連結, 如術語彙編、交互參照、索引典以及目錄等。它結合了傳統索引、圖書館學與人工智慧等領域的優點, 在資訊世界中, 有如 GPS 般提供定址與連結的功能(註 3), 可以有效的表達與組織知識, 以利於探索、推理。

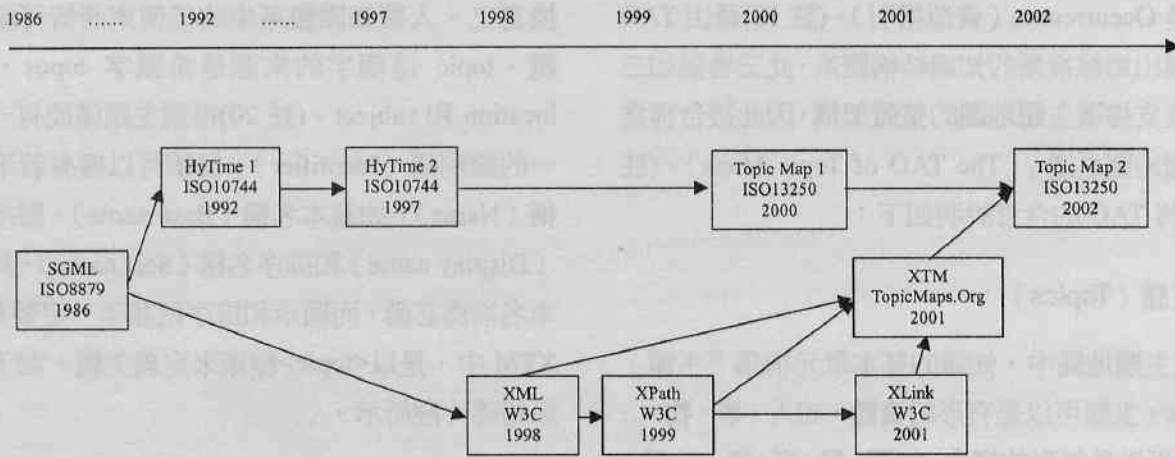
由於主題地圖之知識表達是結合語意網路的基本模型與書後索引, 因此, 在本研究中, 我們嘗試以主題地圖來表現索引典的詞彙結構, 利用主題地圖提供知識管理者自行定義結構的方式, 訂定索引的詞彙關係, 建立各個詞彙後, 再建立彼此間的關連性, 提供使用者以瀏覽的方式來瞭解語彙間的關連性。本文首先從主題地圖之發展切入, 探討其相關標準及技術內涵; 接著闡述知識組織、書後索引、術語彙編與索引典等相關概念; 然後提出以主題地圖建構索引典之模型、方法與步驟; 再者, 實際以「教育論文線上資料庫」(Educational Documents Online, 簡稱 ED-Online) 中的中文教育類詞庫, 選擇十九個主題詞為例, 利用其詞間關係建構語意網路圖, 並據以構成主題地圖; 最後, 我們在 Server 端以 Ontopia 公司的 Omnigator 軟體(註 4)作為主題地圖的處理器, 提供一個實驗性質的導覽式線上索引典。

貳、主題地圖

一、主題地圖的發展

主題地圖是 2000 年由國際標準組織 (International Organization for Standardization, 簡稱 ISO) 和國際電機技術委員會 (International Electro-technical Commission, 簡稱 IEC) 聯合制訂的國際標準 (ISO/IEC 13250:2000) (註 6), 並於 2002 年修訂發佈第二版 (ISO/IEC 13250:2002 Second edition)。其發展與 SGML 和 XML 息息相關, 圖一顯示其演進歷程。一方面採用承襲自 SGML 的 HyTime 語法來定義; 另一方面則另發展出以 XML 為基礎的 XTM, 以更適用於網路環境。





圖一：主題地圖的演進歷程(註 5)

(一) SGML/HyTime 與 ISO Topic Maps

眾所周知，SGML(Standard Generalized Markup Language)是由 ISO 於 1986 年頒佈的國際標準(ISO 8879)(註 7)，具有跨平台、可攜性、結構化、移植性、重覆使用等特性，提供了一個製作電子文件的國際標準；它設計週詳、功能龐大，以致很快地便成為各界遵循規範。HyTime 則是 SGML 的一個應用(Application)，提供多媒體與超連結(Hyperlinking)功能(註 8)，於 1992 年正式成為 ISO 10744 標準。而 Topic Maps 則是由圖形傳播學會(Graphic Communication Association，簡稱 GCA)(註 9)贊助的研究計畫 CapH(Conventions for the Application of HyTime)所發展出比 HyTime 更容易瞭解與應用的子集(註 10)，該計畫提出一個可以記錄內含於書後索引的知識架構，稱為「主題導航地圖」(Topic Navigation Maps)，乃「主題地圖」的前身。這項工作於 1996 年轉移至 ISO 委員會，並於 2000 年 1 月正式頒訂編號為 ISO 13250 的主題地圖標準。(註 11)由於 ISO 13250 採用 ISO 10744 HyTime 標準來定義主題地圖的語法，是一個 SGML DTD，因此又被稱為 HyTM(HyTime Topic Maps 的簡稱)。

(二) XML 與 XML Topic Maps (XTM)

當 HyTM 完成後，由於不是採用 XML 語法，使其應用上頗受限制。因此乃由 TopicMaps.Org 組織(註 12)著手制訂一個採用 XML 的新語法，即是 XTM(XML Topic Maps)(註 13)，並為多數人所採納。XML 是由 W3C(World Wide Web Consortium)於 1998 年 2 月發佈的標準(註 14)，旨在簡化 SGML 並使其更切合網路應用。由於 XML 的出現與所受到的支持，促使了 XTM 的產生。在 ISO 13250 出版的同時，制訂 XTM 規格的行動也同步開始，由 IDEAlliance 贊助、Newcomb 與 Biezunski 創立的 TopicMaps.Org 所負責，在不到一年的時間內，於 2000 年 12 月 4 日於華盛頓 DC 所舉辦的 XML2000 會議宣布初稿，並在 2001 年 3 月 2 日發佈 XTM 第一版(註 15)；到了 2001 年 10 月，XTM DTD 被 ISO 13250 所採納，因此在 2002 年修訂的 ISO 13250 第二版同時包含有 HyTM 與 XTM 兩種語法。(註 16)

二、主題地圖之道

主題地圖的主要概念是 TAO 三要素(註 17)：T 是 Topics(主題)，A 是 Associations(關聯)，而



O 則是 Occurrences (資源指引)。(註 18)藉由 TAO 可建構出錯綜複雜的知識結構體系,此三者猶如三大支柱支撐著主題地圖的整體架構,因此被合稱為「主題地圖之道」(The TAO of Topic Maps)。(註 19)茲將 TAO 的含意說明如下:

(一)主題 (Topics)

在主題地圖中,知識的基本單元稱為「主題」一 topic。主題可以是有形的實體,如人、事、物 ... 等;也可以是無形的概念,如喜、怒、哀、樂 ... 等;

換言之,人類知識體系中的任何東西皆可視為主題。topic 這個字的來源是希臘字 topos,代表 location 和 subject。(註 20)每個主題僅能有一個唯一的識別碼 (Identifier),但卻可以擁有若干個名稱 (Name),如基本名稱 (Base name)、顯示名稱 (Display name) 和排序名稱 (Sort name),其中基本名稱為必備,而顯示和排序則並不一定要有。在 XTM 中,是以 <topic> 標籤來定義主題,如下圖的原始碼片段所示。

```
<topic id="sclin">
  <baseName>
    <baseNameString>林信成</baseNameString>
  </baseName>
</topic>
```

圖二: XTM 原始碼片段: 以 <topic> 定義一個主題

主題可以被歸類成群,稱為主題類型 (Topic types)。換言之,主題類型就是主題所歸屬的類別。一個主題可以歸屬一個以上的主題類型,主題類型在主題地圖中也被定義為一個主題。比方說,

