

國立臺灣大學電機資訊學院資訊工程研究所

碩士論文

Graduate Institute of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

佛學書目自動化標記與後分類應用研究

Automated Tagging and Post-Classification Analysis of

Bibliographies of Buddhist Studies

蔡鴻謨

Hong-Mo Tsai

指導教授：項潔 博士

Advisor: Jieh Hsiang, Ph.D.

中華民國 111 年 7 月

July, 2022

致謝



很快兩年過去了，每次到一個旅程的終點時，回過頭看總是覺得時光過得好快，還記得初入校園時與學長姊及同學們談話時，大家的學識及見識都相當豐富，使我深深地感覺到自己的無知卻又隱隱有些興奮，而我更加確信這就是我所期望的。

首先，非常感謝項老師給了我進實驗室的機會，雖然每次與老師的談話我都壓力很大深怕下一刻就要被罵了，但是我真的很感謝老師的指導，並且真心地感受到老師對於學術的熱忱以及貢獻，由衷的感到敬佩，雖然常常使老師生氣，但老師還是很有耐心的指導，再次感謝老師的包容以及教導。

再來要感謝各位博士的指導，與各位博士的相處能感受到杜老大對於工作的熱忱、一梅姐的熱心、曹博的隨和以及胡博的溫文儒雅，各位博士的指導使我受益良多，尤其感謝一梅姐幫助我一起修改投影片及對於論文的指導。

也很感謝培峰學長教導了我許多的知識與技術，亦花費許多心思在指導我的論文，由衷感謝學長的用心。

感謝志揚學長、旭恩學長、昱甄學姊給了我不只課業上的建議，也給了我許多人生上的建議，感謝冠霖、譽騰、鈺杰讓我增長了許多見識，一起吃飯聊天的時光也都是十分開心，一起修課討論作業的日子也十分充實。

最後要感謝我的父母帶我到這世界上，讓我能在成長的過程中找尋自己想要的人生，也感謝我爸願意每個月提供生活費給我。

因為疫情許多課程都是線上教學，進校園的次數太少也算是一個遺憾吧，畢竟校園真的很美，這兩年的真的很豐富，很慶幸自己能成為實驗室的一份子，與大家一起並肩前行。

摘要

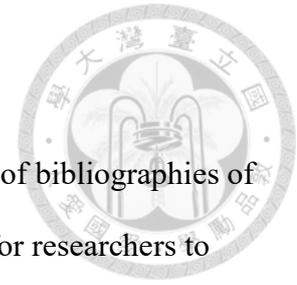


臺大佛學數位圖書館 (NTU Digital Library of Buddhist Studies) 收錄了大量
的佛學書目資料，並且擁有完善的書目檢索系統，提供研究人員做佛學資料的蒐
集。基於書目 metadata 的欄位，臺大佛學數位圖書館的書目檢索系統對檢索結果
提供了 6 種後分類：出版年份、資料類型、出處題名、關鍵詞、著者、語言，這
些後分類都是書目匯入時既有的 metadata 欄位，若是能為書目做佛學類別專門詞
彙的標記，並且將標記做為後分類提供給使用者，將更有利於研究人員對於書目
資料的篩選及整理。

本研究旨在於為臺大佛學數位圖書館的書目資料進行標記，新增三個標記類
別：佛教宗派、佛教人物、佛教經典。事先收集這些類別的專門詞彙進行字串比
對實現自動標記，並建立停用詞表，以利為字串比對的專門詞彙進行篩選，搭配
人工輔助檢核標記、專門詞彙、停用詞表以進行更新，並為臺大佛學數位圖書館
的書目檢索系統新增三種不同的後分類：「提及：佛教宗派」、「提及：佛教人
物」、「提及：佛教經典」，隨著專門詞彙及停用詞表的完整，期望能夠帶給使用
者較為準確的標記及後分類，為使用者帶來更佳的使用體驗。

