

知識觀和知識組織的幾個觀察和思考

吳美美

meiwu@ntnu.edu.tw

臺灣師範大學圖書資訊學研究所

大綱

基本概念

- 認識決定知識的寬度 -- 從馮道說起
- 實體物件相對於虛擬物件的知識組織
- 管理相對於獲取的知識組織

分類理論

- Sparck Jones的八個分類類型
- 知識分類、文獻分類、網路分類
- 分類和知識本體

從案例中 學習

1. 美國國會圖書館
2. 中央研究院詞彙表
3. 口述歷史

問題討論

從馮道說起 – 認識決定知識的寬度

- 馮道的啟發 -- 馮道是誰？(他在中國文化史，卻不在國人的圖書資訊學知識體系中) – 思想史和學術史
- 華盛頓大學圖書館在規劃設計和建造的過程中，匠心獨運、知識淵博的選了六位對人類知識出版和印刷卓有貢獻者，並依時代排序，馮道之後，立即接著古騰堡，這位十五世紀德國的活字印刷創始人就排在這一位十世紀的中國官員之後。
- 最早建築是在1926年完工，該館建築頗負盛名，最著名的大約就是外牆支柱有由教授推薦的24位全球著名的哲學家、藝術家、思想家等，以及高處另有三座代表思想、啟發和專精等三座雕像。但是有趣的是文獻上很少看到對這六道門，以及對門上銅版鏤刻的描述，在舊金山的Tim McCormick注意到了！。



“Strangely, I can’t find mention of the inscriptions which are easily the most visible to any person actually entering the library: the brasswork directly over the six doors, representing six famous figures in printing and publishing: **Tao Feng, Gutenberg, Caxton, Aldus, Estienne, and Elsevier.**”

實體物件 v.s. 於數位物件的知識組織

	實體物件	數位物件
特質	<ul style="list-style-type: none"> • 盤點 • 排序 • 有實體存放位置、對號入座 	<ul style="list-style-type: none"> • 原生或數位化 • 無法盤點 因為太大量 • 不須事先排序 • 獨特的辨識號 • 在不同系統間交換 • 複製容易 • 無單一實體存放位置
方法	<ul style="list-style-type: none"> • 編目規則 • 圖書分類表 • 控制詞彙、索引典等 	<ul style="list-style-type: none"> • 結構 -- 數位文件編碼標準和工具 Dublin Core, TEI... • 內容 -- RDA • 語意 -- 詞彙表 分類表 知識表徵 LCSH, AAT... • 管理應用和交換標準 • 知識本體和鏈結資料

管理 v.s 獲取觀點的知識組織

	管理的觀點	獲取的觀點
概念	<ul style="list-style-type: none">• 以盤點為主• 事先訂定的分類表• 邏輯相關 (logical relevance)• 分類是依照知識前組合• 以知識集結為主要方法	<ul style="list-style-type: none">• 以獲取為主• 依當時的資訊需求檢索• 需求相關 (pertinence relevance)• 檢索是依照後組合• 以知識分析為主要方法
方法	圖書十進分類法、國會分類法	索引法、主題詞表、多分類號法

Sparck Jones的分類理論

- Sparck Jones(1970, 2005) 從資訊檢索觀點提出知識分類的思考，她分析分類的三個要素：類別、物件和屬性。

- 類別和類別的關係有次序性或是沒有次序性？
- 物件和類別的關係是互斥或是可重複入類別？
- 屬性和類別的關係是單一屬性或是多重屬性？

並依此思考，得到八種分類的類型，如右圖

	1		2		3	
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
I	X		X		X	
II	X		X			X
III	X			X	X	
IV	X			X		X
V		X	X		X	
VI		X	X			X
VII		X		X	X	
VIII		X		X		X

Sparck Jones 和8種分類的類型

分類理論

類別 (class)

物件 (object/entity)

屬性 (attributes)

	類別(次序)	物件(互斥)	屬 性	舉 例
	有 無	是 否	單一 多重	
I	x	x	x	學術分類表(Taxonomies) 十進分類法
II	x	x	x	附分法
III	x	x	x	多分類號法
IV	x	x	x	附分法+多分類號法
V	x	x	x	國會分類法、四部法
VI	x	x	x	多面向分類法、附分法
VII	x	x	x	國會分類法+多分類號法
VIII	x	x	x	鏈結資料法

中學生物課所學到的第一個
分類法：界門綱目科屬種

大學在圖書館學系學習杜威十
進分類法、美國國會分類法、
主題詞表等

在美國的研究所讀書時，學知
識組織、分類，老師一定出一
個練習題：將波赫士著名的中
國某個百科全書中的14類動
物重新分組...

傅柯《詞與物：人文科學考古學》(*Order of Things*) 該書前言就引用波赫士的一段話，關於「中國某部百科全書」的分類體系：

「1、屬皇帝所有，2、有芬芳的香味，3、馴順的，4、乳豬，5、鰻鯪，6、傳說中的，7、自由走動的狗，8、包括在目前分類中的，9、發瘋似地煩躁不安的，10、屬不清的，11、渾身有十分精緻的駱駝毛刷的毛，12、等等，13、剛剛打破水罐的，14、遠看像蒼蠅的。」

- 波赫士(Jorge Luis Borges) 阿根廷詩人、阿根廷國家圖書館館長 著有 *The Analytical Language of John Wilkins*
- Wilkins 1614--1672, principal of both Oxford and Cambridge, one of the founders of Royal Society of London.
- 傅柯《詞與物：人文科學考古學》(The order of things, 1966)前言

波赫士在一篇討論17世紀數學家、語言學家 John Wilkins，寫了一篇〈約翰 威爾金斯的分析語言〉時借用、聲稱他聽 Franz Kuhn(1884-1961，德國漢學家，翻譯過紅樓夢和金瓶梅) 說有一本中國的百科全書(應該是類書，虛構的) ‘*Celestial Empire of benevolent Knowledge*’ (翟本瑞翻譯成《天朝仁學廣覽》) 將動物分為14類：