人、哺乳類、動物分別都是主題,但人同時也分屬於哺乳類、動物這兩個主題類型。在 XTM 中,也是以 <topic> 標籤定義主題類型,如下圖的原始碼片段所示。

```
<!-- A Topic Type -->
<topic id="teacher">
  <baseName>
    <baseNameString>老師</baseNameString>
  </baseName>
</topic>
<!-- An Individual Topic -->
<topic id="sclin">
  <instanceOf>
    <topicRef xlink:href="#teacher"/>
  </instanceOf>
  <baseName>
    <baseNameString>林信成</baseNameString>
  </baseName>
</topic>
```

圖三: XTM 原始碼片段: 以 <topic> 定義一個主題類型並加以引用



(二)資源指引 (Occurrence)

對某特定主題更進一步加以描述的相關資源，即稱為該主題的「資源指引」。一個主題可擁有一個以上的資源指引，連結至可定址 (Addressable) 的資訊資源，如某人的出生證明、結婚證書、出版著作、個人網頁或一小段簡介文

字 ... 等。資源指引可以內含在主題地圖內，也可以獨立在主題地圖之外，透過諸如 HyTime addressing(註 21)或 XPointer(註 22)等機制來定址。在 XTM 中，是以<occurrence>標籤來定義主題之資源指引，如下圖的原始碼片段所示。

```
<topic id="sclin">
  <instanceOf>
    <topicRef xlink:href="#teacher"/>
  </instanceOf>
  <baseName>
    <baseNameString>林信成</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#website"/>
    </instanceOf>
    <resourceRef xlink:href="http://mail.tku.edu.tw/sclin"/>
  </occurrence>
</topic>
```

圖四：XTM 原始碼片段：以<occurrence>定義主題之資源指引

(三)關聯 (Association)

關聯主要是用以描述不同主題之間的語意關係，例如「羅貫中」和「三國演義」兩主題之間具有「寫作」關係。不同於資源指引連結到文件來源，關聯表現出一個包含資訊本質、呈現資訊主要價值的知識基礎，一個主題關聯並未限制相關主題的數量。在主題地圖中，關聯也同樣被視為一個主題，也有關聯類型 (Association type)，如「寫作」即可視為一種關聯。關聯類型把具有相同關係的主題匯集成群，有助於增加主題地圖的表達能力。在 XTM 中，係以<association>標籤來定義主題之間的關聯，如圖五的原始碼片段所示。

總而言之，TAO 以主題描述知識組織及架構，

主題可以被對應到其他主題或現實世界中的媒體物件，並利用關聯建立主題間的關係，而資源指引則是指引到一個屬於該主題資源的可識別載體。(註 23)

三、XTM 標籤集及其層級結構

主題地圖就其本質而言是很簡單的，在 XTM 規格書(註 24)中所定義的標籤總共只有十九個，如表一所示。然而，善用這十九個簡單的標籤，卻可以交織出主題之間錯綜複雜的關聯，進而表達出知識的網狀拓樸。



```

<association id="advisor-student-association">
  <instanceOf>
    <topicRef xlink:href="#advrel"/>
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#advisor"/></roleSpec>
    <topicRef xlink:href="#sclin"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#student"/></roleSpec>
    <topicRef xlink:href="#ssw"/>
  </member>
</association>
    
```

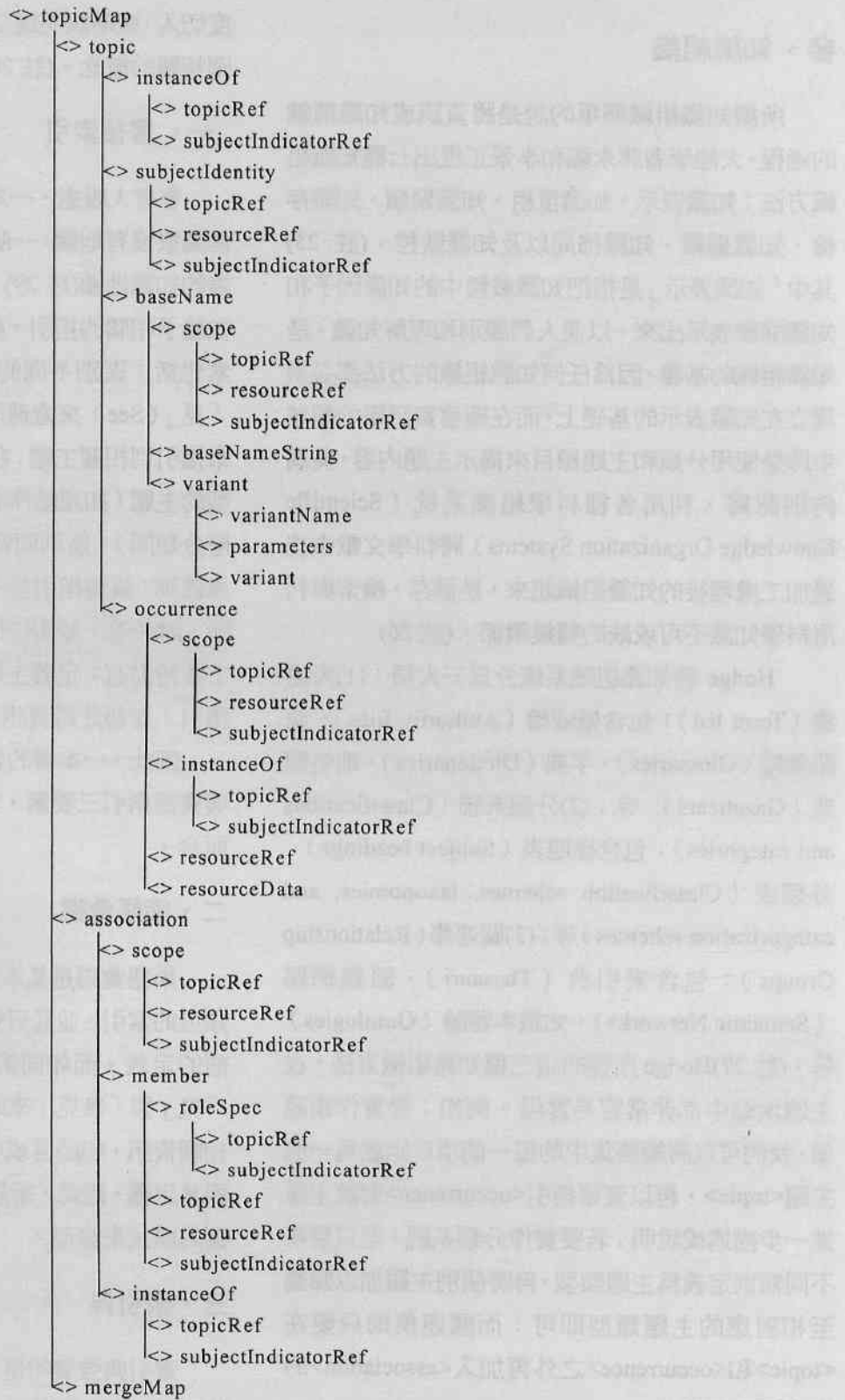
圖五：XTM 原始碼片段：以<association>定義主題之間的關聯

表一：XTM 標籤集

標 籤	用 途
<topicMap>	XTM 的根元素 (Root element)
<topic>	用來定義主題或主題類型
<topicRef>	作為參照到某一個特定主題之用
<instanceOf>	用來將某一主題歸屬於某一類型，以表達主題與類別之間的上下層級關係
<baseName>	定義主題之基本名稱 (Base name)
<baseNameString>	標示主題之基本名稱字串 (Base name string)
<variant>	定義主題之別名
<parameters>	說明主題之別名為何種用途：顯示用、排序用 ... 等
<variantName>	實際的主題別名
<scope>	定義範圍
<occurrence>	標示主題所對應的資源指引
<resourceData>	用以標示一段描述資源的文字資料
<resourceRef>	用以標示一個描述資源的參考位址
<subjectIdentity>	標示某主題為具備唯一識別碼之全域性公用主題
<subjectIndicatorRef>	用以將主題參照至某一個具唯一識別碼之全域性公用主題
<association>	用以標示主題與主題之間的關聯
<member>	標示關聯中的成員
<roleSpec>	關聯中的成員所扮演的角色
<mergeMap>	將一個以上的主題地圖合併在一起