關鍵字：臺大佛學數位圖書館、字串比對、後分類、停用詞、標記

Abstract



NTU Digital Library of Buddhist Studies has a large collection of bibliographies of Buddhist studies and a comprehensive bibliographic search system for researchers to access Buddhist materials. Based on bibliographic metadata, the bibliographic search system of NTU Digital Library of Buddhist Studies provides six types of post-categories of query results: publication years, media types, source topics, keywords, authors, and languages. These post-classifications utilize the existing metadata attributes of the bibliographies. If the bibliographies can be tagged with specific vocabularies of new categories, the categories and taggings can also be used for additional post-query classification. This will provide more beneficial for researchers to filter and organize the bibliographic data.

The purpose of this study is to tag the bibliographic data of NTU Digital Library of Buddhist Studies by add three new tag categories: Buddhist sects, Buddhist persons, and Sutras. The terms of each categories are collected in advance for string matching to obtain automatic tagging. A list of stopwords is also created to facilitate the filtering of specific words for string matching and correction of tags. The lists of vocabulary and stopwords can be modified manually. We added three different categories to the bibliographic search system of NTU Digital Library of Buddhist Studies. They are "Mentions: Buddhist Sect", "Mentions: Buddhist Person", and "Mentions: Sutra", each with a specialized vocabulary and a list of stopwords. Through post-query classification using these categories, we hope to provide better user experience.

Keywords: NTU Digital Library of Buddhist Studies, string matching, post-classification, stopwords, tagging

目錄



| | |
|--------------------------------|-----|
| 口試委員會審定書 | i |
| 致謝 | ii |
| 摘要 | iii |
| Abstract..... | iv |
| 圖目錄 | vii |
| 表目錄 | ix |
| 第一章 緒論 | 1 |
| 1.1 研究背景與動機 | 1 |
| 1.2 研究目的與問題 | 1 |
| 1.3 論文架構 | 2 |
| 第二章 現況探討與相關文獻 | 3 |
| 2.1 臺大佛學數位圖書館 | 3 |
| 2.2 專門詞彙資料收集參考 | 5 |
| 2.2.1 《佛教圖書分類法(2011 年版)》 | 5 |
| 2.2.2 DILA 人名規範資料庫 | 5 |
| 2.2.3 CBETA 中華電子佛典協會 | 6 |
| 2.3 標記工具參考 | 7 |
| 2.3.1 Markus 半自動標記平台 | 7 |
| 2.4 本章小結 | 8 |
| 第三章 研究方法 | 9 |



| | |
|------------------------|----|
| 3.1 研究對象與範圍 | 9 |
| 3.2 研究流程與方法 | 9 |
| 第四章 功能設計與實作 | 14 |
| 4.1 架構設計 | 14 |
| 4.2 資料處理 | 15 |
| 4.2.1 書目資料表欄位新增 | 15 |
| 4.2.2 停用詞資料表 | 18 |
| 4.2.3 佛教宗派專門詞彙 | 18 |
| 4.2.4 佛教人物專門詞彙 | 20 |
| 4.2.5 佛教經典專門詞彙 | 24 |
| 4.2.6 資料處理實作 | 25 |
| 4.2.7 小結 | 26 |
| 4.3 自動標記實作 | 27 |
| 4.4 輔助人工檢核 | 29 |
| 第五章 UI 功能設計與範例演示 | 32 |
| 5.1 輔助人工檢核範例演示 | 32 |
| 5.2 書目後分類及標記呈現 | 37 |
| 5.3 本章回顧 | 41 |
| 第六章 結論與未來工作 | 42 |
| 6.1 研究結論 | 42 |
| 6.2 未來工作 | 43 |
| 參考文獻 | 44 |

