【2019 佛教藏經會議專稿】

以資料鏈結發展智慧時代佛教經典數位研究資源

洪振洲

法鼓文理學院佛教學系副教授

摘要

佛教於二世紀傳入中國之後，在中土流傳、傳播的過程中，產生了大量的漢譯經典。這些經典經過後世多次集結與整理，成為今日漢傳佛教最重要的文獻集合，也就是所謂的「漢文大藏經」。自 1980 年代開始，由中華電子佛典協會所起始的漢文電子佛典的製作工作，不僅創造了廣為學術研究界使用的「CBETA 電子佛典集成」，更引領台灣佛學研究界陸續產出極具應用價值的佛學數位典藏。這些結合佛學知識與數位技術的成果，不僅達到利用現在數位技術，長久保存重要佛學資料的目的，也更進一步發展出便利佛學研究應用的電子資源。然而可惜的是，這些工具還是以處理文獻表面的文字資訊為主，並未進一步著手處理隱藏在文獻背後的深層意義。因此，目前人文學者實際在應用這些工具時，還是相當倚賴於關鍵詞搜尋。但這些單一而零碎的搜尋結果，仍需由研究者逐項解讀、判斷與比對，才能整合組成具有關連意義的資料。近年來，隨著人工智慧技術的發展，電腦機器於正確處理人類語言的工作上，已經取得突破性的進步。但詳究其技術內涵，仍然是以數位統計為基礎的

推論成果，並非真實理解文字背後的意義概念。也因此，如要利用人工智慧技術，讓電腦系統能理解佛典文獻內容，並使其擁有進行內容分析與歸納的能力，我們並非只要關注人工智慧的演算方式，而是需要提供大量能夠使機器閱讀、理解的知識條目資料，作為電腦系統進行知識推論時的根據。為解決這個問題，全球資訊網（World Wide Web）發明者提姆・柏納－李（Tim Berners-Lee），呼籲資訊界重視「語意網」（Sematic Web）與「鏈結資料」（Linked Data）的概念，以便製作讓電腦機器也可以理解與進行推論的資料網絡。而綜觀目前的佛教數位系統，雖然大部分的系統與資料內容皆十分開放，但是系統間少有互相鏈結的狀態。因此本研究嘗試以資料鏈結方式來進行現有佛學數位資料的整合，並嘗試進一步以資料鏈結的概念，來發展人工智慧時代佛教經典數位研究資源。

一、前言

佛教於二世紀傳入中國之後，在中土流傳、傳播的過程中，也興起了長達千年的佛經翻譯熱潮，完成了大量佛典的漢譯。這些漢譯佛教文獻，經過後世多次集結，逐步形成文字記錄，成為我們今日所見的「漢文大藏經」。現今佛教研究者，無不以大藏經為主要研究文獻。今日我們進入了數位時代，數位工具技術之發展一日千里，自 1980 年代開始，由中華電子佛典協會所起始的佛典數位化的風潮，不僅創造了廣為漢傳佛教研究界使用的「CBETA 電子佛典集成」，也引領台灣佛學研究界陸續產出極具應用與參考價值的佛學數位典藏成果。[1]這些結合佛學知識與數位技術的成果，不僅讓這些重要佛學資料，得以利用現代數位技術，達到長久保存的目的，也更進一步發展出便利佛學研究應用的各式電子資源，使得研究者得以利用數位方法來更進一步探索佛教經典的內涵，尋找突破傳統研究瓶頸的契機。[2]有鑑於各式佛學電子資源雖然好用，但缺乏良好整合的問題，法鼓文理學院團隊於 2013 年開始著手建置漢籍佛典數位研究平台，並於 2016 年首度展示該計畫的成果——「CBETA 數位研究平台」。[3]該平台主要以大藏經資料為基底，建設出：（一）易於操作使用的 CBETA 經典線上閱覽界面 CBETA online、[4]（二）用於分析藏經詞彙現象的 CBETA 詞彙分析平台，[5]與（三）今年甫公開

1 參杜正民，〈佛學數位資源的建置與開展〉，《法鼓佛學學報》10，2012 年，頁 147-210。

2 參洪振洲，〈由資料庫到數位研究平台——談佛典文獻數位研究工具之發展與演變〉，《漢學研究通訊》35：1，2016 年，頁 1-14。

3 參洪振洲，〈數位時代漢譯佛典之研究利器——CBETA 數位研究平臺〉，《數位典藏與數位人文》1，2018 年，頁 149-174。

4 法鼓文理學院，數位計畫專案小組，「CBETA 線上閱讀」網頁，見 http://CBETAonline.dila.edu.tw，2019/7/1。

5 法鼓文理學院，數位計畫專案小組，「CBETA 詞彙搜尋與分析」網頁，見

的 DEDU 對讀資料編輯平台。[6]整體來說，這個平台是希望以「資料閱讀」、「知識編輯」、「內容分析」等角度提供研究者新一代的佛學研究資源與研究工具。這個平台的建置，確實讓研究者在進行佛學研究資料探索時有了更便利的工具。但實際上，該工具所提供的功能，還是侷限於提供文獻內容字面上的顯示、搜尋功能為主，並未進一步著手處理隱藏在文獻背後的深層意義。因此，此工具的主要用途，仍僅限於資料查找與少數使用字面統計分析就可以回答的問題。這樣的結果導致人文學者在實際上應用這些數位工具進行研究工作時，還是非常倚賴於使用關鍵詞搜尋到各種單一而零碎的訊息，後續再將這些碎片化的資料來源，逐項解讀、判斷與比對，才能組成具有關連意義的資料。但現今大量資料在網際網路湧現，且在不同時代、不同領域、不同用語的情況下，這樣的問題將會變得更嚴重。以關鍵詞搜尋網絡，不僅耗費大量人工時間，也不免出現毫無意義或重複訊息、乃至天差地別的錯誤結果。對於人文研究者來說，自然希望數位系統能夠提供更精確的搜尋結果，甚或能夠剖析文獻內容的意義內涵，減低在研究過程中所需的蒐集與整合相關資料的繁複作業，讓研究者能夠更專注於研究問題之上。

