

# 多領域文件集之詞彙概念擴展與知識架構之建立

## Conceptual Expansion and Ontological Mapping of Multi-domain

### Documents

陳永祥 Yong-Xiang Chen

中央研究院語言學研究所 臺灣大學資訊工程研究所  
[yxchen@gate.sinica.edu.tw](mailto:yxchen@gate.sinica.edu.tw)

柯綉玲 Xiu-Ling Ke

中央研究院語言學研究所  
[vitake@gate.sinica.edu.tw](mailto:vitake@gate.sinica.edu.tw)

陳克健 Keh-Jiann Chen

中央研究院資訊科學研究所  
[kchen@iis.sinica.edu.tw](mailto:kchen@iis.sinica.edu.tw)

黃居仁 Chu-Ren Huang

中央研究院語言學研究所  
[churen@gate.sinica.edu.tw](mailto:churen@gate.sinica.edu.tw)

### 摘要

傳統資訊檢索過程經常透過查詢擴展技術來增加檢索結果的數量。在大型數位博物館中，考量存放的典藏品項具有領域特殊性，以及不同品項由各自適用的文字描述所造成之差異性，因此傳統查詢擴展方式不盡合用。本研究提出概念擴展之構想，首先由數位博物館典藏品項的標題中抽取出成分詞集，再透過中英雙語知識本體（BOW）將成分詞對應至建議上層知識本體（SUMO），藉此可將成份詞轉換成為概念並對應至具有結構的知識本體節點上，再由集群演算法計算相近節點並將分散之對應節點聚合成具有代表性的群集，最後以構詞分析所得之規則進行群集縮減，決定出符合構詞分析規則的群集用以進行概念擴展。研究成果除提出知識概念擴展流程外，亦以數位典藏國家型科技計畫典藏項目為例，歸納出保存多領域典藏品之數位博物館中文標題構詞樣式，進行分析探討，研究成果可作為機器自動處理之基礎。

關鍵詞：概念擴展，構詞樣式，知識本體，群集縮減

## 一、前言

傳統資訊檢索過程經常透過查詢擴展技術來增加檢索結果的數量。在大型數位博物館中，考量存放的典藏品項具有領域特殊性，以及不同品項由各自適用的文字描述方式所造成的差異性，因此傳統查詢擴展方式不盡合用。概念擴展的構想是以詞彙所代表的知識概念在知識本體中結構中找到相似的概念進行擴展，以使原本只能對應到少數概念的詞彙透過概念延伸得以對應至較多的詞彙，進而從知識概念層次的擴展提升數位博物館典藏品資料的資訊檢索效能。

「數位典藏國家型科技計畫」自 2002 年開始推動，旨在將珍貴的重要文物典藏加以數位化，建立國家數位典藏，以保存文化資產、建構公共資訊系統，促使精緻文化普及、資訊科技與人文融合，並推動產業與經濟發展。因規模龐大，因此目前開發整合型的成果查詢介面提供各界使用者查詢應用，分別為聯合目錄及公共展示系統。

透過知識本體的結構系統，可以比較嚴謹的將知識結構系統建立起來，本研究即以數位典藏國家型科技計畫所提供之 39,765 個典藏品標題（2~5 字詞）為實驗資料，透過中文詞彙網路（Chinese Wordnet; CWN）及中英雙語知識本體詞網（Sinica BOW）所建立的中文詞義分析與知識本體架構，將標題詞彙對應至 SUMO ontology 節點上進行概念擴展與群集，再以中文構詞樣式為標準進行群集縮減，以詞彙語義分析方式提供特殊領域典藏資料庫中資訊檢索可行之方案。

## 二、相關研究

### （一）語義相似度與查詢擴展

距離導向的相似度方法是從大的文字語料庫中去學出分布的相似度來建立模型，Leacock & Chodorow [1]，Resnik [2]，Lin [3]所提出的是三種在自然語言處理應用上很標準的方法。這些公式都是定義用來測量概念(Concept)上的相似度，而非詞彙(word)上的，但在轉換上可以用一對詞與詞之間，多組概念對概念相似度中最高的那組來作為語義相似度的代表。因此可以簡單轉換成詞與詞的相似度計算。