- (a) belonging to the emperor, 屬於皇帝所有的
- (b) embalmed, 塗香料防腐處理的
- (c) tame, 馴養的
- (d) sucking pigs, 乳豬
- (e) sirens, 唱歌的海妖
- (f) fabulous, 傳說中的
- (g) stray dogs, 迷途離群的狗兒
- (h) included in the present classification, 包括於本分類法中
- (i) frenzied, 騷動如瘋子的
- (j) innumerable, 不可勝數的
- (k) drawn with a very fine camelhair brush, 用駱毛細筆畫出來的
- (l) et cetera, 等等
- (m) having just broken the water pitcher, 剛打破水罐子的
- (n) that from a long way off look like flies. 遠看像蒼蠅的

知識分類 文獻分類 網路分類

- **邏輯知識分類法(taxonomy)**
 - 又可稱為邏輯知識組織
 - 類別涵蓋完整、依次序排列、入甲類即不入乙類
- 例如生物學分類為界門綱目科屬種等七大層級

文獻分類法

- 又可稱為文獻知識組織
- 在實體物件時代，分類號用於組成一組獨特號碼以指引所在位置，因此無法指定多重分類號
- 在數位物件時代，只需一組獨特代號指引所在位置，可以採用多種知識表徵，協助檢索

語意網路知識組織工具

- 以邏輯知識分類法、文獻分類法等為基礎，建立節點和關聯的知識表徵

- 例如杜威十進分類法、美國國會分類法、多面向分類法、標題表法等

- 例如知識本體(ontologies)、主題地圖(topic maps)、鏈結資料(linked data)等。

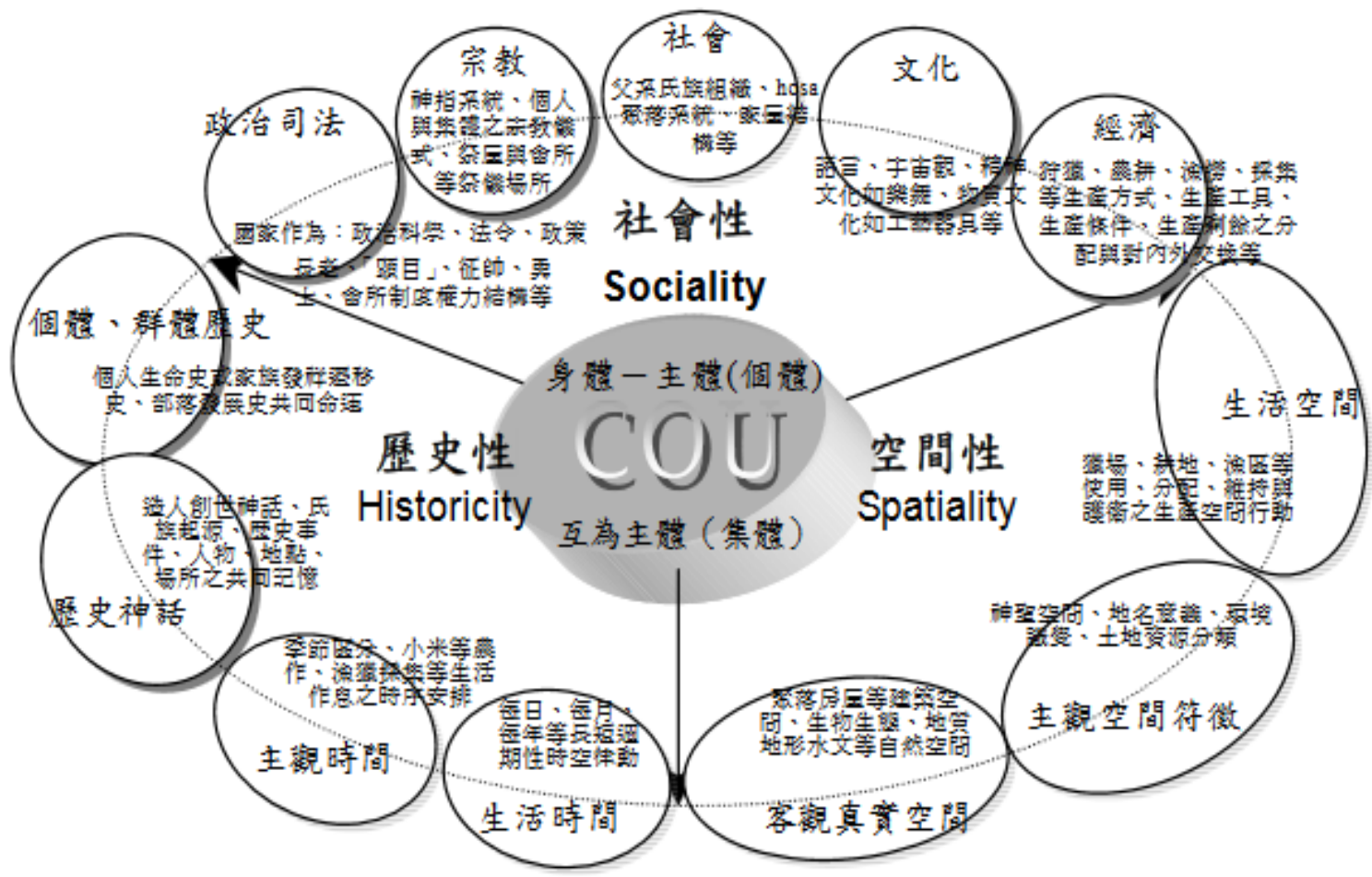


圖 6：鄒族存有及知識本體架構

汪明輝(2009)提出鄒族存有及知識本體架構，分為社會、歷史(時間)、空間三類。

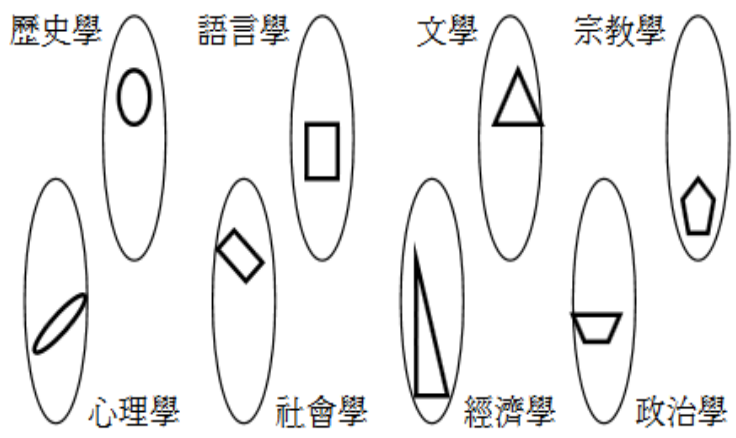


圖 9：孤島式的客家研究

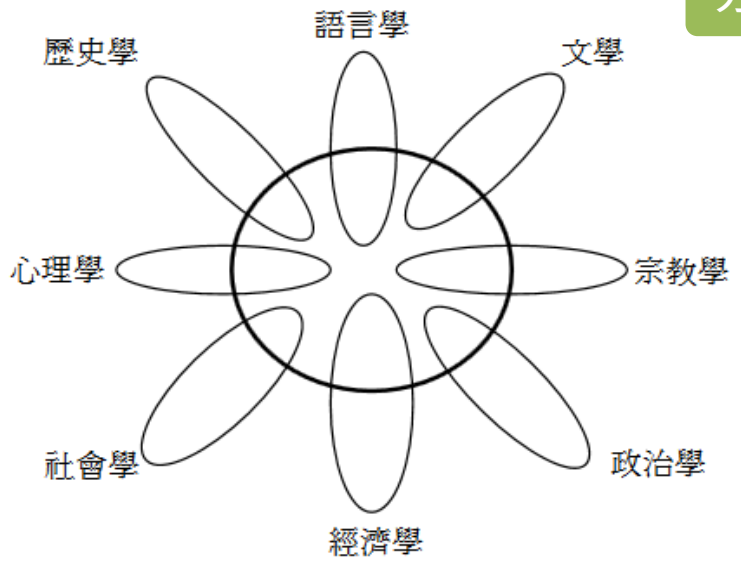


圖 10：跨領域的客家研究

- 施正鋒(2013)期許要將「客家研究」(Hakka Studies) 提升為「客家學」(Hakkaology)，提出三種研究的知識觀。

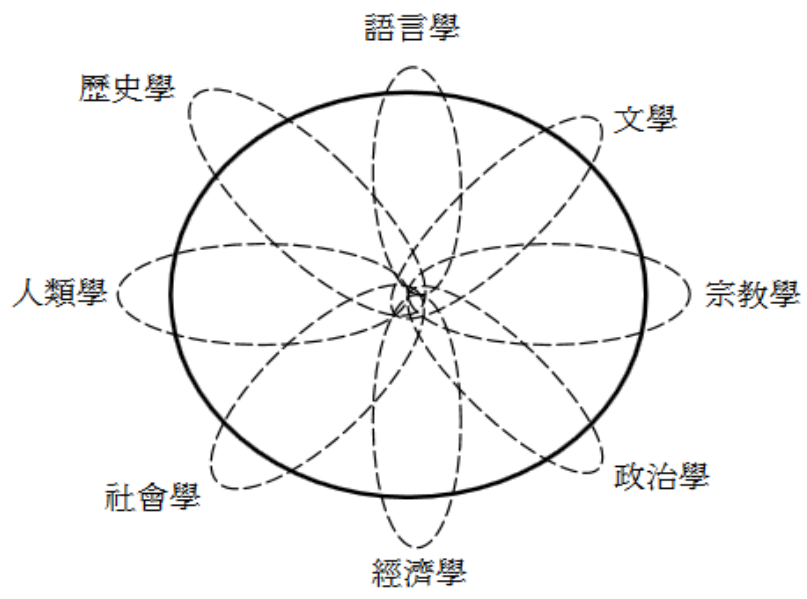


