

「廣群芳譜」—中國歷代植物詞彙知識庫與領域知識本體之建立

謝舒凱 張舒茗 張如瑩 張淳涵 羅鳳珠* 黃居仁

中央研究院語言學研究所

*元智大學中文系

{churen, shukai, candyc}@gate.sinica.edu.tw

摘要

本文介紹了一種建構領域知識的方法。以廣群芳譜為例，我們利用詞彙網路與建議上層知識本體作為知識架構，以進一步分析其知識內容。我們認為此法可作為跨時代之領域知識本體建構之參考。

關鍵詞：領域知識本體 領域詞彙 詞彙網路 建議上層知識本體。

1. 前言

1.1. 研究《廣群芳譜》的動機

當我們在郊外看到一朵盛開的蓮花或是一棵松樹，即使我們不知道他們叫什麼名字，但我們能夠分辨「啊！這是花，這是樹」，這就是我們開始在利用語言對我們所認知的世界做概念的分類。語言的使用是背後概念的表徵，「植物」、「花」、「樹」等都是概念，概念的分類可粗可細，也可因人而異，因領域知識系統不同而有別。此外，時間、空間也都是重要的影響因素。我們說語言是概念的表徵，是說語言在一定程度上反映了說話者認識這個世界的方式，我們懂一種語言，也就是表示我們能夠掌握該語言的概念系統。同樣的，如果我們想要了解一種知識體系，就必須先求能理解其概念系統。

我們對《廣群芳譜》知識分類的方式感到好奇，是因為這本書專門蒐羅、分類中國文獻上有紀錄的植物，可說是一部中國歷代植物詞彙的知識庫，但是，書中的植物概念系統及其內容有許多地方與現代人預期的植物概念不同：到底這本書是用怎樣的概念架構起來的？編纂者如何界定概念的內容範圍？我們應如何了解書中的概念架構而得以了解、使用它呢？

1.2 斷代與領域知識本體研究

在研究《廣群芳譜》之前，我們已經分別建立了《唐詩三百首》及《蘇軾詩》¹的知識本體。此兩知識本體皆以文本中之植物、動物、人造物等三項為研究標的，前者研究唐朝，後者研究宋朝，皆屬斷代研究。相較於二者，《廣群芳譜》收錄的植物概念跨

越朝代，打破過去斷面的研究，而形成時間軸上縱的研究。這項意義豐富了《廣群芳譜》的研究價值。

2. 研究方法

2.1 『莎翁花園』研究法(Shakespearean garden approach)

莎翁花園意即蒐羅、種植莎翁作品中提及的植物所建成的花園，這種作法常見於西方博物館。莎翁花園的建立不僅幫助人們更容易瞭解莎翁作品的涵意，更有助於一窺莎翁時代英國植物的概貌，後者遂成為我們借用莎翁花園研究法建構領域知識本體的導因。

2.2 實際步驟

我們研究的方法是，首先建立一個《廣群芳譜》的線上詞彙資料庫，以《廣群芳譜》中的植物詞彙為基礎，結合中央研究院中英雙語知識本體詞網（簡稱研究院知識詞網，Sinica Bow）²，將語言資訊連結到上層共用知識本體的概念架構上，利用此知識本體的架構，進一步分析《廣群芳譜》的知識內容。同時，我們也嘗試利用此上層知識本體的架構，構建屬於《廣群芳譜》的領域知識本體³，以作為跨時代領域知識本體建立的參考、研究。

3. 簡介廣群芳譜

3.1. 《廣群芳譜》的成書背景

西元1708年出版的《廣群芳譜》，是清朝康熙皇帝內閣學士汪灝等人以明朝王象晉《群芳譜》⁴為基礎

² 中央研究院中英雙語知識本體詞網 (The Academia Sinica Bilingual Ontological Wordnet, Sinica BOW)，是以詞彙語意關係 (Wordnet) 結合知識本體架構 (Suggested Upper Merged Ontology, SUMO) 所開發的多功能檢索與應用系統。目前系統介面雛型開放的功能除了提供中英語雙語查詢、詞彙語意關係的檢索，還可進一步將語言資訊連結到上層共用知識本體的概念架構上，利用此知識本體的架構，可對知識內容做分類，與簡單推理，例如：蜘蛛為節肢動物，船為運輸工具等。網址為：<http://bow.sinica.edu.tw>