如果仔細加以分析，則可進一步歸納出 XTM 的標籤結構，如圖六所示。





圖六：XTM 標籤結構圖



參、知識組織

所謂知識組織簡單的說是將資訊或知識精鍊的過程。大陸學者蔣永福和李景正提出七種知識組織方法：知識表示、知識重組、知識聚類、知識存檢、知識編輯、知識佈局以及知識監控。(註 25) 其中「知識表示」是指把知識載體中的知識因子和知識關聯表示出來，以便人們識別和理解知識，是知識組織的基礎，因為任何知識組織的方法都必須建立在知識表示的基礎上。而在圖書資訊學的領域中即是使用分類和主題標目來揭示主題內容。吳萬鈞則認為，利用各種科學組織系統 (Scientific Knowledge Organization Systems) 將科學文獻中經過加工處理後的知識組織起來，是儲存、檢索與利用科學知識不可或缺的關鍵環節。(註 26)

Hodge 將知識組織系統分為三大類：(1)術語集 (Term list)：包含權威檔 (Authority files)、術語彙編 (Glossaries)、字典 (Dictionaries)、地名詞典 (Gazetteers) 等；(2)分類系統 (Classifications and categories)：包含標題表 (Subject headings)、分類表 (Classification schemes, taxonomies, and categorization schemes) 等；(3)關連集 (Relationship Groups)：包含索引典 (Thesauri)、語義網路 (Semantic Networks)、知識本體論 (Ontologies) 等。(註 27)Hodge 所提的這三種知識組織方法，在主題地圖中都非常容易實現。例如：要實作術語集，我們將術語集中的每一個項目定義為一個主題 <topic>，再以資源指引 <occurrence> 對該主題進一步描述或說明；若要實作分類系統，則只要將不同類別定義為主題類型，再將個別主題加以歸屬至相對應的主題類型即可；而關連集則只要在 <topic> 和 <occurrence> 之外再加入 <association> 的概念便可完成。

在圖書館學中，書後索引、術語彙編和索引典是三種常見的知識組織法，以下便從主題地圖的角

度切入，概略說明這三種知識組織法及其與主題地圖相關的概念。(註 28)

一、書後索引

曾有人說過，一本書若是沒有索引，就好比一個國家沒有地圖。一個傳統的索引事實上就是一本書的知識地圖(註 29)，羅列了這本書所有的主題，並給予相關的指引，如頁碼、同義詞等。其構成要素包括：區別不同的主題類型及資源指引；採用「見」(See) 來處理同義詞；採用參見 (See also) 來指引到相關主題；在次要項目提供連結到其他類型的主題 (如連結作者與其作品、或上層分類與下層分類間)；區別同音異義、同形異義、同音同形異義詞；資源指引並不僅止於頁碼，可能還包括章節、附註等，並用不同的字體標示等。總結來說，主要特點有：定義主題、主題間的關係、主題資源指引，並藉此將資訊傳遞給使用者。

因此，一本書的索引即同時兼具有主題、關聯與資源指引三要素，與主題地圖的 TAO 模型不謀而合。

二、術語彙編

術語彙編是基本詞彙與定義的列表，如同單一類型的索引，並且只有一種資源指引，也就是該術語的定義。而如同索引，一個術語彙編同時具有「見」與「參見」來連結相關主題，有的還會提供相關資訊，如語言或發音等，但重點在於主題名稱與其定義。因此，術語彙編也非常適合使用主題地圖的模式來實現。

三、索引典

索引典強調的是索引的另一層面，它是特定領域中相互關連詞彙的語意網路，重點在於詞彙間的關連性 (Relationship or association)，選擇一個特定詞彙後，索引典會提供相同意義、同類型事物的



更大類與更小類、以及在其他方面相關的詞彙。更重要的是索引典將詞彙歸類，不僅說明詞彙間具有關連性，更指出如何或為何相關，使得指引更為容易。目前在索引典標準中通用的關連類型有廣義詞、狹義詞、不用、相關詞等。(註 30)

在以下的論述中，我們首先探討與分析索引典的定義、目的、結構與編排，接著闡述本研究如何以主題地圖來建構一個實驗性質的線上索引典。

肆、索引典之探討與分析

一、索引典的定義

索引典的英文為“Thesaurus”，其希臘與拉丁的字源本義為「寶典」(A treasury)，自 1950 年代開始發展，廣泛運用於資訊儲存與檢索系統(註 31)，較常見之定義有：

- (一)索引典是控制索引語言正式組織的詞彙，以便將概念間的重要關係(如廣義和狹義)明確的顯示出來。(註 32)
- (二)就功能而言，索引典是一種控制詞彙的工具，其用途是將文獻、索引人員或使用者的自然語言，轉譯成更規範的「系統語言」(文獻工作語言、資訊語言)；就結構而言，索引典是一部含有特定知識領域的詞彙，詞彙間有語意或從屬上的關係，且詞彙是控制的、動態的。(註 33)
- (三)就資訊儲存與檢索的範疇而言，索引典乃收集足以表示知識概念的字或詞，並將之以特定的結構加以排列，這些字彙控制了同義字，區別了同形異義字，並顯現各相關詞彙間階層及語意互屬上的各種關係，以做為索引者在分析處理資料及讀者在檢索資料時能選用一致的、經過控制的詞彙。(註 34)

二、索引典的目的

依據 ANSI Z39.19，索引典的目的有以下四點：(註 35)

- (一)轉化：提供一個讓作者、編制索引者與使用者將自然語言轉換成編索引及檢索用的控制詞彙。
- (二)一致性：促進索引詞彙指派的一致性。
- (三)指示相關性：指出詞彙間的語意關係。
- (四)檢索：作為檢索時的輔助工具。

D. Soergel 認為索引典的概念結構 (Conceptual structure) 在索引與檢索時能將主題概念表現得恰到好處，而索引典將同義詞聚集亦有詞彙控制 (Terminological control) 的功能。(註 36)

三、索引典的結構

一般而言，索引典的詞彙分為標目 (Heading) 及參照款目 (Cross reference entries) 兩種。通常標目被認可為可以使用的詞彙，稱之為敘述語或述語 (Descriptors)；參照款目則為不可以使用的詞彙，稱為非敘述語 (Non-descriptors) 或被替代語 (Use reference)，亦即圖書館書目資料處理時採用的參見 (See) 作法。不被認可的參照款目利用被替代 (Use) 與認可述語連結起來，以便指引索引典的使用者參考以選用適切的敘述語。每一個敘述語之下列有各種關係詞彙，這些詞彙以各種記號表明其間的關係，如範圍註 (Scope note)、替代 (Used for)、被替代 (Use)、廣義詞 (Broader term)、狹義詞 (Narrower term) 及相關詞 (Related term) 等。詞間關係則包括等同 (Equivalence)、層級 (Hierarchical) 和聯想 (Associative) 三種關係。(註 37)(註 38)

四、索引典的編排

索引典的基本編排方式有字順、分類與圖形三



種：(註 39)

(一)字順：拼音語系依字母序或字序排列所有的主題詞；表意文字如漢語，可用「四角號碼」、「注音符號」、「羅馬拼音」或其他拼音等來排列，若首字同音則可配合筆劃區分排列次序。

主題詞

注釋或定義 (SN)

同義詞 (用：USE；代：UF)

廣義詞 (BT)

狹義詞 (NT)

相關詞 (RT)

(二)分類：除分類表外，還配合引領到分類部分的字順索引。在分類部分，每個描述詞都有一個分類代碼，群詞結構只有主題詞、注釋或定義、同義詞及相關詞，層級關係則由分類法和縮行空隔表示。

(三)圖形：包括圖形與字順兩部份，圖形只有描述詞，其他詞間關係在字順部分表示，一般分為樹形結構及箭頭關係兩種。

伍、以主題地圖建構索引典

本節首先界定本研究所採用的詞彙範圍，再據以建構索引典的語意網路圖，最後以 XTM 語法實作成主題地圖，並透過 Ontopia 的 Omnigator 處理器提供線上索引典之導覽。