近幾年來，隨著人工智慧在處理人類語言、影像方面等技術的顯著突破，越來越多具有初步人工智慧的產品出現於世上。這樣的技術發展，也似乎讓我們在古文獻的處理上，看見一道曙光。因此，我們應該開始思考，如何利用人工智慧技術，讓電腦系統能夠具有理解佛典文獻內容、並進一步具有進行分析與歸納的能力。要製作這樣的系統，並非只要關注人工智慧的演算方式，實際上目前所發展的人工智慧技術，仍然是以數位統計為基礎的推論成果，尚無法真實理解文字背後的深層意

http://CBETAConcordance.dila.ed.tw，2019/7/1。

6 法鼓文理學院，數位計畫專案小組，「DEDU 對讀文獻製作工具」網頁，見 http://DEDU.dila.edu.tw，2019/7/1。

義概念。也因此，如要利用人工智慧技術，讓電腦系統能理解佛典文獻內容，並使其擁有內容分析與歸納的能力，我們並非只要關注人工智慧的演算方式，而實際上仍然需要提供大量能夠使機器閱讀、理解的知識條目資料，作為電腦系統進行知識推論時的根據。為解決這個問題，全球資訊網（World Wide Web）發明者提姆・柏納－李（Tim Berners-Lee）2001 年在《科學人》雜誌專文宣告「語意網」（Sematic Web）將取代全球資訊網，並於 2009 年 TED talk 以「The next web」（下一代的網路）為題，呼籲資訊界重視「語意網」與「鏈結資料」（Linked Data），以推進網絡的發展方向。「語意網」的概念，就是希望網際網路上的資料，並非只考慮到人類閱讀的方便，而是也應該提供機器可以閱讀與理解的資料，並且相關的資料必須互相連結，成為一個「資料的網絡」（Web of data）。透過「語意網」建構發展，資料可被電腦有效讀取後，電腦可以自行判斷資料的相關性。近期軟體巨擘微軟（Microsoft）也發表了「Microsoft Concept Graph」，其內容紀錄超過 540 萬條核心概念，與 8400 萬筆的概念關係資訊，其目的就是要讓下一代的人工智慧系統，藉此能理解人類的知識體系，進而整合不同網絡的訊息，讓電腦系統能達到如同人類具有理解與推論的能力。[7]

而綜觀目前的佛教數位資源系統，雖然大部分的系統皆十分開放，且也有部份系統考慮到提供結構化的資料，甚至是便利的「開放程式存取界面」（API）來提供機器可理解且方便存取的資料模式。但是大部份的數位資源，其內容仍以單一資料類型為主要設計目標，系統間且少有互相鏈結的狀態。[8]這主要是因為，每個資料庫各有其不同的設計目的

[7] 請參考：MicroSoft Concept Graph, https://concept.research.microsoft.com, 2017/2/4。

[8] CBETAonline 平台可說是提供最多資料整合範例之處，在該平台中，除全文資料外，也整合了相當完整的經典背景資料。但是對於辭典資料、研究書目、人物時間地點資料等，便是以簡單的「轉交查詢」的方式在系統中呈現相關資料，

與不同的資料涵蓋範圍，這些因素造成各資料庫間的資料異質性高、範圍差異大，所以往往僅能達到部份整合或無整合的狀態。而鏈結資料的概念，本身提供了一個相當方便的資料連結機制。因此本研究希望以資料鏈結方式來進行現有佛學數位資料的整合，並進一步以資料鏈結的概念，來發展人工智慧時代佛教經典數位研究資源。我們希望能以此初步成果為基礎，期望透過創造、共享與應用更多互相鏈結的語意資料，提供人文學者更具前瞻性與突破性的數位研究平台服務。

二、資料鏈結技術與相關應用

「鏈結資料」一詞，是由 Tim Berners-Lee 於 2006 年 W3C 之 Design Issue 中所提出的概念。[9]其主要精神是藉由一個簡單的設計，建立開放的、外顯的、容易理解的方式來描述資料間的關係，最終目的是要將現有的數位資料串連，以產生一個讓電腦能夠直接或間接處理的資料網，幫助電腦正確解讀人類的知識內容。更細節的來說，鏈結資料資料的概念，是由以下的元素所組成：