圖目錄



| | |
|-------------------------------|----|
| 圖 2.1 書目檢所篩選範例 | 3 |
| 圖 2.2 網頁書目資訊範例 | 4 |
| 圖 2.3 DILA 人名規範資料欄位資訊範例 | 6 |
| 圖 2.4 Markus 標記提供的方式 | 8 |
| 圖 3.1 研究流程圖示 | 9 |
| 圖 3.2 後分類範例圖 | 13 |
| 圖 4.1 架構設計圖示 | 14 |
| 圖 4.2 比對結果交集示意圖 | 18 |
| 圖 4.3 DILA 人名規範資料庫查詢結果 | 21 |
| 圖 4.4 DILA 人名規範資料庫 xml | 22 |
| 圖 4.5 資料匯入流程 | 26 |
| 圖 4.6 已收錄的書目標記流程 | 27 |
| 圖 4.7 新匯入或更新之書目比對流程 | 28 |
| 圖 4.8 人工檢核平台操作流程 | 30 |
| 圖 4.9 專門詞彙管理平台操作流程 | 31 |
| 圖 5.1 人工檢核平台操作範例 1 | 32 |
| 圖 5.2 人工檢核平台操作範例 2 | 32 |
| 圖 5.3 人工檢核平台操作範例 3 | 33 |
| 圖 5.4 人工檢核平台操作範例 4 | 33 |
| 圖 5.5 人工檢核平台操作範例 5 | 34 |
| 圖 5.6 人工檢核平台操作範例 6 | 34 |
| 圖 5.7 專門詞彙管理平台操作範例 1 | 35 |
| 圖 5.8 專門詞彙管理平台操作範例 2 | 35 |
| 圖 5.9 專門詞彙管理平台操作範例 3 | 36 |

| | |
|----------------------------|----|
| 圖 5.10 專門詞彙管理平台操作範例 4..... | 36 |
| 圖 5.11 書目後分類及標記呈現範例 1..... | 37 |
| 圖 5.12 書目後分類及標記呈現範例 2..... | 38 |
| 圖 5.13 書目後分類及標記呈現範例 3..... | 38 |
| 圖 5.14 書目後分類及標記呈現範例 4..... | 39 |
| 圖 5.15 書目後分類及標記呈現範例 5..... | 40 |
| 圖 5.16 書目後分類及標記呈現範例 6..... | 40 |
| 圖 5.17 書目後分類及標記呈現範例 7..... | 41 |



表目錄



| | |
|-----------------------------|----|
| 表 4.1 書目比對欄位與標記及後分類結果 | 17 |
| 表 4.2 天台宗比對結果 | 17 |
| 表 4.3 佛教宗派專門詞彙資料 | 20 |
| 表 4.4 佛教人物專門詞彙資料於資料庫 | 23 |
| 表 4.5 佛教經典專門詞彙資料於資料庫 | 25 |
| 表 4.6 各類別資料處理統計 | 26 |



第一章 緒論

1.1 研究背景與動機

臺大佛學數位圖書館 (NTU Digital Library of Buddhist Studies) [1]-[2] 成立已有 20 多年，截自目前收錄的書目已有 40 多萬筆，其中包含全文的書目共 10 萬多筆，包括 47 種語言及 15 種資料類型，臺大佛學數位圖書館在其收集的豐富書目下提供了書目檢索系統來做書目的搜尋，而其書目檢索系統針對查詢結果提供了許多的後分類、相關經律論及 DILA 佛學術語字辭典的外部連結。

臺大佛學數位圖書館在檢索系統的搜尋結果上提供了後分類來做進一步的篩選，目前共有 6 種後分類：出版年份、資料類型、出處題名、關鍵詞、著者、語言，而這 6 種後分類有一個共通性，都是基於匯入書目的 metadata 欄位所做的後分類，都是較為通用的後分類，若是針對佛學特定的類別收集專門詞彙，為書目做佛學類別相關的標記，並且將標記做為後分類提供給使用者，使得後分類能夠針對一些佛學特定的類別去做篩選，更能幫助使用者做資料清理及篩選，讓查詢結果能夠有更多的選擇及多樣性。