圖 11：超領域的族群研究

分類法和知識本體

Table 1. Comparison between classification schemes and ontologies

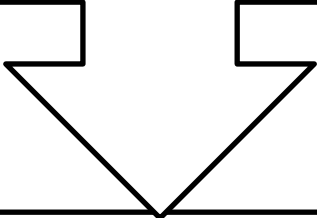
Category	Classification Schemes	Ontologies
Users	Humans	Machines
Purpose	Organization of (large) document collections	Modeling of a domain
Language	Natural language, e.g. English	Formal language, e.g. OWL
Nodes	Usually represent complex concepts or individuals	Usually represent atomic concepts
Edges	Do not have well defined semantics	Have well defined semantics
Instances	Are not necessarily instances of the class in which they are populated	Are instances of the class in which they are populated
Examples	DDC, LCC, Colon classification	Gene ontology ^a , OpenCyc ontology ^b , MeSH ontology

^a <http://www.geneontology.org/>

^b <http://www.opencyc.org/>

Noy, Natalya F. and McGuinness, Deborah L. (2001).
Ontology Development 101: A Guide to Creating Your
First Ontology

知識本體是研究者為了共享一個領域中的資訊而建立的共同語言，包括概念和期間的關係。



建立知識本體的目的

幫助研究者或軟體之間了解共同的語言	領域知識可以重複使用	領域中的假設可以更明確	有助於區分領域知識和操作知識	有助於分析領域知識
-------------------	------------	-------------	----------------	-----------