（一）資源必須具有唯一 URI（Uniform Resource Identifier，統一資源識別碼），用以表達與存取資料

（二）資源的存取方式是透過 HTTP 協定（Hyper Text Transfer Protocol，超連結傳送）與該資源的 URI 而達成。

（三）當使用者存取該資源的資料時，應該透過使用開放性的標準，例如 RDF 表示方式、SPARQL 查詢語言，來提供與該資源相關的資料內容。

並未有完整整合。

9 Tim Berners-Lee, “Linked Data”, Design Issues. W3C (2006), 2010/12/18.

（四）當資料發佈於網路上時，必須提供與網路上其他資料集合的連結。

鏈結資料的概念，已經被 W3C 正式收錄為網路資料建置的標準項目之一。近年來也陸續訂定了許多與鏈結資料相關的網路標準。[10]由於鏈結資料概念具有相當的突破性、重要性且實做上並不困難，因此在此一概念近年來得到許多大型機構的支援，陸續推出具有鏈結資料的概念與相關服務。同時間，近年來於各國發展火熱的開放政府資料之計劃，也多採用鏈結資料模式來進行。以下簡列一些網路上的大型的鏈結資料應用專案。

（一）DB Pedia 專案，由位於德國萊比錫的自由大學的團隊於 2007 年建立，網站上所有資料都利用開放式鏈結資料形式來提供。[11]其內容主要是將維基百科（Wikipedia）的內容經過電腦系統進行自動關聯匹配與連結所產生。網站除提供線上服務外，也提供資料下載。在其正式發行的資料版本 2016-04 內，共提供約 4233000 個概念的資料內容，且彼此間具有相當綿密的聯結。

（二）紐約時報（*The New York Times*）於 2010 年發表了含有約略 10000 筆標籤（Tag）的鏈結資料集，並發表相關個人化擴充服務、個人化標記與連結服務。[12]現今該個人化服務已經中止，但相關的鏈結資料

[10] 參閱 W3C-ALL STANDARDS AND DRAFTS, https://www.w3.org/standards/techs/linkeddata#w3c_all, 2019/7/4。

[11] 參 Sören Auer, Christian Bizer, Georgi Kobilarov, et al., DBpedia: A Nucleus for a Web of Open Data The Semantic Web, In *The Semantic Web. ISWC 2007, ASWC 2007. Lecture Notes in Computer Science* 4825, eds Aberer K. et al. (Berlin, Heidelberg: Springer, 2007), https://doi.org/10.1007/978-3-540-76298-0_52。

[12] 參閱：Build Your Own NYT Linked Data Application, https://open.blogs.nytimes.com/2010/03/30/build-your-own-nyt-linked-data-application/?_r=1, 2019/7/4。

已轉換至 data.io 網站，繼續提供下載服務。[13]

（三）英國廣播公司（BBC）在 2012 年，以開放式鏈結資料為基礎，打造 2012 年的奧林匹克網站，並在 2013 年發表鏈結資料平台，用來聯結 BBC 的各項資料。[14]而自 2014 起，BBC 成立了 ontology 服務，用以持續發展鏈結資料與相關服務。[15]

（四）European Union Open data Portal，為歐盟的開放資料整合網站。[16]其內容提供歐盟政府相關的開放資料集合下載，並內容互相連結。目前已經整合了約三十個歐盟的政府機構。

（五）Wikidata 專案是由維基媒體基金會（Wikimedia Foundation）於 2012 年 10 月所啟動的資料建置計畫。[17]其原本的目的是將維基百科內的對於各種不同語言的條目結合在同一個資料庫的條目中，並給予一個唯一的編號（因為編號都是以 Q 開頭，因此稱為 QID），用以將維基百科內所紀錄的概念，匯集出一個比較有結構文件資料庫。此一項目目前已經變成維基百科內提供每個實體進一步解釋的資料來源。不僅提供 RDF 格式之資料下載，並於 2015 年開始提供 SPARQL 的查詢界面，供各界進行資料查詢與連結。

（六）中央研究院數位文化中心鏈結開放資料平台，由中央研究院數位文化中心，利用中研院歷年來台灣數位典藏計畫所完成的資料內

---

13 New York Times - Linked Open Data , https://datahub.io/dataset/nytimes-linked-open-data, 2019/7/4.

14 Oliver Bartlett, Linked Data: Connecting together the BBC's Online Content, http://www.bbc.co.uk/blogs/internet/entries/af6b613e-6935-3165-93ca-9319e1887858, 2019/07/04.

15 詳見：BBC – ontologies, http://www.bbc.co.uk/ontologies, 2019/7/4。

16 詳見：EU Open Data Portal, https://data.europa.eu/euodp/en/linked-data, 2019/7/4。

17 詳見：Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page, 2019/7/4。

容，與中研院各研究所多年來製作的數位資料，進行整理與 LOD 資料轉置。[18]資料類型包含圖片、影像、聲音、文字，類別則涵蓋生物、人類學、宗教、藝術、影音、歷史等各種項目。截至 2019 年七月為止，共提供超過十一萬六千條資料。同時也提供資料下載與 SPARQL 的資料查詢界面。

（七）上海圖書館開放數據平台，內容包含人名、姓氏、歷史紀年、地理名詞、機構名錄、印章、避諱字等內容，並且資料內容秉持相當開放的態度，除了 RDF 資料下載、SPARQL 查詢服務外，也提供 REST API 的方式讓第三方能進行資料介接。[19]

（八）Buddhist Universal Digital Archive（BUDA），這是由 Buddhist Digital Resource Center（佛學數位資源研究中心，以下簡稱 BDRC）所製作的鏈結開放資料服務。[20]由於 BDRC 專注的資料項目在於西藏佛教文獻資料的整理，因此這個資料庫的內容涵蓋藏傳佛教相關的人物、地點與文獻資的描述。[21]