資訊檢索一般可針對檢索的資料類型區分為兩種，第一種是針對網際網路上所有的資料內容所進行的檢索，由於檢索的範圍太過廣泛，因此必須透過許多不同的策略來針對查詢關鍵字進行擴展以求找出使用者有興趣的內容，大多數的網際網路搜尋引擎網站所提供之服務皆屬於此類。第二種則是針對特定範圍資料所進行的資訊檢索，此類資訊檢索的使用者是在資料內容固定的情況下進行查詢，例如新聞媒體網站或是數位博物館網站。

一般資訊的檢索是用詞彙來代表概念。但概念與詞彙的關係並非都是一對一的，如同義詞（**Synonym**）用來表示多個詞彙都具有相同的概念，即一個概念可對應到多個詞彙。因此，在檢索時若能建構概念與詞彙間的明確關係，將有效提升檢索效益。在一般搜尋引擎的設計上，最常見的策略是利用詞形比對的方式作為資訊檢索的基本方法，再輔助以各種的查詢詞彙擴展或是相關統計運算結果來找出使用者感興趣的資料。

除了利用關鍵詞進行全文檢索（**Full-Text Search**）外，有些資訊檢索系統尚針對文件的內容進行分析，給予文件資料檢索標識（如主題詞彙或分類號），並使用索引詞彙來表示文件內容，資訊使用者與資訊檢索系統之間藉由索引詞彙與檢索詞彙之間的對應來達到擷取與過濾資訊的目的。查詢問句的擴展通常以使用者提供的檢索詞彙為基礎，當原始查詢問句的檢索效益不好時，則可以追加更多的詞彙來改善。關於查詢問句的擴展，尚有相關研究提出利用相關回饋（**Relevance Feedback**）或是使用知識架構（**Ontology**）的元知識（**Atom Knowledge**）來進行[4]。**Stiles** 是最早提出利用相關詞彙來改進檢索效益理論的學者之一[5]。

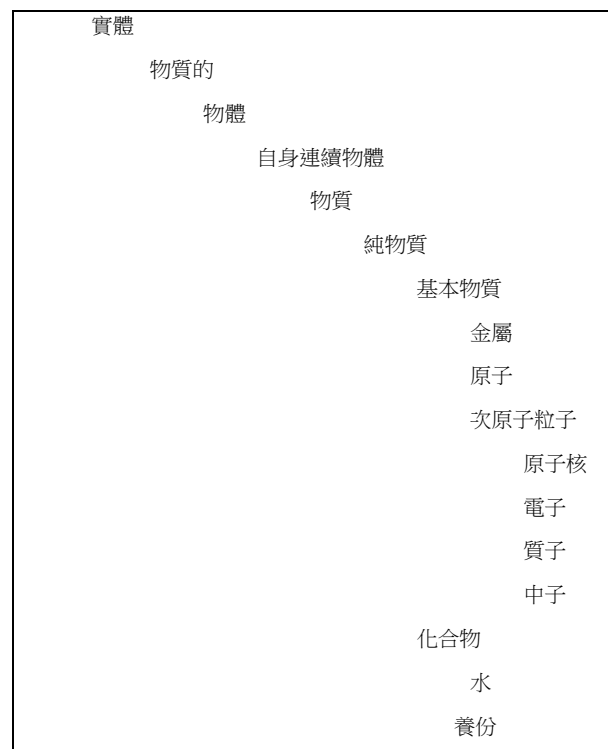
## （二）建議上層共用知識本體（**Suggested Upper Merged Ontology ; SUMO**）