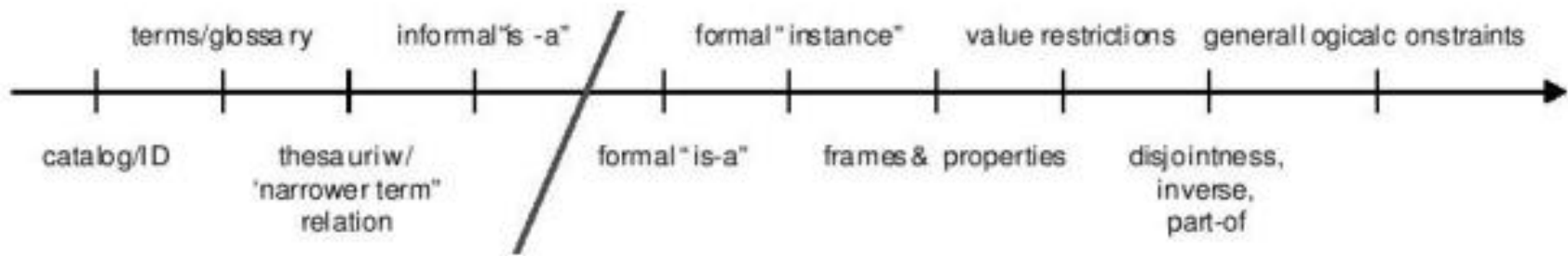


Figure 2.1: An Ontology Scale by Lassila and McGuinness, 2001

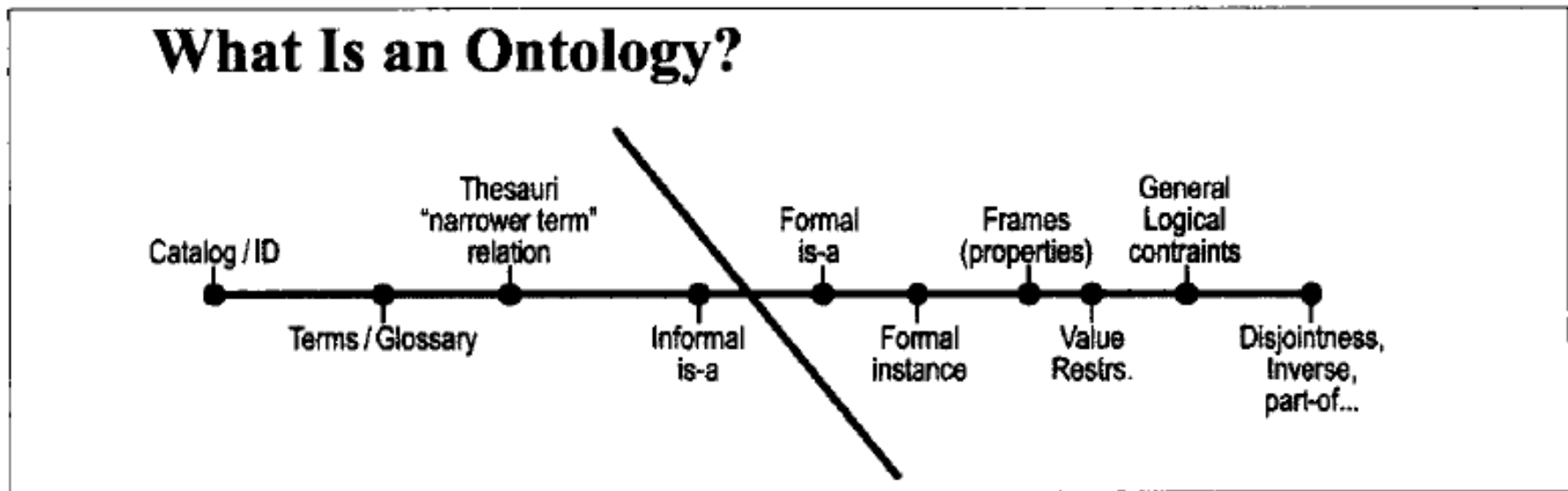


Figure 9.3 McGuinness's taxonomy of ontologies.

(Legg, 2007)

語意網的Ontology工具

- Dublin Core Metadata Element Set

Thesaurus
Ontologies

- Topic Maps
- RDF Schema
- DAML+OIUO
WL
- SHOE

Ontologies
with
(Largely) DL
Expressivity

- Cyc
- SUMO
- SUO

Ontologies
with First-
Order Logic
Expressivity
(or Higher)

從案例中學習 – 知識本體和鏈結資料的應用

1. 國會分類法的知識本體建構

2. 中央研究院領域詞彙分類法 (Domain Lexico-Taxonomy; DLT)、建議上層共用知識本體(Suggested Upper Merged Ontology SUMO)與中央研究院中英雙語知識本體詞網 (Sinica BOW)

3. 口述歷史鏈結資料和知識本體之應用

Ontology for LC Classification

從案例中
學習

The Library of Congress > Linked Data Service >

Ontology for Library of Congress Classification

Search

Enter Keyword or Phrase

All
LC Subject Headings
LC Name Authority File
LC Classification
LC Children's Subject Headings

Search Reset

**Please Note: LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the scheme. This is temporary while the impact of adding LCC to the current system is better understood.