相對於鏈結資料概念的被接受與大量運用，後續利用鏈結資料開發之應用程式，就仍待各方積極努力。不過現今仍有不少讓人眼睛為之一

---

18 詳見：中央研究院數位文化中心，「鏈結開放資料平台」，http://data.ascdc.tw/, 2019/7/4。

19 詳見：上海圖書館數據開放平台，http://data.library.sh.cn/index, 2019/7/4。

20 詳見：BUDA—linked data, https://www.buddhistarchive.org/linked-data-tool, 2019/7/4。

21 實際上，BUDA 所製作的鏈結資料庫的內容與本研究欲發展的漢傳佛典連結資料庫有高度相關性。為避免重複工作，並且能達到資料互通之利，因此我們與該組織合作，並利用用該組織並將所發展出來的藏傳佛教文獻知識本體（BDRC Ontlogy）為基礎，並加以修正，使之能夠用來承載漢傳佛典連結資料。詳細本體樣貌，請參考第四小節內容。

亮的應用程式陸續出現，以下簡列一些相關應用：

1. Google Knowledge Graph：在 google 搜尋「名詞」時，右方會出現的一個「名詞說明欄」，而其背後的技術，便是 Google 利用處理連結資料而達成，稱為「Google Knowledge Graph」的知識搜尋技術。[22]

2. RelFinder，http://www.visualdataweb.org/relfinder.php，用來找出兩個概念間的隱藏關係，並且視覺化，使用者可以利用這個簡易應用，找出兩個看似不相關的實體之間的隱藏關聯。

3. ImageSnippets，http://www.imagesnippets.com，主要是讓使用者上傳相片，並提供可於線上編輯該照片的詮釋資料（metadata）。此外，在其詮釋資料的設計內，納入含有連結至 DBpedia 的開放鏈結資料的概念。所以使用者在編輯完成後，相關資料就進入與全世界鏈結的資料集合內，也可以藉著這些連結資料找到其他相關照片。

4. Quepy，http://quepy.machinalis.com，這個開發應用實際上是個程式語言的函式庫，並非最終產品。其目的是讓使用者可以利用語意問句來與資料庫互動。在它的範例之中，你可以問問「Who is the president of Argentina?」它會透過後端的鏈結資料庫來回答。

5. dataTXT semantic text API，https://dandelion.eu/semantic-text/entity-extraction-demo/，一組文句剖析服務。它利用比對後端的語意鏈結資料，來比對分析出文句中的語意單元。

6. Music Genre Map，http://www.bradybutterfield.com/musicGenreFDG/，將 DBpedia 內音樂類型實體資料關係視覺化的一個專案網站。

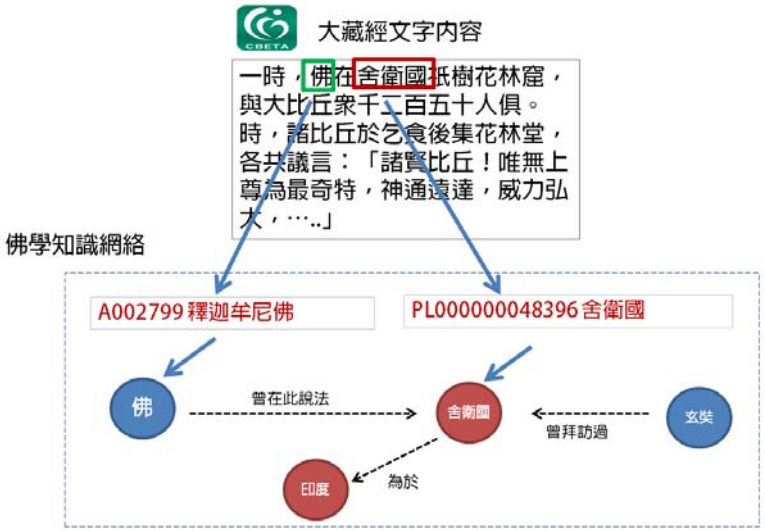
[22] 詳見：Google, Introducing the Knowledge Graph: things, not strings, https://googleblog.blogspot.tw/2012/05/introducing-knowledge-graph-things-not.html, 2019/7/4。

7. Linked Taiwan Artister，由中研院數位文化小組所製作。利用該所所建立的鏈結資料，提供一個完整探索台灣藝術家生平與作品的網站。在該網站中的所有概念實體間皆有相互連結，無論使用者由何角度切入，皆可以連結到其他相關資源。[23]

三、人工智慧時代佛教經典數位研究資源

本研究嘗試以資料鏈結的概念，來發展人工智慧時代佛教經典數位研究資源，除利用鏈結資料技術嘗試建立可以匯集各種與佛學研究相關知識的佛學知識網絡外，也由於佛學研究仍是以大藏經的內容作為最主要的研究資料來源，因此我們也同時以大藏經文字內容為核心，將知識網絡的節點，鏈結於大藏經文字之上，提供大藏經文字更多的解釋內容。因此，本研究所提之人工智慧時代佛教經典數位研究資源，其結構如下圖一所示。

[23] 詳見：中央研究院數位文化中心，“linked Taiwan Artists”, http://linkedart.ascdc.tw/index.php, 2019/7/4。