**SUMO**（**Suggested Upper Merged Ontology**，建議上層共用知識本體）[6]是由 **IEEE** 標準上層知識本體工作小組所提出的知識本體架構，目的是發展成標準的上層知識本體，這將促進資料互通性、資訊搜尋和檢索、自動推理和自然語言處理。知識本體（**ontology**）類似於一組字典或術語表，但能夠使電腦處理更多內容的細節和其結構。透過知識本體可將人們有興趣的領域正規化為一套概念、關係和定理（**axiom**）。上層的知識本體被限制在 **meta** 的概念、一般、抽象或者哲學，因此足夠一般提出（在一定水準上）一個涵蓋廣闊範圍的領域區域[7]。特殊領域具體的概念不被包括在上層知識本體中，但是這樣的知識本體可提供特殊領域（例如：藥、財政、專案...等等）的知識本體結構的建立。**SUMO** 藉由最高層次的知識本體，鼓勵其他特殊領域知識本體以其為基礎衍生出其他特殊領域的知識本體，並為一般多用途的術語提供定義。目前 **SUMO** 已經和英語詞彙網路 **WordNet1.6** 版本作初步的連結。**SUMO** 中的節點以階層樹方式連結，如圖一所示。

## （三）中央研究院中英雙語知識本體詞網（**Sinica BOW**）

中英雙語知識本體詞網（**Sinica BOW**）[8]是一結合詞網（**WordNet**）知識本體與領域標記的詞彙知識庫，由中央研究院語言所中文詞彙網路小組與資訊所中文詞知識庫小組合作建置，從語言工程的角度，以台灣地區的語言使用為經驗基礎，提供語言和語言、語言和概念以及語言和領域的資訊，甚至是跨語言間的訊息。**Sinica BOW**以建立一完整精確的中英對譯資料庫及檢索介面為目的，作為數位典藏知識國際化的基礎；並持續建立各領域之雙語領域辭典，以作為各領域／典藏之雙語控制詞彙參考標準。中英雙語知識本體詞網同時提供具領域判斷能力之資訊檢索應用。此外，建立附加領域標記之雙語辭典及檢索介面使中英雙語知識本體詞網成為一知識加值雙語電子辭典。

Sinica BOW 主要使用的資源包含 WordNet、ECTED (English- Chinese Translation Equivalents Database) 以及 SUMO (Suggested Upper Merged Ontology, 建議上層共用知識本體)。其中 WordNet[9]是 1985 年普林斯頓大學認知科學實驗室以現代心理語言學理論所述的人類詞彙記憶為啟發所開發出的語意式電子字典,以每個同義詞集表達一種詞彙概念,將同義詞集區分為四種英文詞類:名詞、動詞、形容詞、副詞,並以二十幾種詞義關係組織同義詞集。由中研院資訊所與語言所合作建構的 ECTED 以 WordNet 為基礎,經由現有英中或中英電子辭典的詞形對應,為每個同義詞集詞義找出可能相對應的中譯詞組,再經由人工檢驗。尋找對譯盡可能的以詞彙而非描述性短語表達,目的在於讓每個同義詞集都有最適當的一至三個左右的中文對譯。[10]

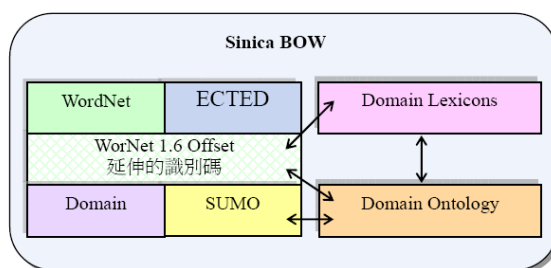


圖一、 SUMO階層節點示例

依據SUMO 2002年版資料,黃居仁等人[11]將系統介面以及概念節點進行中文化納入Sinica BOW之中並進行對應連結,其涵蓋11大類的概念,每大類又區分為二至五個類別,總共囊括3,912個概念。SUMO已經與WordNet1.6版本結合,且以同義(synonymy)、上位(hypernym)、體例(instantiation)這三種類別顯示同義詞集和SUMO概念間的對應關係。除此,更以「中國圖書分類法」為基準,並參考各知識分類與實際研究經驗,提出:包含九大類的知識分類(Knowledge Content),涵蓋427個領域。另外,並因應語言資源特性加入下列語言使用(Language Usage)的各類訊息:專名(說明文字符號的指涉)(Proper Name)、語體(說明文字符號的使用)(Genre/Strata)、各種語言/詞源(Language/Etymology)、各國地名(Country Name)。領域階層的建立在於替不同詞義中的詞彙項目區別其使用的領域。加註領域信息可降低詞彙歧異性,增加資料交換時的互通性,輔助領域詞彙庫之建構。Sinica BOW透過WordNet1.6 offset延伸所產生的識別碼作為媒介,進行串連,將每個資源以及各類訊息連結。