1. LC subject headings
2. LC Name Authority File
3. LC Classification
4. LC Children's Subject Headings
5. LC Genre/Form Terms
6. LC Medium of Performance
Thesaurus for Music
7. LC Demographic Group Terms
8. Thesaurus for Graphic Materials
9. AFS Ethnographic Thesaurus
10. Cultural Heritage Organizations

11. MARC Relators
12. MARC Countries
13. MARC Geographic Areas
14. MARC Languages
15. ISO639-1 Languages
16. ISO639-2 Languages
17. ISO639-5 Languages
18. Extended Data/Time Format
19. Identifiers
20. Carriers

21. Content Types
22. Media Types
23. Resources Types
24. MARC Genre/Form
25. MARC Subject Schemes
26. Classification Schemes
27. Description Convention
28. Publication Frequencies
29. BIBFRAME Ontology
30. Preservation Vocabularies(all)

31. Action Granted
32. Agent Type
33. Content Location Type
34. Copyright Status
35. Cryptographic Hash
Functions
36. Environment Characteristic
37. Environment Purpose
38. Event Related Agent Role
39. Event Related Object Role
40. Event Type

41. Format Registry Role
42. Hardware Type
43. Inhibitor Target
44. Inhibitor Type
45. Object Category
46. Preservation Level
Role
47. Relationship
SubType
48. Relationship Type
49. Right Basis
50. Rights Related Agent
Role
51. Signature Encoding
52. Signature Method
53. Software Type
54. Storage Medium

http://id.loc.gov/ontologies/lcc.html#p_isTableFor

“Although a little time has been spent linking LCC resources with LCSH and LC Names resources, more work can be done here too. *Naturally, linking LCC to Dewey would be a high priority endeavor. There has also been a long-standing desire to use the Library of Congress Classification as an entry point to the bibliographic catalog versus merely a means to locate a book on a shelf*

- 將LCC和DCC互相連結很重要，不是只在指引到書架找書，而是在書目資料的指引...

Ford, Kevin (2013). Library of Congress Classification as linked data. JLIS.it. Vol. 4, n. 1 (Gennaio/January 2013). DOI: 10.4403/jlis.it-5465