圖一　人工智慧時代佛教經典數位研究資源之結構設計示意

這個佛學知識網絡，簡單來說，就是要將佛教重要參考資料規範化、網絡化的一個重要知識結構。並且在完成這樣的知識網絡後，將其於大藏經進行結合，如此將可能表達文字背後更多的意涵，未來以人工智慧演算方式進行相關推論分析時，我們將有更多的知識可以利用。因此，我們也將進行大藏經文字內容的比對，找出文字內容與佛典知識庫對應的部份，鏈結至建立的佛學知識網絡中。在這樣的工作規劃中，我們實際上必須克服三個困難，包含有：

（一）如何應用鏈結資料技術來建立佛學概念的鏈結知識網絡。

（二）如何蒐集與製作更多相關的佛學概念資料，以完善佛學知識網絡。

（三）如何結合大藏經文字資料與佛學知識網絡。

其中第二點比較屬於經營策略的問題，並不在本論文的討論範圍，而第一點與第三點則較屬於技術上的困難，我們在後續兩小節中，將詳細說明我們的初步嘗試與目前成果。

四、建置佛學知識網絡

佛學知識牽涉甚廣，欲製作一個完整的佛學知識網絡，並非一蹴可幾之事。因此本研究目前所進行之部份，僅是非常初步的嘗試，其目的是要先建立一個可能的雛型，為未來進行更大規模施作的計畫作準備。本研究之佛學知識網絡，並不是先建立一個空白的資料庫，再慢慢累積內容。而是由現有的可用資料集來進行轉換，以便減低製作的難度與所耗費的時間。因此，我們首先選定由法鼓文理學院所製作的佛學規範資料庫，作為轉換佛學知識網絡的首要目標。[24]該資料庫內容可以分為人物、時間、地點、經典目錄等四個資料庫，內含佛教相關人物、曆法資料、中國歷史地名、佛教相關地點與佛教經典書目資料等等。規範資料庫內容是極具結構的條目式資料，每一個資料庫的條目，皆獨立且完整的表示對應實體的完整資訊。此外，每一個規範資料庫的條目，早已被賦予一個唯一且不變的資料規範編號，這樣的設計，與連結資料適用的資料模型十分相符，以此資料庫作為轉換目標，將可收事半功倍之效。

而在這四個資料庫中，我們更進一步選定以內容最為明確且單純的人物規範資料庫開始進行。此人物規範資料庫的內容，並非來自某一個特定資料集合的內容，而是在法鼓文理學院其他製作相關專案的製作過程中，遇到需要查找人物、時間、地點相關內容時，由人工逐一查證資料內容建置而成。[25]規範資料庫專案由 2008 年開始進行，截至 2019 年

[24] 佛學規範資料庫網址為：http://authority.dila.edu.tw, 2019/6/20。

[25] 人物規範資料庫的內容，主要經由資料：高僧傳數位化專案（http://

六月為止，資料庫內容共收錄有 42000 筆人名資料。要進行鏈結資料的轉換，則必須釐清人物規範資料庫內容的所有細節，下表一列出規範資料庫的資料欄位與其意義。

表一　「人物規範資料庫」欄位整理

屬性	意義	補充說明
Authority ID	規範號碼	每筆資料都不同的規範資料編碼
Name	常名	每筆資料在每一種語言會有一個最正式或最被人接受的名稱，稱之為常名。
Alternative Name	別名	其他相關稱呼，也就是所謂的別名。
Gender	性別	
Birth	生年	紀錄之精確度為日。若相關資料可能找出該人物明確生日，則紀錄該日期。例如：0413-05-29。若找出精確生日，則以最精確的文獻紀錄，以「不早於」與「不晚於」的兩個時間點來表達。例如：鳩摩羅什知道應該出生於公元 344 年，但無從得知為何月何日，因此紀錄為：0344-01-01 ~ 0344-12-31。
Death	卒年	紀錄方式與生年欄位相同
Birthdate reference	生年紀錄來源	含說明文字與參考來源
Deathdate reference	生年紀錄來源	含說明文字與參考來源
From	籍貫	出生地，連結至地點規範資料庫
Death place	圓寂地	死亡地點，連結至地點規範資料庫
Category	高僧傳分類	該筆資料主人於歷代高僧傳記中有獨立傳記，則以此欄位紀錄該傳記之分類。

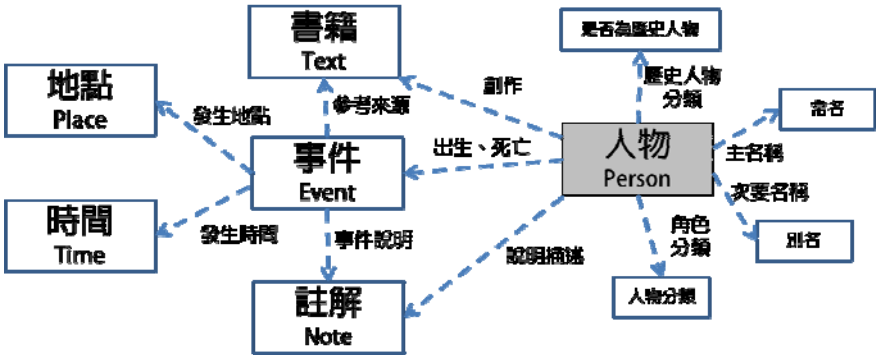
buddhistinformatics.dila.edu.tw/biographies/）、中國佛寺志數位化專案（http://buddhistinformatics.dila.edu.tw/fosizhi/）、《宋高僧傳》之校勘與數位化版本（http://buddhistinformatics.dila.edu.tw/songgaosengzhuan/）、禪宗世系暨禪師名號檢索系統（http://zenlineage.dila.edu.tw/）、法鼓全集（http://ddc.shengyen.org/）等等數位化計畫執行過程中連帶產生。