因WordNet1.6 offset延伸的識別碼可獲得原本WordNet存在的詞類、解釋、英文例句、同義詞集、各同義詞集間的詞義關係及其所屬詞彙。而SUMO概念與WordNet的連結，使得可透過該識別碼獲取詞義與概念搭配的訊息。以WordNet為基礎所建置的ECTED與針對WordNet同義詞集的各詞彙項目所給予的領域值，也是透過該識別碼獲取。而特殊領域詞彙庫，加上相對應的Sinica BOW識別碼，也可保留原始資源的資料庫格式和WordNet連結。

因領域知識本體則是在SUMO某些概念下進行延伸發展。每個特殊領域詞彙庫中的詞彙一樣具有所屬的概念，其所屬概念可能是SUMO或特殊領域知識本體的某一概念，特殊領域詞彙庫和領域知識本體的結合，使得透過該識別碼又串起所有的訊息。Sinica BOW的資源和架構如圖二所示。由於透過WordNet可以和同是以WordNet為基礎架構所建置的其他語系WordNet資源加以連結，例如：EuroWordNet[7]，因此以此基礎架構可編製成多語的詞彙網路，成為多語環境中所需之語言知識結構的基礎資料。



圖二、 SINICA BOW 架構圖

### 三、知識本體對應與構詞分析

#### (一) 研究資料

對於儲存多樣領域知識的數位博物館而言，將典藏品項涵蓋的知識概念對應至知識結構可作為許多延伸應用與研究之基礎，例如查詢擴展與跨語言知識交換。而典藏品項之標題名稱可用來作為典藏品特性的具體描述。本研究採用中央研究院中文詞知識庫小組所開發的斷詞系統對數位典藏國家型計畫中的典藏品標題進行斷詞[12][13]，取其中二至五字詞所構成的典藏品標題作為研究資料，共 39,765 筆。這些標題分屬於生物、考古、地質、人類學、檔案、拓片、器物、書畫、地圖與遙測、善本古籍、新聞、漢籍全文、影音、建築等十四個主題領域，其中 96%的標題詞為詞典中未收錄之項目，如馬銀花、嘉義中學與瑪瑙雙耳杯等。

對這些詞典中未收錄的標題詞而言，由於缺乏足夠之上下文資訊，不易以計算詞彙共現度的方式來進行詞彙擴展。因此，本研究在詞彙層次上對這些未知詞進行分析與處理。其中，以詞彙所包含的詞義概念進行概念擴展以找出構成詞彙的概念在知識本體上的位置並延伸是本文研究的主要方向。以標題詞所包含的概念作為基本單位則可將之與儲存知識概念的知識本體進行對應連結。而在進行概念分析上，因詞彙與概念的對應過程可能產生歧義，所以需要納入處理歧義的機制。