1.2 研究目的與問題

本論文主要目的為研究如何利用臺大佛學數位圖書館的書目資料進行標記，新增三種標記類別：佛教宗派、佛教人物、佛教經典，並為臺大佛學數位圖書館的書目檢索系統新增三種不同的後分類：「提及：佛教宗派」、「提及：佛教人物」、「提及：佛教經典」。

研究的問題主要為：以什麼方式對書目進行標記、如何篩選佛學類別的專門詞彙及書目標記如何達到準確、如何將書目標記做為後分類、如何將整個功能架構與臺大佛學數位圖書館原有架構介接並將標記及後分類呈現給使用者、同

樣的模式能否沿用到新的類別上。



1.3 論文架構

本論文共六章，分別為以下結構及目的

第一章 緒論：研究背景、動機、目的與問題及論文架構。

第二章 現況探討與相關文獻：講述目前臺大佛學數位圖書館的現況及細節與相關文獻的回顧。

第三章 研究方法：整體研究的流程與方法。

第四章 功能設計與實作：整體功能的架構與實作方法。

第五章 UI 功能設計與範例演示：UI 功能的設計介紹與整體功能的範例演示。

第六章 結論與未來工作：總結整體研究並提出未來發展的可能性及方向。



第二章 現況探討與相關文獻

2.1 臺大佛學數位圖書館

臺大佛學數位圖書館目前的書目檢索系統是採用 solr 做為搜索引擎，目前提供以下欄位查詢（輸入查詢詞，查詢方式採用部分符合）：題名（書名）、出處題名、著者、關鍵詞、摘要、目次、叢書名、ISBN/ISSN、不限欄位，除顯示查詢結果外，也會將書目中出現的查詢詞以紅字(顯色)標示，而針對查詢結果可以再做後分類，目前有 6 種後分類：出版年份、資料類型、出處題名、關鍵詞、著者、語言，而其後分類採用的是同種類別為聯集而不同類別為交集的方式，例如：查詢”戒律”，後分類選擇出處題名的”香光莊嚴”及”海潮音”、著者的”釋悟因”進行篩選，此時查詢結果將會是，出處題名（”香光莊嚴”|”海潮音”）&著者（”釋悟因”），&為交集、|為聯集。查詢結果的單筆書目顯示時，其資訊內容並沒有將查詢詞做顯色。圖 2.1 為臺大佛學數位圖書館檢索頁面的概況。

The screenshot displays the search interface for '戒律' (Vinaya) in the National Taiwan University Buddhist Digital Library. On the left, there are search filters for '出處題名' (Source Title) and '著者' (Author), with '香光莊嚴' and '海潮音' selected under source titles, and '釋悟因' selected under authors. The main results list shows four items, with the third item, '談戒說律-- 戒律是生活型態的選擇', highlighted in orange. On the right, a detailed view of this item is shown, including its title, author, publication information, and a summary.

圖 2.1 書目檢所篩選範例

臺大佛學數位圖書館為每一筆收錄的書目著錄有詳細的欄位，目前的欄位有：seq（序號）、topic（題名）、author（作者）、source_topic（出處題名）、page（頁次）、press_time（出版日期）、publisher（出版者）、publisher_location（出版地）、publisher_url（出版者網址）、bilogyraphy_language（使用語言）、media_type（資料類型）、remark（附註項）、keyword（關鍵詞）、summary（摘要）、tablecontent（目次）...等，計 76 欄，其中有些欄位為多重值欄位如：author（作者）、publisher（出版者）、bilogyraphy_language（使用語言）、keyword（關鍵詞）...等，而目前這些欄位儲存方式為以“;”分隔開。但並非每一筆書目在所有欄位中皆會有資料，在檢索頁面上僅會呈現該筆書目有資訊值的欄位，以序號為 120032 的書目為例在網頁上顯示如圖 2.2。

The screenshot shows a search interface with a search bar containing '阿閼佛' and a search button. Below the search bar, there are navigation links for '進階查詢', '查詢歷史', and '使用說明'. The search results are displayed in a table-like format with various fields and values.