Histroical Person	是否為歷史人物	用以表達該人物是否為實際存在的歷史人物，或是虛構人物。
Comment	註解	資料相關補充資料。
Teacher	師承	此人物的師承，連結至對應的人物規範資料
Students	弟子	此人物的弟子，連結至對應的人物規範資料庫
Works	作譯資料	此人物的作譯紀錄，連結至書目規範資料庫
Occurs in	提及的文獻	此人物被哪些文獻所提到，連結至書目規範資料庫

（一）建立知識本體

在鏈結資料建立的過程當中，最重要也是最困難的步驟，便是建立一個能夠描述完整資料內容的知識本體（ontology）。[26]在建構知識本體的工作中，最主要的便是識別此知識本體中的概念類別，並標示概念間的關係。由於我們目前僅進行了以人物規範資料的鏈結資料轉換工作，因此我們的知識本體，也是以描述人物規範資料為核心來進行設計，該知識本體的細節如下圖二所示。待未來需要擴增到其他三者後，再來繼續進行知識本體之擴增。

[26] 在此，我們並非意指要產生一個意圖完整描述所有人類知識內容的一般性知識本體（generic ontology），而僅是用來完整涵蓋此資料庫內容的專業領域知識本體（domain specific ontology）。



圖二　以人物規範資料為核心之資料本體概念示意

由圖二之中所表達的知識本體，可以發現到，我們的知識本體利用了事件（event）概念為核心，來串連起人物資料的相關資料。實際上，在表一中所表達的人物規範資料的內容，有許多欄位都與人物出生事件（包含：生年、籍貫、出生資料參考）與死亡事件（卒年、圓寂地、死亡資料參考）相關。因此將這些資訊，歸類至兩大事件內，將可以使整體資料結構變得容易。而人物的其他資訊（包含：常名、別名、人物分類、著作、是否為歷史人物…等），便以人物類別的相關屬性來表達。

（二）挑選描述語彙

在鏈結資料的實體紀錄格式方面，一般選擇以 RDF 規範作為資料的主要表達方式。[27]所謂的 RDF 規範，簡單來說，就是利用所謂的三元組（tripe）：subject（主體／類別）、predicate（謂詞／屬性）、object（客體／屬性值）三部份來進行來進行資料的表達。這三個主要部份，組成為一個完整描述（statement）。其中 subject 與 object 通常會是資料庫內

[27] 請參閱：W3C – RDF, https://www.w3.org/RDF/, 2019/6/20。

的兩個不同物件，而 predicate 就是用來描述兩個物件之間的關係。雖然物件間的關係表達方式，在 RDF 規範中並沒嚴格限定需要用那些詞彙來表達。但根據開放鏈結資料最佳製作實例指引的建議，我們應該要盡量採用別人已經定義過的詞彙，以便增加不同資料集之間的一致性與互通性。[28]在我們廣泛比較目前常見的詞彙規範後，我們決定利用 BDRC 所製作的知識本體綱要（ontology schema）。該資料綱要雖是以藏傳佛教知識為主軸的設計，但因為藏傳佛教資料與漢傳佛教資料內容上有相當關聯，資料結構也十分相近，因此主體可以減低我們在設計資料綱要與挑選描述詞彙上的困難。因此，我們於描述語彙挑選過程中，盡量採用 BDRC 鏈結資料小組進行所發表的規範，若遇相關疑問時，則與該小組進行規範討論。[29]目前在我們所挑選的語彙中，共包含有使用了 rdf、skos、bdo 等等語彙集的相關描述，整理如下表二。

表二　人物規範鏈結資料使用之語彙集

前置詞	命名空間	語彙集名稱
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	RDF 標準語彙
skos	http://www.w3.org/2004/02/skos/core#	簡單知識組織系統語彙 (Simple Knowledge Organization System, SKOS)
bdo	http://purl.bdrc.io/ontology/core/	BDRC 語彙

目前所製作完成的描述語彙內容，整理如下表三。

[28] 請參閱：Best Practices for Publishing Linked Data, https://www.w3.org/TR/ld-bp/, 2019/6/20。