³ 建立領域知識本體的具體做法：1. 建立文本詞彙資料庫 2. 將詞彙庫結合研究院知識詞網，Sinica Bow。相關領域知識本體構建包括：魚類知識本體、唐詩三百首知識本體、蘇軾詩知識本體構建，請見研究院知識詞網網頁。

⁴ 《群芳譜》是十七世紀初的一部植物學兼農學巨著，體例沿

¹ 請見<http://bow.sinica.edu.tw/ont/>

加以改編而成。康熙四十七年成書，定名為《御定佩文齋廣群芳譜》。《廣群芳譜》致力於蒐集、考證歷代文獻上出現過的植物資料，以便於後人檢索、研究。《廣群芳譜》依清朝皇室所藏圖書，經過汪灝等人集體對原書的錯誤加以更正，又刪去冗文，增加許多新的內容⁵，根據研究，《廣群芳譜》是一本具有文藝性質的農園藝著作，對於農學與植物學都有很大的影響。

3.2 《廣群芳譜》的體例和內容

《廣群芳譜》是一本有關植物的譜錄，全書共八十卷，除天時譜「記四時長養之理、萬彙榮枯之候」之外，分為穀譜、桑麻譜、蔬譜、茶譜、竹譜、花譜、果譜、木譜、卉譜、藥譜十譜。所記載的都是中國文獻上有紀錄的植物資料，並彙整、考證多家的資料，包括：經、史、傳記、方書、地志、詩經、爾雅、歷代詩文等。每類後都附有歷代詩詞等文藝作品或相關典故。篇幅長短不一，涉及知識包羅萬象，包括有神話、傳說、藝文、園藝、飲食、醫學等。

除茶譜、竹譜外，每譜之下有各種譜名。譜名就是植物的名稱。開頭概括說明植物特徵、外觀，然後在每一大類中又分為若干種。另外，譜名有時分見於兩處說明：例如，梅、杏、桃、李之類在花譜、果譜兩處都有。花譜著重於「花」或「枝幹」的形狀、大小、顏色的描述，果譜則對「果實」的著墨較多。以下舉例說明：

例：卷第二十二 ● 花譜

梅花 [梅實別見果譜]

原澆以塘水則茂忌肥水種類不一白者有綠萼梅[凡梅花附蒂皆絳紫色惟此純綠枝硬亦青特為清高…]增 范成大梅譜 …早梅[花勝直腳梅吳中春晚二月始爛漫獨此品于冬至前已開故得早名錢塘湖上亦有一種…]官城梅[吳下圃人以直腳梅擇他本花肥者接之遂敷映唐所謂官梅止謂在官府園圃中非此官城梅也]

例：卷第五十四 ● 果譜

梅 [梅花別見花譜]

原…赤者材堅白者材脆種類不一白者有綠萼梅[實大五月熟]…冠城梅[實甚大五月熟]…早梅[四月熟]

3.3. 《廣群芳譜》的植物名稱

《廣群芳譜》沒有泛稱的譜名，如：草、花、木、穀、瓜等。植物有時以特定名稱出現，如花譜中的巖桂、春桂；有時卻不特別指出是何種植物，如：木譜中有譜名「桐」但沒有「梧桐」，在「桐」中說明「梧桐」是「桐」的一種，後面這一類的植物

名稱還有麥、稻、芝等等。

另外，《廣群芳譜》在每種植物後都記述有多種別名，如芍藥又稱解倉、沒骨花；羊桃又稱鬼桃、羊腸。須注意的是，有時名稱相同，指涉的植物卻大不相同，如穀譜的赤小豆和藥譜的相思子別名都是「紅豆」，而「蘭」一詞可能是「蘭花」也可能是「蘭草」。以上所提到的這些植物名稱和別名，有的反映了歷史的變化，有的顯示出地理的方言差異，有的是由典故演變而來，說明詞彙本身的資訊可能包含三種以上的面向，這些將來都會嘗試在《廣群芳譜》的線上詞彙知識庫中顯示出來。