一、研究範圍界定

本研究自教育論文線上資料庫 (Educational Documents Online, 簡稱 EdD-Online) (註 40) 的中文教育類詞庫中，以主題詞「專家系統」為中心，選取其相關詞彙作為建置實驗用索引典之素材。EdD-Online 是由國立臺灣師範大學圖書館與各師範校院圖書館共同合作建置，內容包含教育論文線上資料庫、教育新聞剪報資料庫、中文教育類詞庫查詢等，收錄國內教育類論文、研討會論文，並提供文獻傳遞服務、教育專題選粹等功能，是一個非營利性資料庫。而中文教育類詞庫係以 ERIC 資料庫 (Educational Resources Information Center, ERIC) 之索引典為基礎，可註記詞彙的適用範圍 (Scope)、廣義詞 (Broader terms)、狹義詞 (Narrower terms)、相關詞 (Related terms)、替代 (Used for) 與被替代 (Use) 等關係語彙。中文教育類詞庫獨立建置為一線上的全文檢索資料庫，使用者透過詞庫之查檢瀏覽，可直接將單一或多個詞彙轉入教育論文資料庫中檢索，相關詞亦可供一併檢索。(註 41)

為避免語意關係範圍的無止無盡，我們從 EdD-Online 中選出十九個與「專家系統」相關的主題詞，我們對這十九個主題詞及其替代詞彙進行關聯性分析後，歸納出其詞間關係，如表二所示。

表二：本研究所採用的十九個主題詞與其詞間關係

主題詞	替代(UF)	相關詞(RT)	廣義詞(BT)	狹義詞(NT)
專家系統	知識本位系統	電腦、資料處理、知識呈現、人機系統、資訊系統	人工智慧	智慧型教學輔導系統
智慧型教學輔導系統	智慧型電腦輔助教學系統、智慧型電腦輔助教學	電腦管理教學、互動式錄影帶、編序教學輔導、教學輔導方案	電腦輔助教學、專家系統	
人工智慧			電腦、知識呈現	專家系統

(續下表)



主題詞	替代(UF)	相關詞(RT)	廣義詞(BT)	狹義詞(NT)
知識呈現		人工智慧、電腦、專家系統、自然語言處理		
自然語言處理	自然語言理解系統	人工智慧、電腦 知識呈現、人機系統	資料處理	
人機系統	人機對話、人機介面	電腦輔助教學、電腦、專家系統、自然語言處理、線上系統		親使用者介面
電腦管理教學	生涯成熟態度問卷、電腦輔助教學管理、電腦本位教學管理	電腦輔助教學、電腦、智慧型教學輔導系統	資訊系統	
互動式錄影帶	智慧型錄影帶	電腦輔助教學、智慧型教學輔導系統、線上系統	線上系統	
編序教學輔導		電腦輔助教學、智慧型教學輔導系統、編序教材	編序教學、教學輔導	
教學輔導方案	教學輔導計畫、教學輔導服務	智慧型教學輔導系統、教學法、教學輔導		
編序教學	編序學習、編序式自我教學	編序教材	教學法	電腦輔助教學 編序教學輔導
電腦	電腦科技	人工智慧、電腦輔助教學、電腦管理教學、資料處理、專家系統、知識呈現、人機系統、自然語言處理、線上系統、資訊系統		
資料處理	自動化資料處理、資料表格化、電子資料處理	電腦、專家系統、線上系統、資訊系統		自然語言處理
教學法	教學方法、教學實務、教學系統、教學技巧、教學方法	教學輔導方案	教育方法	編序教學
教學輔導	輔導教學	教學輔導方案	個別教學	編序教學輔導
電腦輔助教學	電腦輔助學習、電腦本位教學	電腦管理教學、電腦、互動式錄影帶、人機系統、編序教材、編序教學輔導	電腦在教育上的應用、編序教學	智慧型教學輔導系統
資訊系統		電腦、資料處理、專家系統、線上系統		電腦管理教學
線上系統	互動式系統(線上)	電腦、資料處理、資訊系統、人機系統、資訊系統		互動式錄影帶
編序教材	自我教學教材	電腦輔助教學、編序教學、教材 編序教學輔導		

取材自：教育論文線上資料庫中文教育類詞庫 (EdD-Online <<http://140.122.127.251/edd/edd.htm>>)



二、建構語意網路

確定所採用的詞彙及其關聯性後，便可開始著手建構其語意網路。語意網路是 AI 研究領域中的一種表現形式，包含節點 (Nodes) 與連結 (Links)，節點通常代表物件、概念或某特殊領域中的情境，

連結則代表節點間的語意關連。(註 42)圖七呈現了一個語意網路的簡單概念，其中有兩個節點分別表示「廣義詞」與「主題詞」，一個連結則代表「是一個」(is-a link)，表示節點間的關係。因此，其所代表的語意為「廣義詞是一個主題詞」。



圖七：以語意網路描述之簡單事實

若「專家系統」是一個特殊的個體，被我們認定為是某個詞彙的廣義詞，則可以增加一個專家系

統的節點到這個語意網路中，如圖八所示。



圖八：事實的推論(轉換)

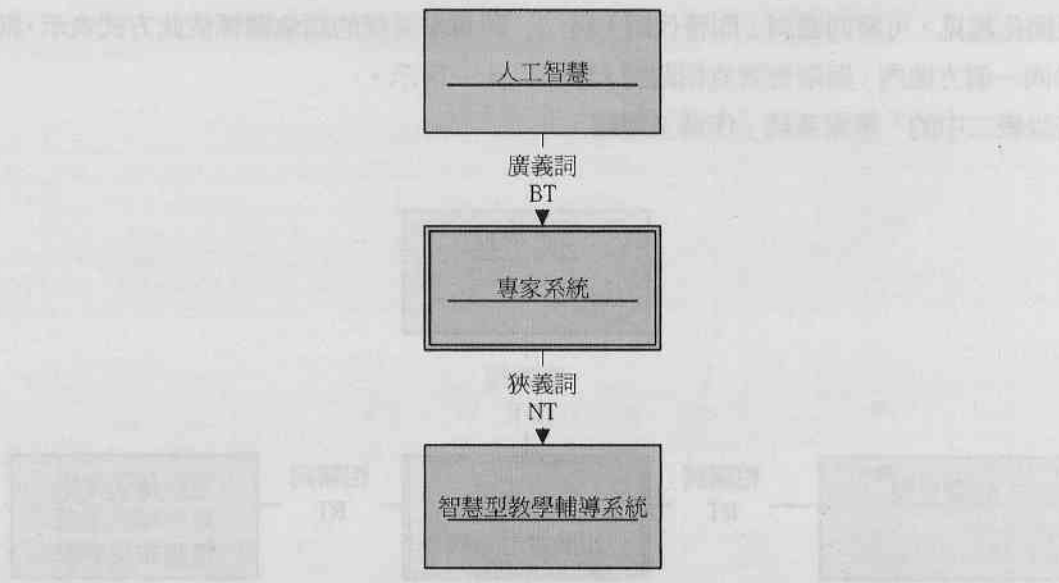
要注意的是，圖八並不只呈現兩個最開始的論據(「專家系統是一個廣義詞」與「廣義詞是一個主題詞」)，經由簡單的連結，還推論出第三個事實「專家系統是一個主題詞，這種基於語意關連而推論出新論據的能力，稱為「可遞性」"Transitivity"，可遞性允許利用新連結建立來產生新關係。然而，可遞性通常都是單向，所以我們可以說「專家系統是一個廣義詞」，但不能說「廣義詞是一個專家系統」，儘管如此，還是可以建立另一種雙向連結，如「這個主題詞可能是一個廣義詞」、「這個廣義詞可能是專家系統」。測試可遞性存在的最好方式是證明以下說明是真實的：「所有主題 A 的例子都與主題 B 有一個專指關係 (Specific relationship)」，當所有的廣義詞都是主題詞，則廣義詞是一個主題詞，然而，主題詞的所