| 欄位名稱 | 值 |
|-------|---|
| 作者 | 林保彝 (著) |
| 出處題名 | 中華佛學研究=Chung-Hwa Buddhist Studies |
| 卷期 | n.8 |
| 出版日期 | 2004.03 |
| 頁次 | 295 - 318 |
| 出版者 | 中華佛學研究所=The Chung-Hwa Institute of Buddhist Studies |
| 出版者網址 | http://www.chibs.edu.tw/ch_html/0_ch04/ch0404_02.html |
| 出版地 | 臺北市, 臺灣 [Taipei shih, Taiwan] |
| 資料類型 | 期刊論文=Journal Article |
| 使用語言 | 中文=Chinese |
| 附註項 | 作者為臺北藝術大學美術史研究所教授 |
| 關鍵詞 | 心靈環保; 阿閼佛 |
| 摘要 | 山東四門塔阿閼佛，是臺灣2002年的重大文化記事，從年初到年末，整整一年，一直牽引著臺灣島下的文化脈絡，只要有些些訊息，瞬間成為關心人們的注視焦點，尤其在法鼓山文教基金會以人類世界文化遺產的關照懷抱，開始為阿?佛頭像探尋「回家」的可能性後，頓時成為海內外關注的盛事。末學因研習佛教藝術機緣，有幸蒙法鼓山聖嚴師父的囑咐，參與了這趟山東神道寺四門塔石雕佛頭像，復歸佛身的「世紀之索」心靈環保與文化洗禮。 |
| 目次 | 一、世紀之索、修行之願 295 二、禮釋專研、決斷疑難 296 三、流轉聚首、世紀啟航 297 四、流轉聚首、再現一波 298 五、文資保存、在資料庫 299 |
| ISSN | 1026969X (P) |
| 點閱次數 | 2225 |
| 建檔日期 | 2005.02.04 |
| 更新日期 | 2021.04.07 |

圖 2.2 網頁書目資訊範例



2.2 專門詞彙資料收集參考

為了針對佛學類別新增一些標記及後分類的種類，需要參考到以下書籍文獻或資料庫，佛教宗派專門詞彙參考自《佛教圖書分類法(2011年版)》、佛教人物專門詞彙參考自 DILA 人名規範資料庫、佛教經典專門詞彙參考自 CBETA 中華電子佛典協會。

2.2.1 《佛教圖書分類法(2011年版)》

《佛教圖書分類法(2011年版)》 [3]-[4] 由香光尼眾佛學院圖書館編撰，其大綱共有 10 個類別為：「00 總類」、「1 教理」、「2 佛教史地」、「3 經及其釋」、「4 律及其釋」、「5 論及其釋」、「6 儀制；修持；布教；護法」、「7 佛教文藝；佛教語文」、「8 中國佛教宗派」、「9 世界各地佛教宗派」。

大綱的 10 個類別再往下細分子類別，例如「9 世界各地佛教宗派」下有：「900 世界各地佛教宗派」、「930 西藏佛教宗派」、「940 韓國佛教宗派」、「950 日本佛教宗派」；每一個子類別再往下細分，例如「950 日本佛教宗派」下再細分：「951 南都六宗」、「952 天台宗」、「953 真言宗」、「954 禪宗」、「955 淨土教」、「956 日蓮宗」、「959 日本其他宗派」，可以再往下細分，例如「951 南都六宗」下有：「951.1 俱舍宗」、「951.2 成實宗」、「951.3 三論宗」、「951.4 法相宗」、「951.5 華嚴宗」、「951.6 律宗」。

2.2.2 DILA 人名規範資料庫

DILA 佛學規範資料庫包含多個資料庫，有：時間規範資料庫、人名規範資料庫、地名規範資料庫、佛經目錄規範資料庫，而其中的人名規範資料庫 (Person Authority Database) [5] 提供佛學相關人物的名稱規範資料查詢。