[29] 在本研究中，關於佛學知識網絡的建置，我們特別感謝 BDRC LOD 小組的 Élie Roux 先生的協助。

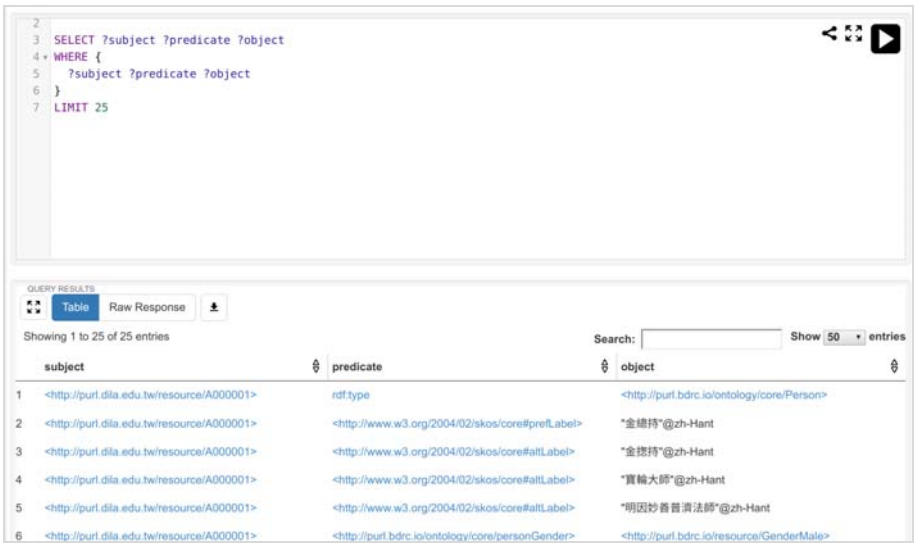
表三 人物規範鏈結資料使用之描述語彙

類別 Class	描述語彙 Vocabulary	註釋 Note
Person（人物）	skos: prefLabel	表示人物的主要名稱
	skos: altLabel	表示人物的次要名稱
	bdo: personGender	表示人物的性別
	bdo: role	表示人物的類別
	bdo: note	表示人物資料的註解
	bdo: personStudentOf	表示該人物的學生
	bdo: personTeacherOf	表示該人物的老師
Person Birth（出生事件）	onYear	出生年代（僅知為那一年時）
	onDate	出生日期
	bdo: note	相關註解
Person Death（死亡事件）	onYear	死亡年代（僅知為那一年時）
	onDate	死亡日期
	bdo: note	相關註解
Note（註解）	bdo: noteText	註解文字
	bdo: noteLocationStatementCBETA	註解來源（以 CBETA 行號紀錄）

（三）網頁系統建置

利用上述的對應資料，我們已經將人物規範資料庫內的內容，初步轉換至符合 LOD 的資料格式，並架設資料相關網站已供運用。[30]在網站內已經初步可以利用 SPARQL 方式進行資料查詢，執行範例如圖三。

[30] http://purl.dila.edu.tw:13180/fuseki/, 2019/6/20。



圖三　以人物規範鏈結資料網頁查詢示意

五、鏈結大藏經文字與佛學知識網絡條目

如同前小節所說明，我們所製作的佛學知識網絡，並不會是獨立運作項目，而是將會結合到大藏經的文字身上，提供文字更多的背景知識。而這樣的工作，其困難點是在於：（一）必須花時間與人力找出文字內所參考到的佛學知識網絡條目，並加以連結文字與參考資料。（二）對於文字載體而言，也需要設計一個良好的機制來承載大量的不同參考訊息。

對於項目一，一般來說除了以人工進行逐字閱讀查找的工作之外，也可以考慮利用新一代人工智慧技術的協助，進行名稱實體的辨識（NER）與進行實體參考項目之連結。本研究的實際執行過程中，我們選擇在過往專案中，已經由人工進行藏經文字與規範資料庫的實體辨

識，並加以鏈結的成果（例如：佛教傳記文學地理資訊系統、中國佛教寺廟志數位典藏資料庫）來進一步加以處理，如下圖四為高僧傳地理資訊系統的文字說明畫面與資料項目連結。因此在本專案中，我們目標首先設定為將這些專案的成果，整合至 CBETA 數位研究平台。這樣的目標設定，預計將可以收到事半功倍之效果。

圖四　高僧傳地理資訊系統的文字說明畫面與資料項目連結

由於 CBETA 的原始資料本身是利用符合 TEI 標記規範的 XML 所製作，雖然在 TEI 規範中，對於各種參考實體，都有提供對應的標記方式。[31]但是若需要把所有紀錄都放回 TEI 之上，則可能需要大量的人工進行資料編輯，加上當多種意義疊加在相同的文字之上時，可能會需要大量的協調，才能找出資料共存之道。因此，我們改採另一種方式，就

[31] 於 TEI 標記標準中，人名以「persName」紀錄，而地名以「placeName」標記，時間資料則建議使用「date」來紀錄。

是將這些相關的參考資料，利用差異資料儲存的方式，也就是將這些資料獨立在 CBETA 的 TEI 之外，而採用相對參考的方式來表達。一個實際的範例如下圖五。

	A	B	C	D	E	F
1	lb	position	tag	type	key	name
2	0003a02	0	place	start	PL13847	普陀
3	0003a02	2	place	end	PL13847	普陀
4	0003a06	9	place	start	PL151	東海
5	0003a06	11	place	end	PL151	東海
6	0004a01	0	place	start	PL13847	普陀
7	0004a01	2	place	end	PL13847	普陀
8	0004a05	11	place	start	PL13847	補陀
9	0004a06	1	place	end	PL13847	補陀
10	0006a04	0	place	start	PL151	東海
11	0006a04	2	place	end	PL151	東海
12	0006a04	4	person	start	A006952	周應賓
13	0006a04	7	person	end	A006952	周應賓

圖五　由佛寺志範例資料抽取出的人物、時間、地點等文字內相關實體所在位置與其規範碼

上圖五中的資料是由《普陀山志》第一卷中所抽出的範例資料，其內容包含實體所在行號（lb）、實體對應標記在該行中的字元位置（position）、該位置為實體的開始或結束位置（type）、[32]實體的型態分類（tag）、對應到規範資料庫內的編號（key）與該位置的文字內容（name）。如此紀錄的目的，也是為了因應可能的底本文字異動。因此當 CBETA 每次變動發行新資料版本時，將會同時確認此差異資訊是否需要對應調整。利用此方式，在底本文字的負擔上相對的小。對於底本