有例子並非都是廣義詞，所以主題詞是一個廣義詞是不正確的。

反身關係 (Reflexive relationships) 發生於節點間的關係屬於「關聯」時，如某詞彙與某詞彙有關，則可以繪成雙向關係。對稱關係則發生於節點的位置不會影響結果說明的真實性，如某兩個詞彙間具有層級關係，「人工智慧是專家系統的廣義詞」、「專家系統是人工智慧的狹義詞」，前面提到的「關聯」關係也是屬於反身關係的一種。

語意網路使得繼承層級更為容易，隨著層級結構，部分關係可被繼承，如圖九即可由「人工智慧是專家系統的廣義詞」與「專家系統是智慧型教學輔助系統的廣義詞」，推論出「人工智慧是智慧型教學輔助系統的廣義詞」。

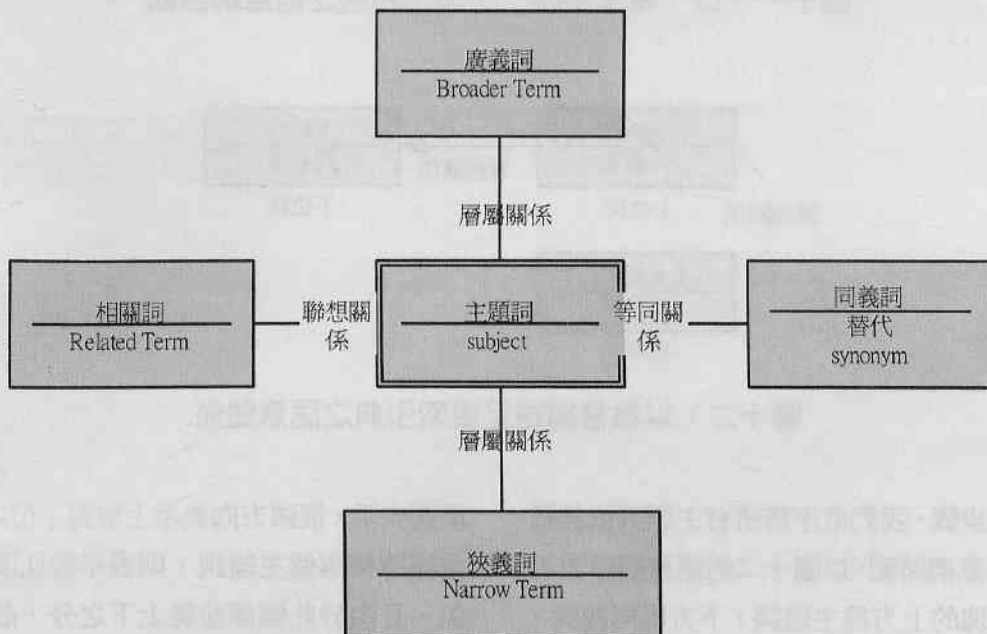




圖九：索引典之層級繼承關係

基於以上的分析，我們可依據索引典詞彙關係，提出如下的語意網路模型：以「主題詞」為中心，其等同關係為「同義詞」；層級關係為「廣義

詞」、「狹義詞」；聯想關係則為「相關詞」。此模型如圖十所示。

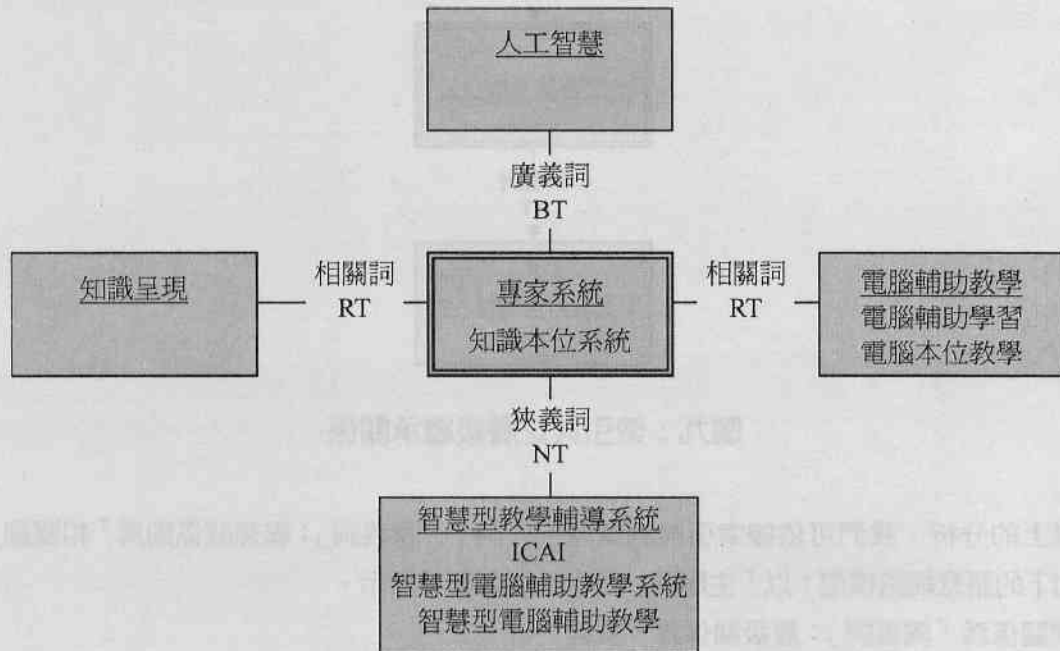


圖十：索引典詞彙關係之語意網路模型

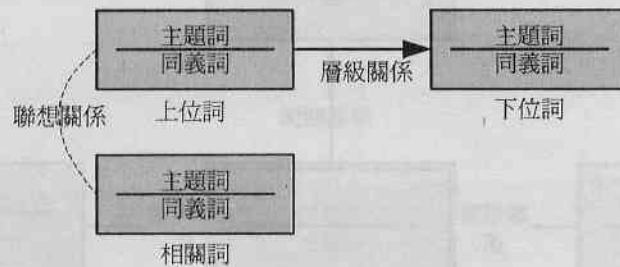


爲了更簡化起見，可將同義詞（即替代詞）與主題詞置於同一個方塊內，而兩旁置放相關詞。舉例而言，若以表二中的「專家系統」作爲主題詞，

將專家系統的語彙關係依此方式表示，則結果如圖十一所示。



圖十一：以「專家系統」主題詞為例之語意網路圖

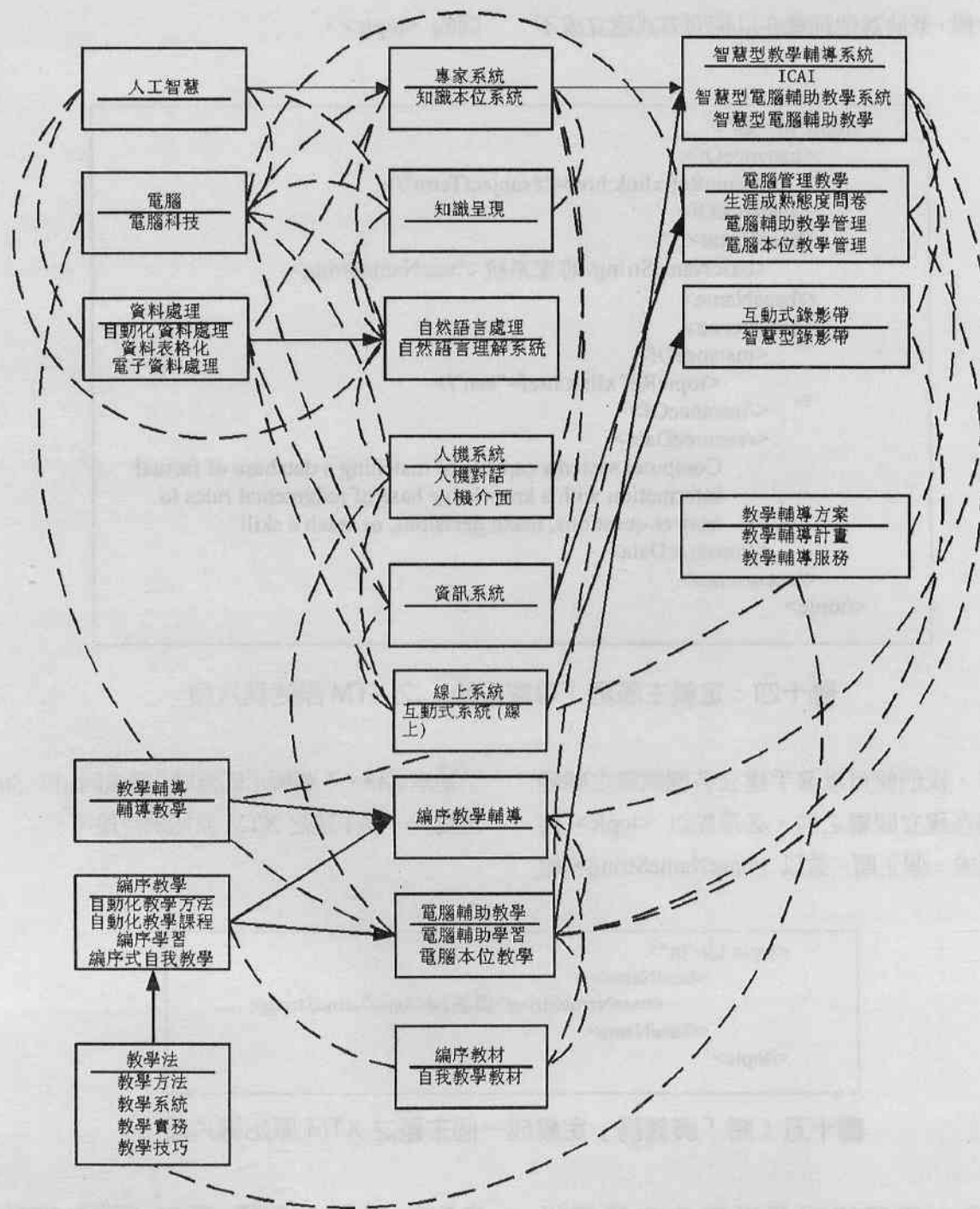


圖十二：以語意網路呈現索引典之語意關係

根據上述步驟，我們依序將所有主題詞依其關聯性建構成語意網路圖，以圖十二的語意網路表示方式，每個方塊的上方爲主題詞，下方爲同義詞，表示同一概念。若主題詞間有層級關係，則以實線

箭頭表示，箭頭方向表示上位對下位之關係。若以虛線連接兩個主題詞，則表示彼此間具有聯想關係，且由於此關係並無上下之分，故並無箭頭方向。最後結果如圖十三所示。





圖十三：以 Ed-D 索引典中的十九個主題詞彙之語意網路圖

三、實作主題地圖

完成以上十九個詞彙之語意網路後，即可遵循 XTM 標準，建構其主題地圖。首先，將各個詞

彙以 <topic> 元素加以標示，並將 <topic> 以 <instanceOf> 加以分門別類，再加上其範圍註 (SN) 作為其資源指引，並以 <occurrence> 加以標示。如下圖所示即為主題詞「專家系統」的 XTM



原始碼片段，至於其他詞彙亦以相同方式建立成不同的 <topic>。

```

<topic id="es">
  <instanceOf>
    <topicRef xlink:href="#subjectTerm"/>
  </instanceOf>
  <baseName>
    <baseNameString>專家系統</baseNameString>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#sn"/>
    </instanceOf>
    <resourceData>
      Computer systems capable of matching a database of factual