人名規範資訊包含有以下欄位：名稱、別名 Alternative Names、規範碼 ID

No、人名群組 Person Group、舊編碼、朝代 Dynasty、籍貫 From、性別 Gender、是否為歷史人物 Historical Person、註解（精簡） Notes（simple）、師徒關係、SameAs、Occurs in...等欄位，並非所有人物都含有所有的欄位，例文殊菩薩如圖 2.3。



文殊菩薩(wén shū pú sà)

別名 Alternative Names :
[正體中文] 三世覺母妙吉祥(sān shì jué mǔ miào jí xiáng),文殊師利(wén shū shī lì),妙吉祥(miào jí xiáng),妙德(miào dé),滿祖室哩(mǎn zǔ shì lì),曼
[藏文 (羅馬轉寫)] jam dpal
[梵文 (羅馬轉寫)] Mañjuśrī

規範碼 ID No : A002144
人名群組 Person Group : 華嚴三聖;十二大士;
舊編碼 : G0145
朝代 Dynasty : 沒有給定朝代;
籍貫 From : 舍衛國(PL000000048396)
是否為歷史人物 Historical Person : 否
註解(精簡) Notes(simple) : 佛教四大菩薩之一，以智慧著稱，與般若經典
衛國多羅聚落梵德婆羅門家，後至釋迦牟尼佛所出家學道。一般稱文殊師
SameAs :

- Wikidata :
1. <https://www.wikidata.org/wiki/Q471696>

Occurs in :

- CBETA T50n2060_p0444c11 唐高僧傳: 釋慧淨傳
- CBETA T50n2060_p0452a10 唐高僧傳: 釋玄奘傳 { 曼殊室利 }
- CBETA T50n2060_p0566b09 唐高僧傳: 釋智顛傳

圖 2.3 DILA 人名規範資料欄位資訊範例

DILA 佛學規範資料庫中的下載區可以下載資料庫中的完整資料，其中 DILA 人名規範資料庫下載檔案有：.rdf、.xml、.rnc 檔案。

2.2.3 CBETA 中華電子佛典協會

CBETA 中華電子佛典協會 [6] 於 1998 年 2 月 15 日正式成立，其宗旨為「收集所有的漢文佛典，以建立電子佛典集成。」、「研發佛典電子化技術，提昇佛典交流與應用。」、「利用電子媒體之特性，以利佛典保存與流通。」

、「期望讓任何想要閱藏的人都有機會如願以償。」。

CBETA 提供了許多佛典的線上閱讀 (CBETA Online)，並以英文字母將經典區分為：「T 大正藏」、「X 卍續」、「A 趙城金藏」、「K 高麗大藏經-新文豐版」、「S 宋藏遺珍-新文豐版」、「F 房山石經」、「C 中華大藏經-中華書局版」、「U 洪武南藏」、「P 永樂北藏」、「J 嘉興藏」、「L 乾隆大藏經-新文豐版」、「G 佛教大藏經」、「M 卍正藏經-新文豐版」、「D 國圖善本」、「N 南傳大藏經」、「ZS 正史佛教資料類編」、「I 北朝佛教石刻拓片百品」、「ZW 藏外佛教文獻」、「B 大藏經補編」、「GA 中國佛寺史志彙刊」、「GB 中國佛寺志叢刊」、「Y 印順法師佛學著作集」、「LC 呂澂佛學著作集」...等，而其中每個類別都可以往下細分（例：阿含部上為 T01、阿含部下為 T02），而其中每一部又可以再往下細分（例：T0001 長阿含經（22 卷）為阿含部上的第 1 本經典共 22 卷、T0002 七佛經（1 卷）為阿含部上的第 2 本經典共 1 卷），可用此規則做查詢，長阿含經（22 卷）為阿含部 T01 的地 1 本經典 T0001，因此可用 T01n0001 查詢到此經典，而 CBETA 有提供所有經典及其清單的下載。