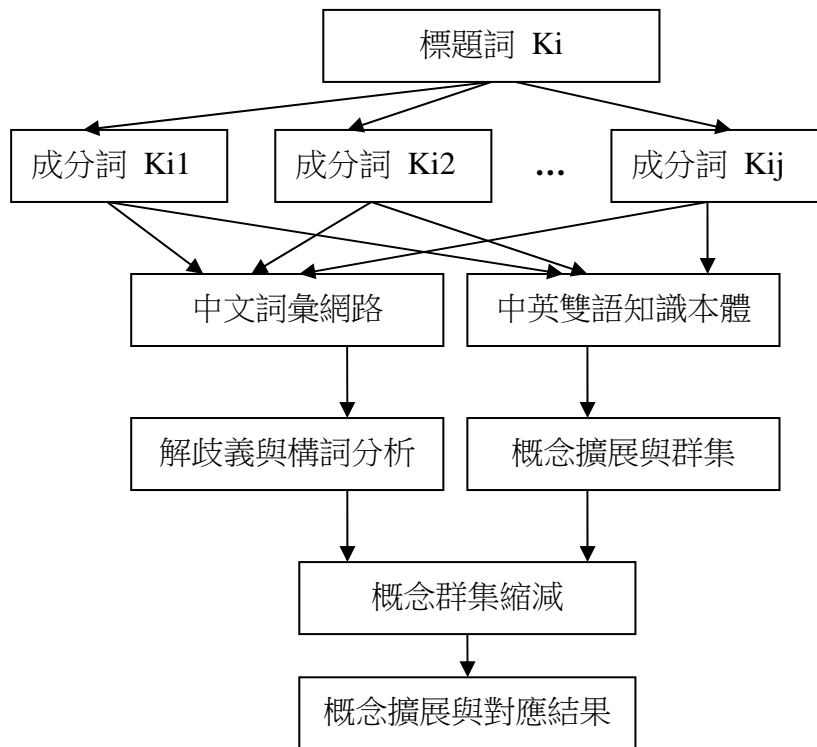
## (二) 知識本體對應策略

現有已開發之知識本體可以用來作為知識交換與知識儲存之基礎，如 SUMO 與 MILO (Mid-Level Ontology)。但一般數位博物館典藏項目均具有其特殊性，因此無法順利在泛用的知識本體上找到對應節點，例如數位典藏國家型計畫中便有超過 96% 的項目無法在 SUMO 中直接找到對應節點。

由於絕大部分典藏品標題無法直接對應至詞典紀錄，因此在建構整個典藏內容的知識系統時，可以選擇的方向主要有二：(1) 自行建立領域特用之知識本體或 (2) 透過對應策略將典藏品項之標題對應至現有知識本體節點上。

依據研究目的，本研究提出一對應策略，透過中文構詞分析，先將無法直接對應之標題詞切分為較小的成分詞，再以詞形比對的方式將成分詞與中文詞彙網路及中英雙語知識本體中現有節點產生比對連結，得到一個對應結果。藉此方式可於不花費高額成本的情況下，在概念連結層次將標題詞彙擴展並對應至知識本體中。詳細步驟如圖三所示。

在成分詞的定義上，本研究將 N 字標題詞切分成二字成分、三字成分、...、(N-1) 字成分，將所有的可能成分定義為成分詞集合，再透過這些成分詞與 149,751 筆 BOW 概念進行詞形比對與連結。



圖三、知識概念擴展流程

### (三) 構詞分析

將標題詞切分為成分詞的對應方式具有增加對應結果數量之效果，然而亦有納入太多歧義資料的副作用。因此，為了在知識本體對應過程過濾過多的非目標結果，本研究透過 CWN 中的詳細詞義資料進行中文構詞分析解決歧義問題，將成分詞構成標題詞的組合方式以人工方式進行語言分析，歸納出兩種主要的樣式：事件驅動與實體驅動。

#### 一、事件驅動 event driven：(專有名詞(人名) +) 動詞 (+受詞)

##### 1. 專有名詞(人名) + 動詞

例：

羅漢松 / 哭了  
小學生 / 跳舞

##### 2. 專有名詞(人名) + 動詞 + 受詞

例：

团仔 / 騎 / 木馬  
我 / 敬愛 / 國父

### 3. 動詞 + 專有名詞(人名)

例：

回味／蔡惠風

找回／太魯閣

## 二、實體驅動 entirety driven：無動詞的

### 1. 修飾語 + 中心語

表一、修飾語 + 中心語概念分析

修飾語 + 中心語概念分析	例
屬性／專有名詞(地名)／中心語	垂花／蓬萊／葛
屬性／專有名詞(人名)／中心語	黃文／獻公／集
專有名詞(地名)／屬性／中心語	城武縣／稅／銀 湘潭縣／銀／錠
年代／材質／中心語	光緒年／銀／錠 漢／青玉／璫
年代／性質／中心語	全蜀／藝文／志 漢／人物畫／像
年代／結構／中心語	全唐／聲律／論 漢魏／叢書／選
材質／功能／中心語	光纖／連接／器 網結草／雨／衣
材質／樣式／中心語	瑪瑙／雙耳／杯 黃瑪瑙／煙／壺
性質／中心語	印花稅／條例 壞死性／腸炎