陳永祥黃居仁利用領域詞彙分類與雙語知識本體詞網輔助主題資訊搜尋

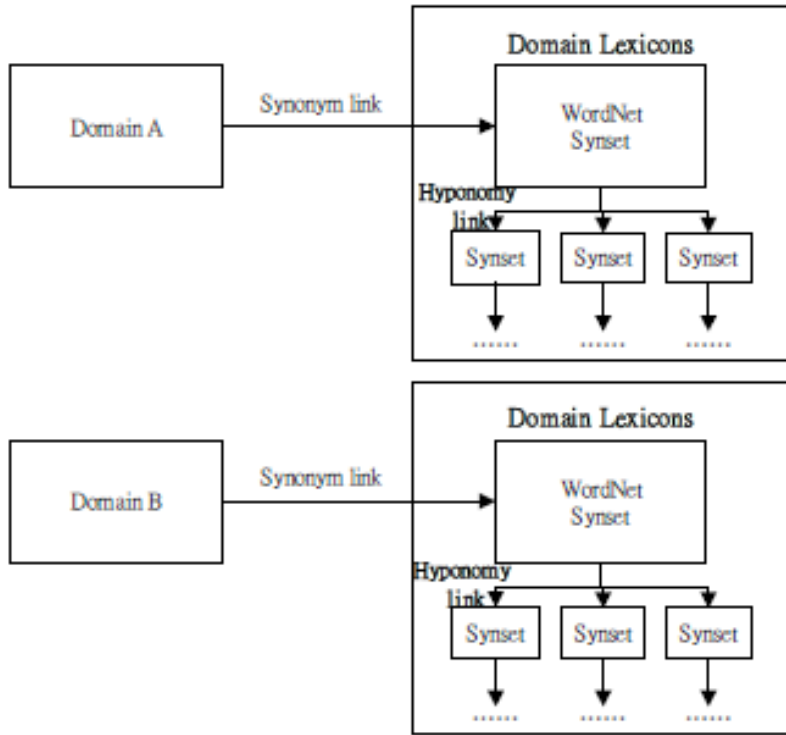


圖 2 DLT 領域詞表建構方法

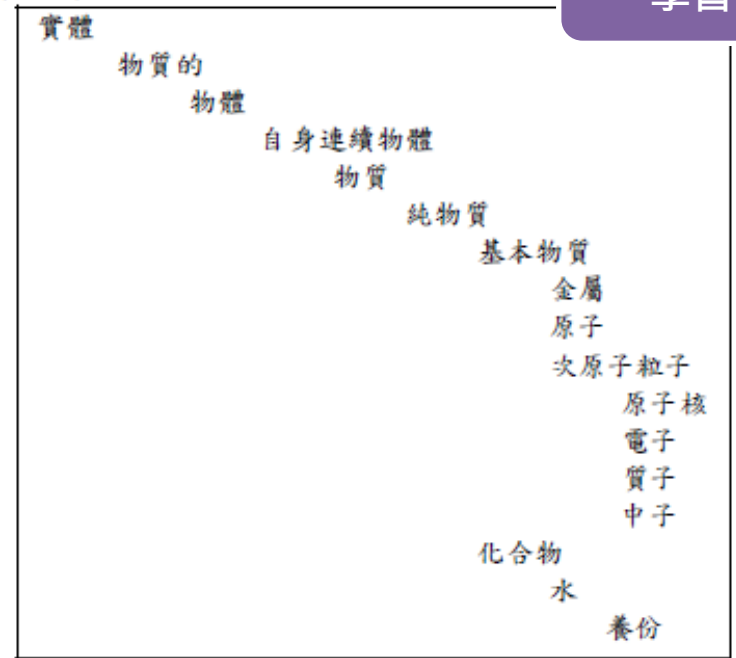


圖 3 SUMO 階層節點示例

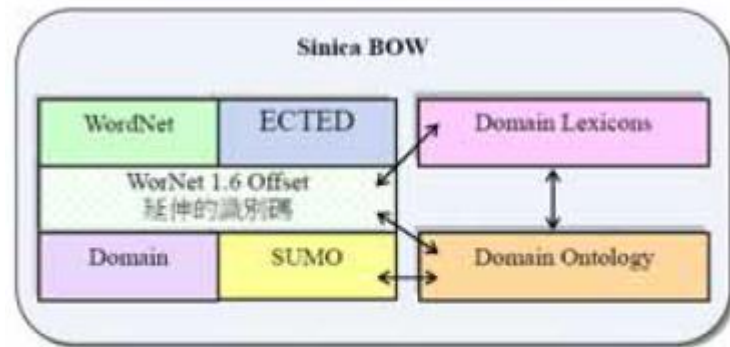


圖 4 SINICA BOW 架構圖

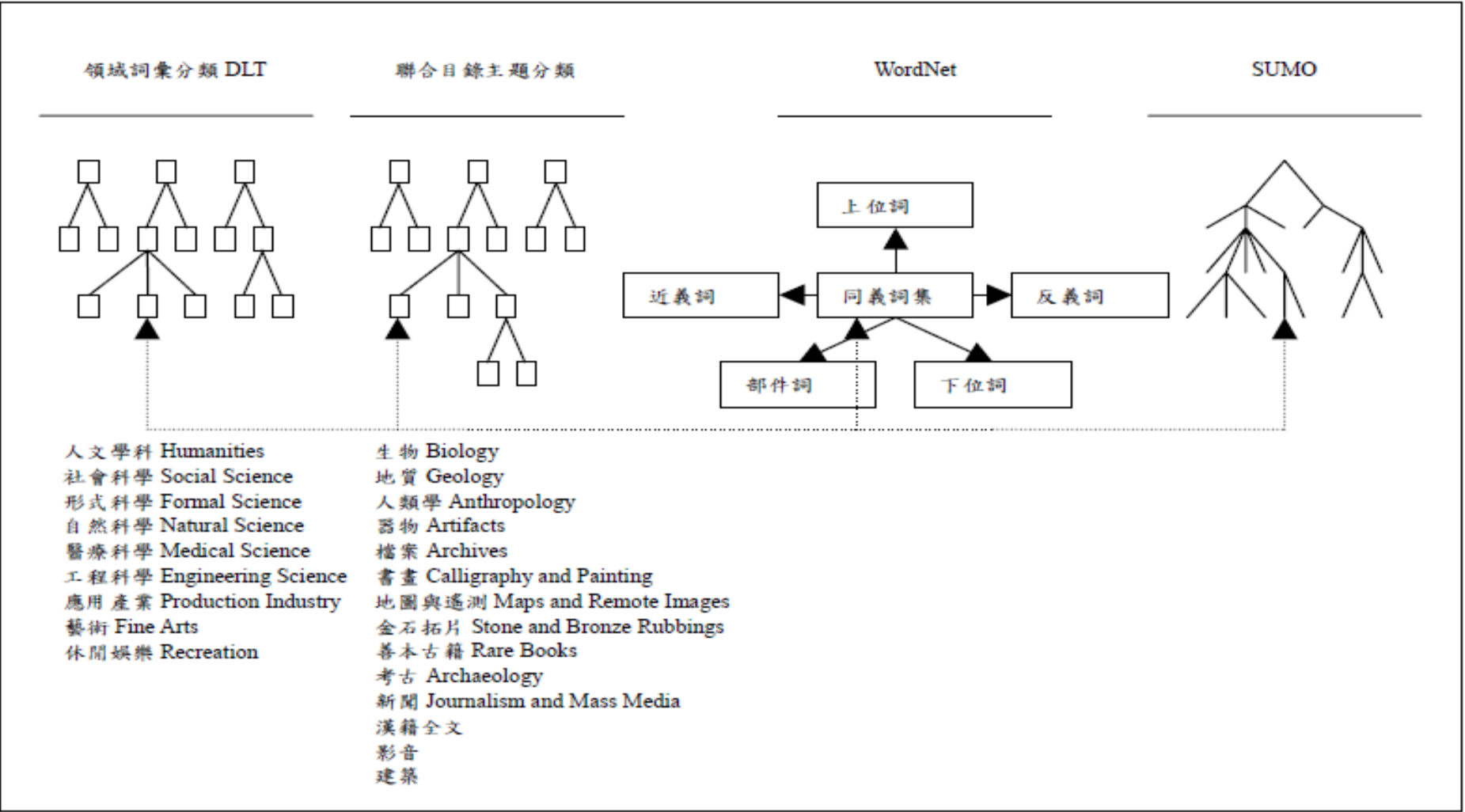


圖 6 領域詞彙對應模式

陳永祥，黃居仁。利用領域詞彙分類與雙語知識本體詞網輔助主題資訊搜尋。第五屆數位典藏技術研討會。<<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.7.18.56698244.2431219.pdf>>

表 2 聯合目錄主題分類資料對應表

聯合目錄主題分類	領域分類	WordNet 同義詞集 (同義詞集 offset)	SUMO 對應關係 ; SUMO www.sumo-project.org/
生物 Biology	自然科學 -> 生物學	biota, biology (05987709)	Hypemym ; Organism
地質 Geology	自然科學 -> 地球科學 -> 地質	geological formation, geology, formation (06691504)	Hypemym ; SelfConnectedObject
人類學 Anthropology	社會科學 -> 人類學	social anthropology, cultural anthropology (04673338)	Instantiation ; FieldOfStudy
器物 Artifacts	藝術 -> 造形藝術	artificiality (03757381)	Hypemym ; SubjectiveAssessmentAttribute
	藝術 -> 裝飾藝術 -> 手工藝		
檔案 Archives	人文學科 -> 圖書館學	archive, archives (02206789)	Hypemym ; EducationalOrganization
	人文學科 -> 史學		
書畫 Calligraphy and Painting	藝術 -> 國畫藝術	calligraphy, penmanship (04826894)	Hypemym ; Text
	藝術 -> 裝飾藝術 -> 書法	painting, picture (03079051)	Hypemym ; ArtWork
地圖與遙測 Maps and Remote Images	藝術 -> 攝影 -> 遠距攝影	Map(02965788)	Hypemym ; Icon
	藝術 -> 國畫藝術 -> 製圖法		
	應用產業 -> 交通		
	自然科學 -> 地球科學		
金石拓片 Stone and Bronze Rubbings	人文學科 -> 古文字學	Rubbing (03259790)	Hypemym ; copy
		Bronze	Hypemym ; ArtWork