      information with a knowledge base of judgmental rules to
      answer questions, make decisions, or teach a skill.
    </resourceData>
  </occurrence>
</topic>

```

圖十四：定義主題詞「專家系統」之 XTM 原始碼片段

接著，我們便可以著手建立各個詞彙之關聯了。不過在建立關聯之前，必須先以 <topic> 將關聯定義成一個主題，並以 <baseNameString> 給

予基本名稱。下圖所示即為將「廣義詞」以 <topic> 定義成一個主題之 XTM 原始碼片段。

```

<topic id="bt">
  <baseName>
    <baseNameString>廣義詞</baseNameString>
  </baseName>
</topic>

```

圖十五：將「廣義詞」定義成一個主題之 XTM 原始碼片段

再來，便可將兩個不同的主題詞以 <association> 建立關聯了。例如，下圖是建立「專

家系統」與「人工智慧」兩主題關聯的 XTM 原始碼片段。




```

<association id="ha1">
  <instanceOf>
    <topicRef xlink:href="#hierarchical"/>
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#bt"/>
    </roleSpec>
    <topicRef xlink:href="#ai"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#nt"/>
    </roleSpec>
    <topicRef xlink:href="#es"/>
  </member>
</association>

```

圖十六：建立「專家系統」與「人工智慧」關聯之 XTM 原始碼片段

依據上述步驟，一一將所有的主題詞以此方式進行定義，並將之關聯起來，便完成了一個以主題詞「專家系統」為中心的主題地圖了。

最後，我們在 Web 伺服器上安裝由 Ontopia 公司開發的主題地圖處理器 Omnigator，再將完整的 XTM 文檔發佈至該伺服器上，即完成此線上索引典的製作了。使用者只要用一般 Web 瀏覽器即可使用此線上索引典，圖十七呈現的是其概觀，分別顯示了主題詞彙、主題詞間的關聯性、在關聯中所扮演的角色、以及資源指引的類別 ... 等。

若以「專家系統」一詞為例（參見圖十八），可以很清楚的畫面左方看到與其相關之主題及彼此間的關聯性，而右方則是其資源指引的內容。

陸、主題地圖的應用議題與未來發展

一、應用議題

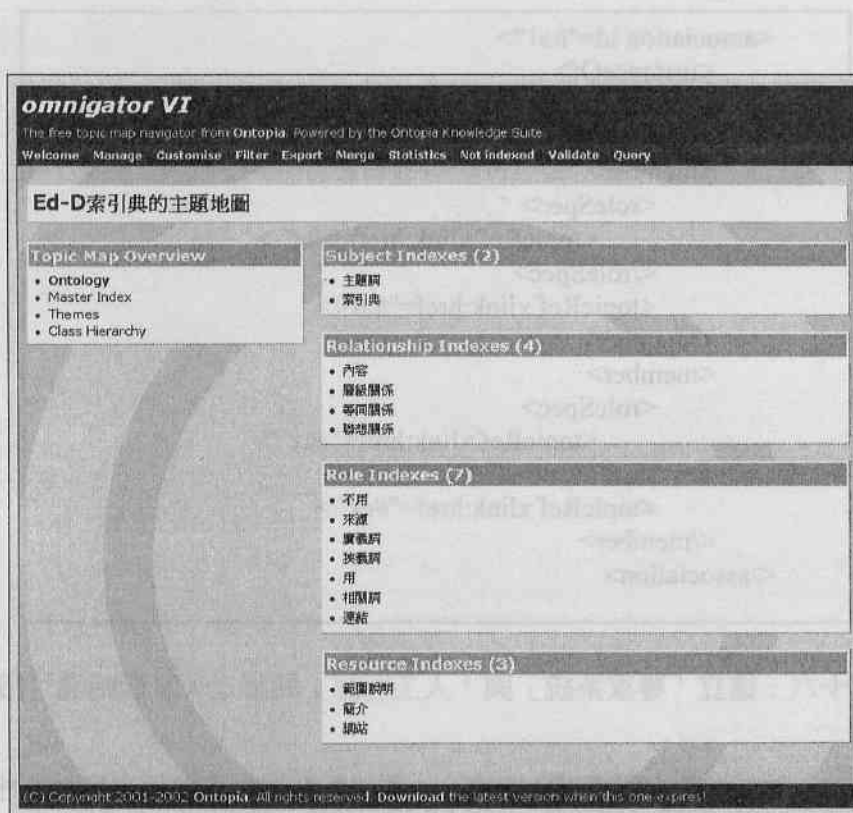
雖然主題地圖的發展可追溯到相當長遠的一段時間，並發展的如火如荼，然而在應用層面上，仍有一些問題亟待解決，M. Biezunski、S. Newcomb

和 M. Bryan 即提到主題地圖相關標準訂定之問題。(註 43)首先是目前的 ISO 13250 定義 XTM 與 HyTM 兩種交換語法，但並未解釋這兩者該如何關聯，當負責開發的人採用不同的方法時，就會造成將來兩者間的交換問題。