### 2. 中心語(+連接詞或介系詞) 中心語

例：

靈芝／和／牛樟

回顧／與／展望



## 四、概念擴展與群集縮減

### (一) 概念擴展與群集

在透過成分詞將標題詞對應至 BOW 概念之後，便可以擴展得到一組標題詞及其成分概念，以及成分概念連結到 SUMO 上的概念節點集合。而因知識本體本身所具備的結構系統，所以可使用節點間的概念距離定義出相似概念，進而計算相似群集，使得成分詞得以擴展對應至知識本體上並透過群集的方式形成群內差異小，群間差異大的概念群集。本研究設定相似概念節點為在 SUMO 架構中距離各成分概念距離為 2 以內的概念節點。

每個經由 BOW 擴展的知識概念都可以在 SUMO 中找到對應節點，由於 SUMO 的樹狀結構，這些概念節點集合所形成的子樹便可視為一個群集。表二以“瑪瑙雙耳杯”作為例子，說明成分詞經過對應至 BOW 的成分概念擴展之後，藉由 SUMO 概念節點位置計算相似度距離所得到的概念群集。由本例中可知瑪瑙雙耳杯的成分詞對應到 SUMO 時，共可在知識本體樹狀結構中形成七個主要群集。

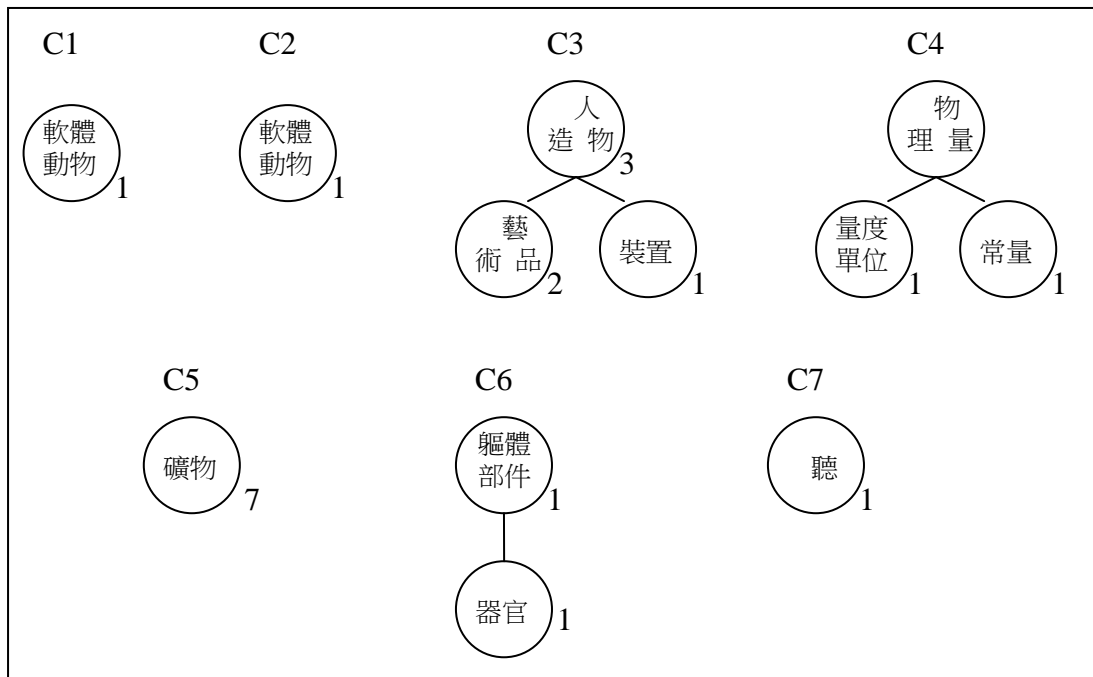
表二、概念群集範例：瑪瑙雙耳杯

瑪瑙雙耳杯					
BOW 擴展結果	WN ID	SUMO node	SUMO 中譯	SUMO 概念節點位置	概念群集
瑪瑙貝	01466296N	mollusk	軟體動物	1.1.1.1.2.4.8.14.15.9.,mollusk,軟體動物,C,_mollusk	cluster 1
瑪瑙貝	01466296N	mollusk	軟體動物	1.1.1.4.11.25.46.57.,mollusk,軟體動物,C,_mollusk	cluster 2
瑪瑙花紋的搪瓷鐵器	02169007N	art work	藝術品	1.1.1.1.2.5.13.,art work,藝術品,C,_art_work	cluster 3
瑪瑙紋搪瓷器	02169007N	art work	藝術品	1.1.1.1.2.5.13.,art work,藝術品,C,_art_work	cluster 3
雙耳瓶	02185088N	artifact	人造物	1.1.1.1.2.5.,artifact,人造物,C,_artifact	cluster 3
雙耳黏土窄瓶	02185088N	artifact	人造物	1.1.1.1.2.5.,artifact,人造物,C,_artifact	cluster 3
雙耳平底酒杯	03291208N	artifact	人造物	1.1.1.1.2.5.,artifact,人造物,C,_artifact	cluster 3
雙耳式耳機	02809404N	device	裝置	1.1.1.1.2.5.16.,device,裝置,C,_device	cluster 3

BOW 擴展結果	WN ID	SUMO node	SUMO 中譯	SUMO 概念節點位置	概念群集
瑪瑙線	09870127N	unit of measure	測量單位	1.2.3.11.50.,unit of measure,量度單位,C,_unit_of_measure	cluster 4