4. 廣群芳譜的植物分類體系

4.1 十個譜

前面介紹過，《廣群芳譜》用了十個譜來給植物作分類，這十個譜大致上區分了植物的自然屬性和親緣關係，但經過研究觀察，發現這十個譜，書中都沒有說明其定義和內容，有些譜的內容分類甚至不太嚴謹，在一定程度上反映了《廣群芳譜》編纂者的觀察角度，可能不是以純植物生長類型的觀點出發。下面提出一些觀察：

- (1) 特定的譜之下的植物，有時不一定只屬於這個譜，有時也可歸到其他譜，例如分到卉譜的「苻」自古以來在中國就是一道名菜⁶，應該也可分到「蔬譜」。此外，花、卉、蔬雖分作三譜，然花中有卉(蘭蕙的附錄有蘭草、蕙草)，卉中有花(菖蒲)，卉中有蔬(苻)，該分到何譜？為何分到該譜？《廣群芳譜》沒有說明其分類標準。
- (2) 譜名分見於兩處說明的有花譜、果譜以及木譜、果譜等，例如，梅、杏、桃、李之類在花譜、果譜兩處都有。但也有例外，如「栗」、「榛」、「椰子」只分配到果譜，木譜的地方查不到。
- (3) 筍放在竹譜之下作為附錄說明，不另外分配到蔬譜。「筍」也是竹，卻歸到卉譜。
- (4) 譜名之下說明各種類。如：「桐」有三種：青(梧桐)、赤、白(泡桐)三種，此三種不另外別列譜名，此三者之外的其他「桐」則別列譜名，如岡桐(油桐)、刺桐、臭桐、胡桐、折桐等；有時譜名之下說明的種類會別列譜名，如：「垣衣」是「苔」的一種(廣群芳譜：苔在牆曰垣衣)，《廣群芳譜》在譜名「苔」之外，又別列譜名「垣衣」。

4.2 細部概念區分

《廣群芳譜》在分類的概念上，除了用十個譜來區分外，在每個譜下，還針對植物的性狀或功能做進一步的細分。如卉譜有靈草(如：芝)、水草(如：蘋)、芳草(如：鴛鴦草)、香草(如：茅草)、染草(如：茜草)、隰草(如：蓬)等；此外，花譜中有草本(如：石竹)、木本(如：杜鵑)、蔓本(如：薔薇)，木譜中有

用宋朝陳景沂《全芳備祖》，在國內外植物學界有很大影響，但《群芳譜》也有不足之處，《四庫全書》提要說明：是所謂「略於種植而詳於治療之法與典故藝文」。《群芳譜》以討論植物為主，兼論金魚和鶴。全書共 28 卷，分為天譜、歲譜、穀譜、蔬譜、果譜、茶竹譜、桑麻葛譜、藥譜、木譜、花譜、卉譜和鶴魚譜等。

⁵用「增」字表新增內容，「原」字表為《群芳譜》原文。

⁶見《詩經植物圖鑒》，潘富俊著作，2001，貓頭鷹出版社。

灌木(如：棘)、藤屬(如：女蘿)等。以上種種概念，有的清楚明白，有的模糊不易分辨，有的看起來簡單易懂，實際上卻和我們的認知不同，例如「果譜」中有一個「水果」的概念，指的是如荔枝、柑、橙一類多汁的果類，而西瓜、甜瓜雖然水分也很多，概念上被歸類為「蔬果」，也就是指草本或蔓生植物所結的果實⁷，而蘋果、葡萄呢？他們是屬於皮薄的「膚果」。可見，如果我們僅用現代的詞彙觀點來理解《廣群芳譜》的分類概念是不夠的，也容易錯誤。

5. 廣群芳譜對應到 SUMO 的概念

5.1. 何謂 SUMO？SUMO 的分類概念有哪些？

SUMO是Suggested Upper Merged Ontology 的簡稱，中文翻作「建議上層共用知識本體」，由IEEE標準上層知識本體工作小組所建置。這工作小組的目的是發展標準的上層知識本體，以促進資料互通性、資訊搜尋和檢索、自動推理和自然語言處理。上層的知識本體提出一個涵蓋範圍廣闊的領域區域，希望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⁸。