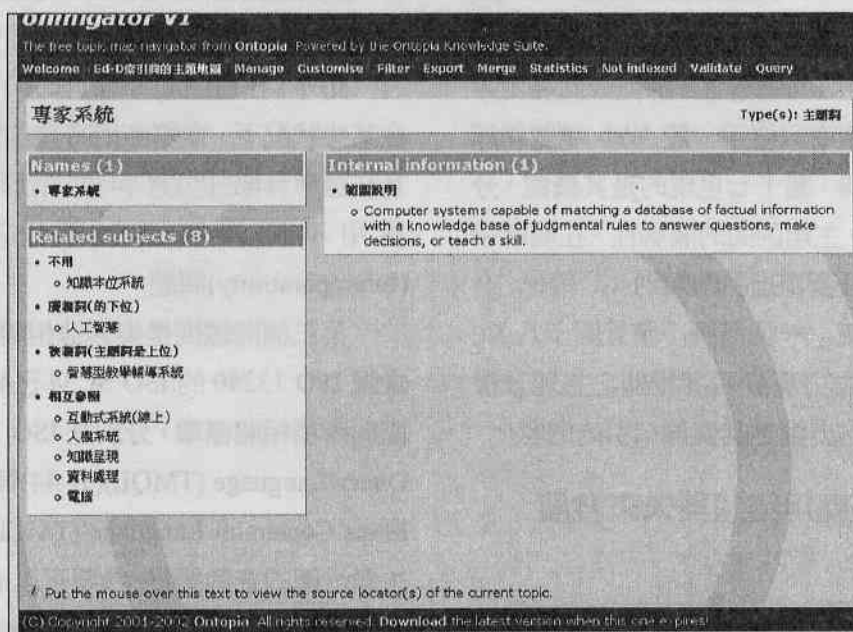
此外，在目前的 ISO 13250 中並沒有具體說明在某些狀況下，實際應用時該如何處理這些部分，且在標準有部分語意不明的情形下，開發人員很容易用不同的解讀而導致另一個互通性 (Interoperability) 問題。

第三個問題則是與其他相關標準的支援，負責維護 ISO 13250 的 ISO SC34 已經制訂了與主題地圖的兩項相關標準，分別是 ISO 18048: Topic Maps Query Language (TMQL)(註 44)與 ISO 19756: Topic Maps Constraint Language (TMCL)。(註 45)前者是主題地圖的查詢語言，希望可以成為主題地圖資料查詢語言(SQL)或是 XML 查詢語言(XML Query)的一種，未來將使主題地圖的應用軟體能夠以更精簡的方式，來擷取主題地圖中的資訊，目前發展至





圖十七：以 Ed-D 索引典中的十九個主題詞彙之主題地圖



圖十八：「專家系統」一詞之詞彙關係及資源指引



1.0 版。TMCL 則是主題地圖的一種概要或限制 (Schema or constraint) 語言，一種主題地圖中物件類別的規則，由關聯類型、主題類型、資源指引類型、關聯角色類型、題材(Theme)、範圍等來定義。

但這兩者都需要 ISO 13250 提供合適的基礎，比方說，在 TMQL 要定義一個問句如 “find all base names of topic X in scope Y”，操作人員必須小心且清楚說明該如何處理，在 XTM 語法中是有定義到這一塊，但在 HyTM 中卻付之闕如，且 XTM 有時也提供相當多的方法來表示同一件事，這些都會造成主題地圖在應用上的問題。因此，從 2001 年 5 月開始，ISO SC34 為了解決相關問題，開始訂定標準應用模型(Standard Application Model, 以下簡稱 SAM)標準，TMQL 與 TMCL 將等到 SAM 完成後，依據該項產出結果再繼續進行。

二、未來發展

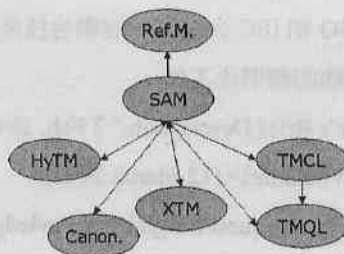
因而 ISO SC34 預計將整個 ISO 13250 發展成多部分，以解決目前現有問題，SAM 即是其中的關鍵所在，SAM 將奠基於 XML Information Set(註 46)，將定義主題地圖的架構，同時也是合併(merge)與副本移動的關鍵操作點，SAM 亦處理了 HyTM 與 XTM 的通透性問題，以及上述提及，TMQL 與 TMCL 的基礎。

而將來的主題地圖標準也會基於 SAM 模型，

來重新處理 HyTM 和 XTM 在主題地圖之關聯，新版的語法規範將會描述如何自一份文件建立 SAM 的實例，但並不會改變語法本身，也就是說，將會有如下的描述：「為了在文件中的每一個 <topic>，建立一個主題項目」、「為了 <topic> 的每一個 <baseName> 子欄位，建立一個基本名稱項目，並加入與其相對應之主題項目的 [base names] 屬性」，等諸如此類。

未來的 ISO 13250 也會加入參考模型 (Reference model)，一個在主題地圖中更為抽象的圖形模式，名稱(Name)與指引資源指引(Occurrence resources)將會成爲一個與主題相同層次的節點，這將使參考模型比 SAM 使用更少的限制 (Constructs)，並在改變時不致影響到超模型 (Metamodel)。

參考模型提供一個架構，爲了闡述不同知識表示法間的關聯，如主題地圖、RDF(註 47)、KIF(註 48)，將使主題地圖與其他知識表示法間的溝通更爲容易。預計在 SAM 部分將定義 SAM 對映到參考模型的規範，TMQL 與 TMCL 也會經由 SAM 與參考模型有所關聯，很明顯的，SAM 與參考模型間的一致性是非常重要的，許多部分都會依此進行。可從圖十九看出未來 ISO 13250 中不同部分之關聯性，例如 TMQL 與 TMCL。



圖十九：未來 ISO 13250 各部分關聯性之概念圖

資料來源：M. Biezunski, S. Newcomb, and M. Bryan, “ISO/IEC JTC 1/SC34 N323 Guide to the topic map standards,” 23 Jun 2002, <<http://www.y12.doe.gov/sgml/sc34/document/0323.htm>>.



新版的 ISO 13250 預計會包括幾個部分：此標準之架構導言、SAM、參考模型、XML Topic Maps syntax (XTM)、HyTime Topic Maps syntax (HyTM) 以及主題地圖的權威語法，目前尚未估計整個新標準的完成時間。

柒、結論

在本研究中，我們成功的提出一個以主題地圖建構索引典之語意網路模型、方法與步驟，並從實際詞庫中選擇十九個主題詞及其替代詞彙作為實驗用素材，歸納其詞間關係並建構語意網路圖，再以 ISO 13250 之 XTM 主題地圖標準語法實作，最後在伺服器端建置主題地圖處理器加以處理，順利

的完成一個實驗性質的線上索引典。

展望未來，主題地圖是一項結合了知識工程和知識組織的新技術，更是一種分散式的知識表式法，若能有效的加以運用可以整合眾多分散於網路上的知識庫，從而建構出一個強大的知識管理系統。不過，主題地圖亦將如同其他新興的標準或技術一樣，一方面在各個領域展露潛力，另一方面則將遭遇各種問題的嚴厲挑戰，是否真能成為知識組織的利器，或是成為知識管理的有效解決方案，實在是一個頗值得探討的議題，亦是我們未來的研究重點之一。

(收稿日期：2003 年 11 月 3 日)