瑪瑙線	09870127N	constant quantity	常量	1.2.3.11.48.,constant quantity,常量,C,_constant_quantity	cluster 4
瑪瑙	10543998N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
(礦)苔紋瑪瑙	10544179N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
瑪瑙	10617402N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
紅瑪瑙	10617402N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
彩紋瑪瑙	10740932N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
紅條紋瑪瑙	10740932N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
纏絲瑪瑙	10740932N	mineral	礦物	1.1.1.1.1.2.4.,mineral,礦物,C,_mineral	cluster 5
雙耳的	00236774A	body part	軀體部件	1.1.1.1.2.4.9.18.,body part,軀體部件,C,_body_part	cluster 6
雙耳心	04189008N	organ	器官	1.1.1.1.2.4.9.18.23.,organ,器官,C,_organ	cluster 6
雙耳的	02509854A	hearing	聽	1.1.2.8.38.84.87.92.,hearing,聽,C,_hearing	cluster 7

## (二) 群集縮減

由於概念擴展後所得到的群集往往包含太過龐雜的內容，因此有必要建立一個群集縮減的機制以濾去與原始標題差異較大的概念群集，僅保留主要相關的概念群集。此時，前述由人工歸納之構詞樣式即可扮演過濾器之角色。以瑪瑙雙耳杯一例，本研究即以構詞原則區辨出此標題詞中主要的中心語以及修飾語，再以此構詞樣式至知識本體中相對應分支中過濾出具有高相關性之對應群集，由已擴展並分群後的概念群集中保留對應正確知識本體分支之群集，而拋棄其餘群集不符合構詞原則之群集。



圖四、SUMO 子樹群集與次數分布（以“瑪瑙雙耳杯”為例）

此方法中關鍵的判斷基礎為各群集所連接的 SUMO 概念節點與構詞樣式的搭配。由“瑪瑙雙耳杯”範例可得知在“修飾語 + 中心語”的構詞樣式下，群集 3 所指涉的 SUMO 概念（人造物、裝置、藝術品）恰可表達中心語的概念。而群集 5 所指涉的 SUMO 概念（礦物），以及群集 6 所指涉的 SUMO 概念（軀體部件、器官）則用來傳達本例的修飾語概念，分別為材質與樣式。因此透過構詞樣式所進行的群集縮減便可將原本擴展至七個群集的所有概念縮減成群集 3、群集 5 以及群集 6 等三個具有代表性的群集。圖四中各 SUMO 子樹節點右側之數字為成分詞相關概念數量，作為選取群集與否的加權條件。經群集縮減後，原始標題詞的相似概念延伸詞集便是兩個縮減後概念群集中的詞彙。