那麼，SUMO 的分類概念有哪些呢？SUMO 所提出的概念很廣泛，可以說涵蓋了人類所有的知識概念，不過它的概念是較概括的、一般的，因此可以容納較特定的領域知識本體。就植物的概念來說，SUMO 僅有「開花植物」、「非開花植物」兩大類，在「非開花植物」的概念節點上，才分有藻類、真菌、蘚類、蕨類四個小類。

5.2. 中國典籍如何對應到 SUMO：應用範例

前面第三節的部分，我們看到，17世紀完成的《廣群芳譜》在知識系統分類上有許多地方不易掌握，主要是因為(1) 作為分類用的概念本身沒有清楚的定義 (2) 作為概念表徵的詞彙與現代理解的概念不同。

將《廣群芳譜》的分類系統對應到 SUMO 的概念架構上，也就是將領域知識系統連結到上層共用知識系統的一種嘗試。我們拿《廣群芳譜》的分類系統和 SUMO 比一比，發現《廣群芳譜》的概念多又複雜，SUMO 的概念少而簡單，這樣兩個性質不同的概念架構，透過詞彙對應的結果，所有《廣群芳譜》的概念都連結到 SUMO 的概念架構上了。請見下面樹圖舉例說明：(粗字體為 SUMO 概念)

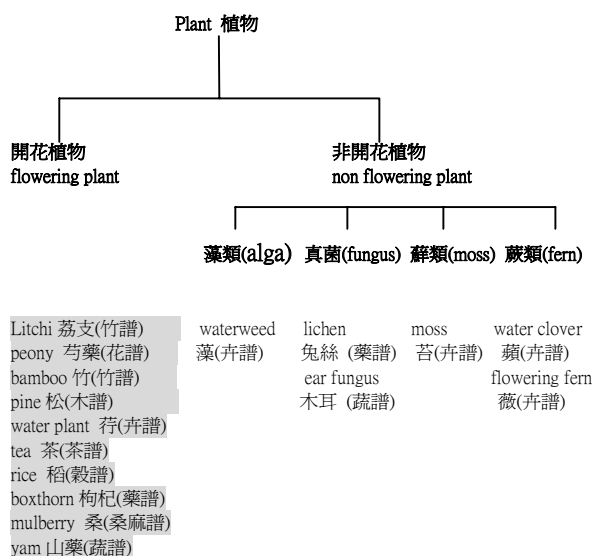


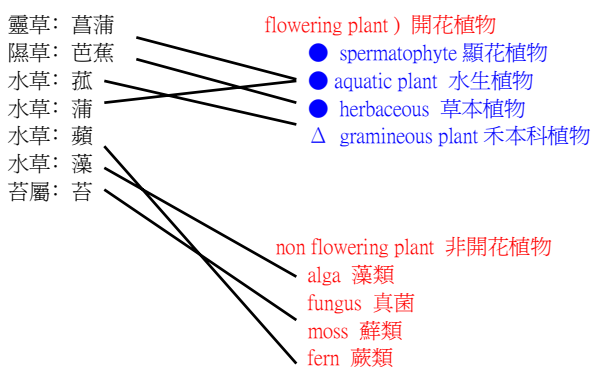
圖 1 廣群芳譜與 sumo 概念之連結

從上圖，我們看到 SUMO「開花植物」的概念遍及《廣群芳譜》十個譜，也就是說，《廣群芳譜》每一個譜都有植物是屬於 SUMO 的開花植物。只有藥譜、卉譜中的某些植物會連到 SUMO「非開花植物」的概念。

如果我們將焦點集中在卉譜的植物詞彙，將其連結到 SUMO 的概念架構上，並加入 WordNet⁹ 的語意關係，則會得到下面的圖：(因篇幅有限，僅舉例說明)