註釋：

- 註 1：本研究部分成果曾刊載於「2003 年資訊科技與圖書館學術研討會，2003/5/29，淡江大學」，本文內容業經作者修訂並加強研究成果，以分享圖書館界同道。
- 註 2：Giarratano and Riley, "Expert Systems: Principle and Programming," PWS-KENT (1989), pp. 63-102.
- 註 3：Pepper, S. (n.d.), "The TAO of topic maps: Finding the way in the age of Infoglut," XML Europe 2000, 12-16 June 2000, <<http://www.gca.org/papers/xml europe2000/papers/s11-01.html>> (22 March 2003).
- 註 4：Omnigator 是由 Ontopia 公司所開發出來的軟體，為一伺服器端的主題地圖處理器，全稱為 omnivorous topic map navigator，以結構化方式顯示主題地圖，可於以下網址下載試用版 <<http://www.ontopia.net/download/index.html>> (18 March 2003).
- 註 5：此圖主要參考 Michel Biezunski, "Introduction to the Topic Maps Paradigm," in XML Topic Maps, ed. Jack Park, (Boston: Addison-Wesley, 2003), p.24.，再經本文作者參酌其他相關資料修訂而成。
- 註 6：為了制訂資訊科技的全球標準，ISO 和 IEC 合組了一個聯合技術委員會 (Joint Technical Committee, 稱為 ISO/IEC JTC 1)，負責資訊科技範疇的標準化工作。
- 註 7：Erik Naggum, "SGML: Erik Naggum's Brief Description," 7 Feb. 1995, <<http://xml.coverpages.org/naggumWhat.html>> (12 March 2003).
- 註 8：Steven R. Newcomb, "A perspective on the quest for global knowledge interchange," in XML Topic Maps, ed. Jack Park (Boston: Addison-Wesley, 2003), p. 38.
- 註 9：GCA 於 1966 年成立，後改名為 IDEAlliance (International Digital Enterprise Alliance) 是一個非營利性組織，致力於各項標準的發展，如 GRACoL、ICE、JIFFI、Mail.dat、papiNet、PRISM、PROSE XML、SPACE XML、SGML 與 XML 等，目前擁有超過 300 個公司會員與上千個個人會員。官方網站為



<<http://www.idealliance.org/index.asp>>。

- 註 10：Steve Pepper and Lars Marius Garshol, "The XML Papers: Lessons on Applying Topic Maps," Oct. 2002, <<http://www.ontopia.net/topicmaps/materials/xmlconf.html>> (5 Mar. 2003).
- 註 11：同註 10。
- 註 12：Topicmaps.Org 是一個獨立性組織，主要負責發展主題地圖的 XML 版本，即 XTM 規範，IDEAlliance 是其
主要贊助者，官方網站為<<http://www.topicmaps.org/>>。
- 註 13：TopicMap.Org, "XML Topic Maps (XTM) 1.0 - TopicMaps.Org Specification," 8 Aug. 2001, <<http://www.topicmaps.org/xtm/1.0/index.html>> (5 Jan. 2004).
- 註 14：W3C, "Extensible Markup Language (XML)," <<http://www.w3.org/XML/>> (20 Feb. 2003).
- 註 15：同註 8，頁 37-40。
- 註 16：M. Biezunski, S. Newcomb, and M. Bryan, "ISO/IEC JTC 1/SC34 N323 Guide to the topic map standards," 23 Jun. 2002, <<http://www.y12.doe.gov/sgml/sc34/document/0323.htm>> (17 Mar. 2003).
- 註 17：同註 3。
- 註 18：目前 Occurrence 一詞的中文譯名並不一致，有：資源指引、資源所在、參考處、出處、關係實例等，本文
採用「資源指引」一詞，取其「指引該主題至相關資訊資源所在處」之意。將來華文世界若有更適切之譯
名則從其用語。
- 註 19：同註 3。
- 註 20：Michel Biezunski, "Introduction to the Topic Maps Paradigm," in XML Topic Maps, ed. Jack Park (Boston:
Addison-Wesley, 2003), p.24.
- 註 21：ISO 10744:HyTime 中的一項功能，主要對於超文件連結與多媒體時間同步性的應用。可參考
<<http://www.ornl.gov/sgml/wg8/docs/n1920/html/n1920.html>> (22 March 2003)。
- 註 22：XPoint，全名為 XML Pointer Language，是 Xlink 的一部份，主要支援 XML 文件內的資源定址，官方網站
URL 為：<<http://www.w3.org/XML/Linking>> (22 Mar. 2003)。
- 註 23：Alschuler, Liora, "Topic Maps," 16 June 2000, <<http://www.xml.com/pub/a/2000/06/xmlleurope/maps.html>> (22
Dec. 2002).
- 註 24：同註 13。
- 註 25：蔣永福、和李景正，「論組織組織方法」，中工圖書館學報 1 期（2001 年），頁 3。
- 註 26：吳萬鈞，「科學知識組織系統」，資訊傳播與圖書館學 5 卷 1 期（民國 87 年 9 月），頁 19-20。
- 註 27：Gail Hodge, "Knowledge Organization Systems: An Overview," in System of Knowledge Organization for Digital
Libraries: Beyond Tradition Authority Files, April 2000, <<http://www.clir.org/pubs/reports/pub91/contents.html>>
(19 Feb. 2003).
- 註 28：同註 3。
- 註 29：同註 3。
- 註 30：索引典標準詳見 Z39.19:1993 [Guidelines for the Construction, Format, and Management of Monolingual
Thesauri]、ISO 5964:1985[Documentation - Guidelines for the establishment and development of multilingual



thesauri] 與 ISO 2788:1986 [Documentation - Guidelines for the establishment and development of monolingual thesauri]。

註 31：黃惠株，「淺談索引典」，佛教圖書館館訊 5 期(民國 85 年 3 月)，頁 2。

註 32：“ISO 2788:1986 Guidelines for the establishment and development of monolingual thesauri,”

<<http://www.nlc-bnc.ca/iso/tc46sc9/standard/2788e.htm#3>> (26 Dec. 2002).

註 33：美國資訊科學學會台北分會、農業科學資訊服務中心、國立中央圖書館主辦，索引典理論與實務（台北市：中國圖書館學會，民國 83 年），頁 5。

註 34：蔡明月，線上資訊檢索：理論與應用（台北市：台灣學生，民國 80 年），頁 177。

註 35：“ANSI Z39.19-1993(R1998) Guidelines for the Construction, Format, and Management of Monolingual Thesauri”，p.1, <<http://www.niso.org/standards/resources/Z39-19.pdf>>(27 Dec. 2002).

註 36：同註 31，頁 2。

註 37：同註 31，頁 4-5。

註 38：同註 33，頁 44-53。

註 39：同註 31，頁 5。

註 40：EdD-Online 網址為 <<http://140.122.127.251/edd/edd.htm>> (26 Dec. 2002)。

註 41：教育論文摘要檢索系統(教育論文線上資料庫)：擴建計畫執行報告，國立臺灣師範大學圖書館，<<http://www.lib.ntnu.edu.tw/Education/eplars-per.html>> (26 Dec. 2002)。

註 42：Eric Freese, “Using Topic Maps: for the representation, management & discovery of knowledge,” XML Europe 2000, 12-16 June 2000, <<http://www.gca.org/papers/xml europe2000/papers/s22-01.html>> (27 Dec. 2002).

註 43：同註 16。

註 44：Hans Holger Rath and Lars Marius Garshol, “TMQL requirements (1.0.0),” <<http://www.y12.doe.gov/sgml/sc34/document/0249.htm>>(22 May 2003).

註 45：Steve Pepper, ‘Draft requirements, examples, and a “low bar” proposal for Topic Map Constraint Language,’ <<http://www.y12.doe.gov/sgml/sc34/document/0226.htm>> (22 Mar. 2003).

註 46：由 W3C 於 2001 年訂定的建議規範之一，主要提供一個定義集，使其他標準在提及 XML 文件中的資訊時能有所依據，官方網站為 URL: <<http://www.w3.org/TR/xml-infoset/>>。

註 47：RDF，全名為 Resource Description Framework，是 W3C 於 1999 年所訂定的建議規範之一，主要用於製作交換以及自動處理網際網路資源之描述，定義資源描述語法與模式，官方網站為 URL: <<http://www.w3.org/RDF/>>。

註 48：KIF，全名為 Knowledge Interchange Format，是一種計算機導向語言，主要作為不同程式間的知識交換，目前由史丹佛大學電腦科學系的知識系統實驗室所負責維護，官方網站為 URL: <<http://www.ksl.stanford.edu/knowledge-sharing/kif/>>。