表 3 SUMO 節點對應表

聯合目錄主題分類	SUMO 節點位置
生物 Biology	1.1.1.1.2.4.8.,organism,生物體
地質 Geology	1.1.1.1.,self connected object,自身連續物體
人類學 Anthropology	1.2.7.32.,field of study,研究領域
器物 Artifacts	1.2.4.13.61.211.,subjective assessment attribute,主觀評價屬性
檔案 Archives	1.1.1.4.12.31.51.,educational organization,教育組織
書畫 Calligraphy and Painting	1.1.1.1.3.8.22.,text,文本 1.1.1.1.2.5.13.,art work,藝術品
地圖與遙測 Maps and Remote Images	1.1.1.1.3.7.,icon,圖示
金石拓片 Stone and Bronze Rubbings	1.2.5.14.62.214.202.142.215.,copy,複製 1.1.1.1.2.5.13.,art work,藝術品
善本古籍 Rare Books	1.1.1.1.3.8.22.45.,book,書籍 1.1.1.1.3.8.22.,text,文本
考古 Archaeology	1.2.7.32.,field of study,研究領域
新聞 Journalism and Mass Media	1.2.4.13.60.210.,occupational role,職業角色 1.1.1.1.3.8.22.,text,文本
漢籍全文	1.1.1.1.3.8.22.,text,文本
影音	1.2.7.32.,field of study,研究領域 1.1.1.1.3.8.22.,text,文本
建築	1.1.2.8.43.93.94.,constructing,建構

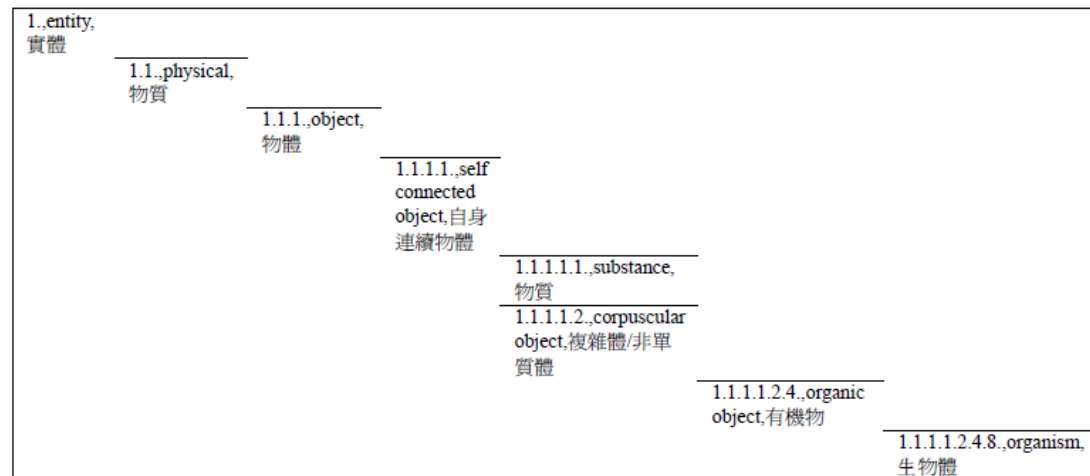


圖 7 “生物”節點在 SUMO 中的位置

從中研院領域詞彙分類與雙語知識本體詞網 輔助主題資訊搜尋計畫學到甚麼？

1. **查詢檢索模式**可以透過**拓展多種知識組織**方法來進行 -- 該研究就是透過連結DLT, Sinica BOW與WordNet，在瀏覽檢索上提供一查詢拓展模式。
2. Sinica BOW所提供之中英雙語對應詞彙，可提供作為聯合目錄展示系統發展**中英雙語查詢介面**之使用。
3. **提供知識地圖** -- 透過主題分類與SUMO節點的對應，可了解數位典藏計畫中典藏品在整體人類知識架構中所在之環境位置，亦對於延伸了解相關知識提供了方向。
4. 可瞭解**自然語言處理技術**與語言資源在數位典藏計畫中的定位及應用方向。

表 5 聯合目錄高頻查詢詞

關鍵詞
胡台麗；陶；王建民；酒器；虱目魚；陶 AND 陶器；西拉雅； <u>蝴蝶</u> ；聯勤；蓮；台中；大乘；董氏針灸；大稻埕； <u>蝶</u> ；文徵明；坐骨 AND 神 AND 痛；棒球；翠玉白菜；鬼；…； <u>紫斑蝶</u> ； <u>昆蟲</u> ； <u>拺蝶科</u> ； <u>紫斑蝴蝶</u> ； <u>蝶科</u> ； <u>鳳蝶總科</u> ； <u>鳳蝶</u> ； <u>鱗翅目</u> ；……

表 6 蝴蝶領域詞表

”蝴蝶”領域詞表
衣蛾(tineid), 衣蛾(clothes_moth), 夜盜蛾 (armyworm), 夜盜蛾 (armyworm), 昆蟲(insect), 昆蟲(bug), 昆蟲(coreid_bug), 枯葉蛾(lasiocampid), 枯葉蛾(eggar), 麥蛾(gelechiid), 麥蛾 (Gelechia_gossypiella), 麥蛾(angoumois_moth), 菜粉蝶(small_white), 菜粉蝶 (southern_cabbage_butterfly), 葉蟲(leaf_bug), 葉蟲 (mirid_bug), 蛹(pupa), 蛹(chrysalis), 蛾(moth), 蛾 (gypsy_moth), 鳳蝶(emperor_butterfly), 鳳蝶 (emperor), 穀蛾(tineid), 穀蛾 (grain_moth), 燈蛾 (arctiid), 燈蛾(tiger_moth), 蠶蛾(bombycid), 蠶蛾 (giant_silkworm_moth)

口述歷史鏈結資料和知識本體之應用

- These relationships include:
 - has met
 - is an acquaintance of
 - is a friend of
 - is a close friend of
 - is influenced by
 - is a mentor of
 - collaborated with:
 - was in a band together with
 - played with
 - was a member of the band of
 - toured with
 - was the bandleader for

More on where these terms come from can be found on our Ontology page (*currently under development*).