[32] 為實際運作上的方便起見，開始與結束標記分開為兩筆資料來紀錄。

文字工作人員而言，這個改變完全不會干擾原有的工作程序，也就是不需要因為上層解釋資料的增加而有所變化。但是在最終界面的呈現上，則必須將 XML 的底本，與本參考資訊結合在一起後，一同呈現。而這個工作，便交給資料轉換程式在資料處理的過程中來處理即可。目前經過資料中心整合後的 HTML 檔案範例如下圖六。

圖六　將佛寺志人物時間地點參考訊息，整合至前端研究平台之示意範例

六、結語

在本論文中，我們描述了利用鏈結資料打造人工智慧時代佛學數位資源的想法，並說明了我們於實作過程中所遭遇的問題與我們的決策及相關處理細節。目前，我們預計進行兩大目標，首先是利用鏈結資料技術，搭配佛學規範資料庫的豐富內容，初步進行佛學知識網絡的雛型建置。另外，我們也預計將佛學知識網絡的條目，結合到藏經文字身上，以便讓文字能夠增加更多隱藏在字面背後的的意含。實際上，第一階段中所打造的佛典知識庫，僅是這裡我們想要由文字連結知識的其中一個對象，而其他尚可能有：辭典內的詞條解釋、經文之內的相互參考與解釋文字、在不同段落重複使用的解釋、現代白話解釋……等等訊息，都是我們整合的目標。也就是說，我們嘗試能夠將之前散諸於各研究專案成果的知識，收納回 CBETA 的文字核心系統中，以便提供每一段文字訊息，能有更多的參考資訊。期望透過創造、共享與應用更多互相鏈結的語意資料，提供人文學者更具前瞻性與突破性的數位研究平台服務。

致謝

本研究感謝科技部數位人文計畫：漢籍語意鏈結的探討與應用研究—以佛典數位資源為例—漢籍「鏈結開放性資料」（LOD）之研究：以佛典經錄與科判為例（106-2420-H-655 -002 -MY3）之補助，讓本研究得以進行。

引用書目

佛教典籍與古籍

《大正新脩大藏經》之藏經資料引自「中華電子佛典協會」（Chinese Buddhist Electronic Text Association，簡稱 CBETA）的電子佛典系列光碟（2011 年版）。

現代專書、論文

杜正民，2012，〈佛學數位資源的建置與開展〉，《法鼓佛學學報》10，頁 147-210。

洪振洲，2016，〈由資料庫到數位研究平台——談佛典文獻數位研究工具之發展與演變〉，《漢學研究通訊》35：1，年頁 1-14。

洪振洲，2018，〈數位時代漢譯佛典之研究利器——CBETA 數位研究平臺〉，《數位典藏與數位人文》1，頁 149-174。

Sören Auer, Christian Bizer, Georgi Kobilarov, et al.. 2007. DBpedia: A Nucleus for a Web of Open Data The Semantic Web, In *The Semantic Web. ISWC 2007, ASWC 2007. Lecture Notes in Computer Science* 4825. Eds Aberer K. et al. Berlin, Heidelberg: Springer, pp. 722-735. https://doi.org/10.1007/978-3-540-76298-0_52.

Tim Berners-Lee. 2006. “Linked Data”, Design Issues. W3C, 2010/12/18.

網路資源

上海圖書館數據開放平台，http://data.library.sh.cn/index, 2019/7/4。

中央研究院數位文化中心，“linked Taiwan Artists”, http://linkedart.ascdc.tw/index.php, 2019/7/4。

中央研究院數位文化中心鏈結開放資料平台，http://data.ascdc.tw/, 2019/7/4。

佛學規範資料庫，http://authority.dila.edu.tw, 2019/6/20。

BBC —ontologies, http://www.bbc.co.uk/ontologies, 2019/7/4.

BUDA—linked data, https://www.buddhistarchive.org/linked-data-tool, 2019/0/-04.

Build Your Own NYT Linked Data Applicationhttps://open.blogs.nytimes.com/2010/03/30/build-your-own-nyt-linked-data-application/?_r=1, 2019/7/4.

CBETA 線上閱讀網頁，http:// CBETAonline.dila.edu.tw, 2019/7/1。

CBETA 詞彙搜尋與分析網頁，http:// CBETAConcordance.dila.ed.tw, 2019/7/1。

DEDU 對讀文獻製作工具網頁，http:// DEDU.dila.edu.tw, 2019/7/1。

EU Open Data Portal, https://data.europa.eu/euodp/en/linked-data, 2019/7/4.

Google, Introducing the Knowledge Graph: things, not strings, https://googleblog.

blogspot.tw/2012/05/introducing-knowledge-graph-things-not.html, 2019/7/4.

MicroSoft Concept Graph, https://concept.research.microsoft.com, 2017/2/4.

New York Times - Linked Open Data, https://datahub.io/dataset/nytimes-linked-open-data, 2019/7/4.

Oliver Bartlett, Linked Data: Connecting together the BBC's Online Content, http://www.bbc.co.uk/blogs/internet/entries/af6b613e-6935-3165-93ca-9319e1887858, 2019/7/4.

Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page, 2019/7/4.

W3C-ALL STANDARDS AND DRAFTS, https://www.w3.org/standards/techs/linkeddata# w3c_all, 2019/7/4.