字體顏色說明：廣群芳譜概念+植物詞彙，SUMO，WordNet 語意關係

卉譜



⁷ 參考教育部國語會國語辭典解釋。

⁸ 此段說明摘錄自研究院詞網知識本體網頁 (<http://bow.sinica.edu.tw/ont/>)，目前 SUMO 已經和英語詞彙網路 WordNet1.6 版本作初步的連結。網站：<http://www.ontologyportal.org/>

⁹ English WordNet，這裡簡稱 WordNet，是一個以詞彙語意關係做基礎所開發的單語詞彙知識資料庫，由普林斯頓大學開發擁有，授權中央研究院研究使用。網站請見：<http://wordnet.princeton.edu/>

圖 2 卉譜、SUMO 知識本體與 Wordnet 語意關係

上圖結果顯示,《廣群芳譜》詞彙對應到SUMO、WordNet概念架構的結果,使得《廣群芳譜》「卉譜」的知識內容更清楚明白:它包含有開花植物的水生植物、草本植物、禾本科植物,以及非開花植物的藻類、蘚類,及蕨類植物¹⁰。這樣的概念架構和結果,也為文本的進一步比較研究提供了很好的基礎。

6. 初步觀察結果

透過觀察 SUMO 和《廣群芳譜》概念體系結合的情形,我們更容易掌握《廣群芳譜》的知識系統。下面提出一些觀察結果:

- (1) 『筍』的概念:前面提到筍放在竹譜之下作為附錄說明,不另外分配到蔬譜。《廣群芳譜》記載:爾雅 筍則竹之初生者故曰筍竹萌也可以為菜穀 說文 筍竹胎也。“bamboo shoot”在 SUMO 的概念是”fruit or vegetable”,但由《廣群芳譜》分類的概念來看,「筍」的概念比較接近「生長growth」或「未完全形成的non fully formed」。
- (2) 『開花植物』的定義:花譜中分有草本(如:石竹)、木本(如:杜鵑)、蔓本(如:薔薇)的區別,草本、木本植物都是開花植物,所以花譜中包含這兩者應是可理解的。但也有些會開花的草本植物會被歸入卉譜。可見並不是會開花的植物就可歸入「花譜」。而 SUMO 中的「開花植物(flowering plant)」的定義範圍顯然比《廣群芳譜》的「花譜」大的多。
- (3) 從人的觀點出發:《廣群芳譜》實際上對植物做譜的分類時,不傾向以植物生長型態作分類,而偏向以人的實用性觀點作分類,如:可以觀賞的歸入花譜,可治病養生是藥譜,可食用的算作蔬譜的一類。至於花則有異花,果有異果,木有異木,這種分法也可說是從人(而非植物本身)的觀點出發。



圖 3 廣群芳譜線上詞彙知識庫查詢介面

7. 未來工作

《廣群芳譜》線上詞彙知識庫,目前提供廣群芳譜譜名、植物詞彙及別名、同義詞林同義詞的檢索,以及教育部國語會國語辭典的查詢功能(見圖三)。未來將考慮增加中英語雙語檢索、內部知識查詢(如:詩文檢索、原文出處等),並連接植物學資料庫,增加圖片和相關的植物學知識。

除此之外,《廣群芳譜》的知識本體建置也是未來要努力的目標。完整的知識本體建構有賴詞彙知識庫的完整性與正確性,一旦建置完成,可以作為跨時代領域知識本體建立的參考、研究,這對於知識的運籌來說是很有意義的。

參考文獻

- [1] 潘富俊.(2001). 詩經植物圖鑒。
- [2] 潘富俊.(2001). 唐詩植物圖鑒。
- [3] 潘富俊.(2001). 楚辭植物圖鑒。
- [4] 黃居仁、羅鳳珠、張如瑩、張舒茗.(2004). 研究院知識詞網與唐詩三百首一雙語知識本體詞網簡介及唐詩知識本體之初步構建。
- [5] 游修齡.(1999). 清代農學的成就和問題, 農史研究文集, 中國農業出版社。
- [6] 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>

¹⁰更深入的研究尚進行中,這裡舉例的僅是廣群芳譜詞彙知識庫的小小部分。