值得注意的是，基於群集中包含節點的數量多寡，我們指定給群集 5 較高的重要程度，這是以詞典觀念來進行的設定，因為詞典中包含較多的項目自然是概念上重要的群集項目。而修飾語的類別經本文分析可得到有屬性、專有名詞、年代、材質、性質、結構、功能、樣式、性質等九大類。

## 五、結論

本研究主要目的在於探討以概念擴展的方式將原先具有領域特殊性的數位博物館典藏品標題進行成分概念分析，並對應至知識本體上的節點。再以構詞樣式將概念群集進行縮減，得到關聯概念群集。建構群集並以構詞樣式篩選關聯群集的一個好處是群集間無互斥性，可避免競爭而犧牲有代表性的群集。結果將使不同領域之典藏品能藉由標題的連結而整合成一知識系統。在研究中以人工歸納方式整理出構詞樣式並提出具有代表性的範例作為說明，可作為未來大量自動化處理之基礎。

由本研究所提出之概念與研究設計可針對儲存大量多領域知識的單一典藏機構文字資料進行概念擴展，對於數位博物館相關研究可有所助益，特別是在漢語數位博物館的資訊檢索應用上。同時亦可作為查詢擴展相關研究之參考。而由於專有名詞辨識及處理上之困難，後續研究上可導入合適的名稱辨識方法以使處理範圍能更臻完整。

## 參考文獻

- [1] C. Leacock and M. Chodorow, "Combining local context and WordNet sense similarity for word sense disambiguation," In *WordNet, An Electronic Lexical Database*. The MIT Press, 1998.
- [2] P. Resnik, "Using information content to evaluate semantic similarity," In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, Canada, 1995.
- [3] D. Lin, "An information-theoretic definition of similarity," In *Proceedings of the 15th International Conference on Machine Learning*, Madison, WI, 1998.
- [4] S. Gauch and J. B. Smith, "An Expert System for Automatic Query Reformation," *Journal of the American Society for Information Science*, vol. 44, no.3, 1993.
- [5] H. Chen, T. D. Ng, J. Martinez, and B. R. Schatz, "A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System," *Journal of the American Society for Information Science*, vol. 48, no.1, pp.17-31, 1997.

- [6] Suggested Upper Merged Ontology, <http://www.ontologyportal.org/>
- [7] I. Niles and A. Pease, "Toward a Standard Upper Ontology," In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Chris Welty and Barry Smith, eds, Ogunquit, Maine, 2001.
- [8] 中央研究院中英雙語知識本體詞網 The Academia Sinica Bilingual Ontological Wordnet (Sinica BOW) , <http://BOW.sinica.edu.tw>
- [9] WordNet, <http://www.cogsci.princeton.edu/~wn/>
- [10] C. R. Huang, E. I. J. Tseng, D. B. S. Tsai, and B. Murphy, "Cross-lingual Portability of Semantic relations: Bootstrapping Chinese WordNet with English WordNet Relations," Language and Linguistics, vol. 4.3, pp. 509-532, 2003.
- [11] C. R. Huang, R. Y. Chang, and S. B. Lee, "Sinica BOW (Bilingual Ontological Wordnet): Integration of Bilingual WordNet and SUMO," 4th International Conference on Language Resources and Evaluation (LREC2004), Lisbon. Portugal, 2004.
- [12] W. Y. Ma and K. J. Chen, "Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff," Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing, pp. 168-171, 2003.
- [13] W. Y. Ma and K. J. Chen, "A Bottom-up Merging Algorithm for Chinese Unknown Word Extraction," Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing, pp. 31-38, 2003.