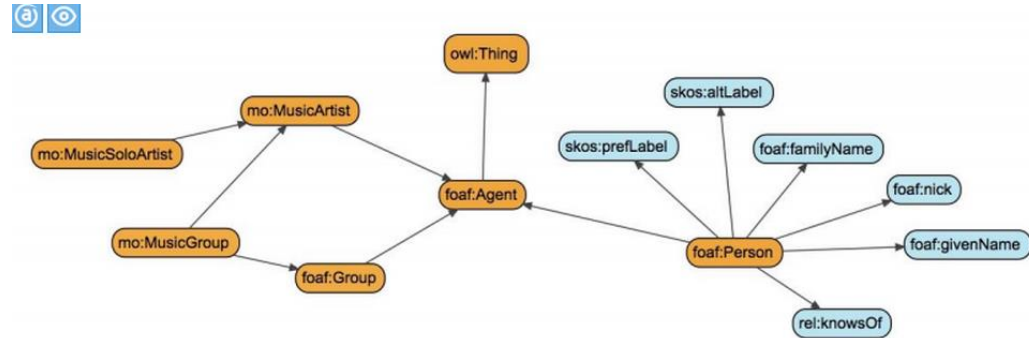
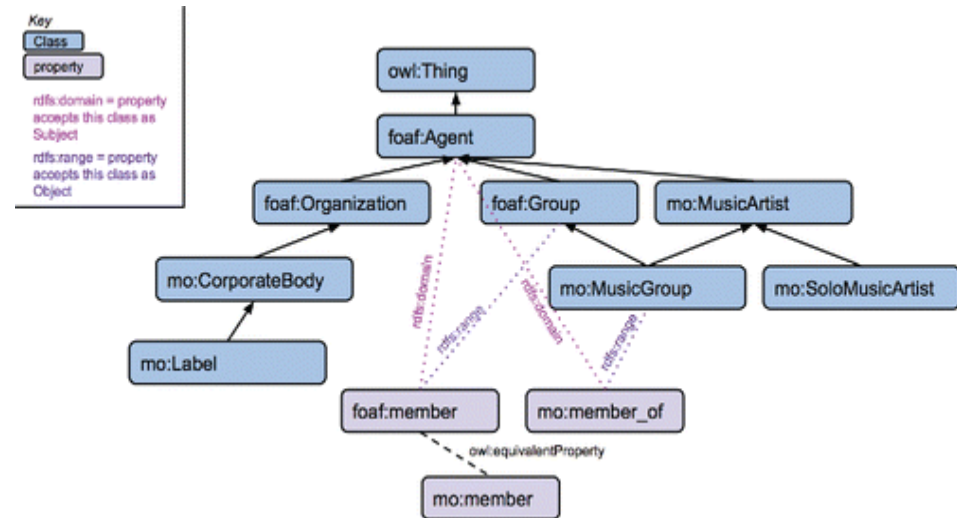


Figure 1 Diagram of core set of classes and properties.

Thorsen, Hilary. Ontologies in the Time of Linked Data
<http://www.iskocus.org/NASKO2015proceedings/Thorsen.pdf>



1. Johanna Drucker(2013)曾經說過「沒有任何分類系統是價值中立的，客觀的，或自證的，所有的分類系統都是承自自己內部生產的思想印記」。承認並擴大思維？
http://dh101.humanities.ucla.edu/?page_id=33
2. 口述歷史的知識本體，從哪裡開始建構？在哪個地方落錨？
 - Mazé, Elinor A. “Metadata: Best Practices for Oral History Access and Preservation,” in *Oral History in the Digital Age*, edited by Doug Boyd, Steve Cohen, Brad Rakerd, and Dean Rehberger. Washington, D.C.: Institute of Museum and Library Services, 2012,
<http://ohda.matrix.msu.edu/2012/06/metadata/>

參考文獻 (Selective)

- Legg, C. (2007). Ontologies on the Semantic Web. *Annual Review of Information Science and Technology (ARIST)*, 41(1), 407-451.
- Noy, Natalya F. and McGuinness, Deborah L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*.
http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- Ora Lassila & Deborah L. McGuinness: "The Role of Frame-Based Representation on the Semantic Web", Knowledge Systems Laboratory Report KSL-01-02, Stanford University, 2001; Also appeared as Linköping Electronic Articles in Computer and Information Science, Vol. 6 (2001), No. 005, Linköping University, 2001
- 施正鋒(2013)。原住民族知識體系與客家知識體系。台灣原住民族研究學報，3(2)：115-142。
- 陳永祥，黃居仁。利用領域詞彙分類與雙語知識本體詞網輔助主題資訊搜尋。第五屆數位典藏技術研討會。
<<http://www.ling.sinica.edu.tw/eip/FILES/publish/2007.7.18.56698244.2431219.pdf>>

謝謝聆聽 敬請不吝指教！

吳美美博士
教授兼所長
臺灣師範大學圖書資訊學研究所
meiwu@ntnu.edu.tw