2.3 標記工具參考

為了規劃標記的方式及標記的規則，本研究參考 Markus 半自動標記平台。

2.3.1 Markus 半自動標記平台

Markus [7]-[8] 半自動標記平台可以協助使用者在中文或韓文的文本標記各類別的專門詞彙，如：人名、地名、時間等詞彙，Markus 提供三種標記方式：自動標記、手動標記、關鍵字標記，通常都先以自動標記的方式，鏈結現有的詞彙資料庫進行標記，例如：人名的 CBDB、DILA 及韓國人名；後以手動標記修正自動標記的結果，或加上使用者的自訂標記，而關鍵字標記則能夠用關鍵字、正規表達式來進行自動標記。



二A：自動標記

以中國歷代人物傳記資料庫 (CBDB) 標記中國人名 (全名、字號) 及官名。
以中國歷史地理信息系統 (CHGIS) 及臺灣歷史地名 (TWGIS) 標記中國地名。
以法鼓佛教學院時間規範資料庫 (DILA Time Authority Database) 標記日期。
以法鼓佛教學院人名規範資料 (DILA Person Authority Database) 標記佛教人名。
以法鼓佛教學院地名規範資料 (DILA Place Authority Database) 標記佛教地名。
以法鼓佛教學院佛學詞彙 (DILA Glossaries of Buddhist Studies) 標記佛學詞彙。

以韓國歷代人物綜合情報系統 (한국역대인물 종합정보 시스템) 及韓國民族文化大百科詞典 (한국민족문화대백과사전) 標記韓國歷代人名 (全名、字號、諱號等)。
以韓國民族大百科詞典及東亞圖 (동아도) 標記韓國地名。
以朝鮮王朝實錄 (조선왕조실록) 及朝鮮王朝實錄詞典 (조선왕조실록 사전) 標記朝鮮時期官職名。
以韓國民族大百科詞典、韓國文集叢刊 (한국문집총간) 以及朝鮮王朝實錄詞典標記書名。

你可以勾選要標記的類別：

Ready

二B：手動標記

通常先經過二A後，再手動標記。你可以用手動標記，修正自動標記的結果，加註新標記，出現多個同名關鍵字時，選定關鍵字，並修正錯誤。不過你也可以直接以手動標記。

Ready

二C：關鍵字標記

你可以上傳自己的語詞列表，或建立自己的關鍵字列表，接著你可以從你的文本，來閱讀或獲取與此列表上語詞相關的資訊。

Ready

三：段落篩選

從檔案中擷取含特定標記的段落

Ready

圖 2.4 Markus 標記提供的方式

2.4 本章小結

Markus 的自動標記以各個資料庫的專門詞彙來對文本進行標記，可以快速且有效的對文本進行標記，對於準確性也有手動標記能夠對自動標記進行修改，藉此半自動的方式可以快速並準確的對文本進行標記。

依照 Markus 的自動標記方式，參考佛教宗派、佛教人物、佛教經典相關的書籍或資料庫並收集專門詞彙。



第三章 研究方法

3.1 研究對象與範圍

主要研究對象是針對臺大佛學數位圖書館提供的 443138 筆書目資訊進行標記及檢索頁面的後分類功能開發。詞彙相關資料收集方面自《佛教圖書分類法(2011 年版)》、DILA 人名規範資料庫、及 CBETA 中華電子佛典協會。《佛教圖書分類法(2011 年版)》中有 10 大類別，佛教宗派的專門詞彙收錄於「8 中國佛教宗派」、「9 世界各地佛教宗派」，因此只取這兩大類別為資料收錄對象，DILA 人名規範資料庫及 CBETA 中華電子佛典協會所收錄的人物專門詞彙及佛典專門詞彙則全數為資料收錄對象。

3.2 研究流程與方法

針對研究問題及研究對象進行一系列的研究流程與方法，並評估研究方法，從初步的功能規劃帶到資料收集、處理及資料篩選，再到標記方法規劃及後分類處理方式，最後將標記及後分類呈現給使用者，依照研究流程逐步說明使用到的方法及整體的設計，以回應相關的研究問題。

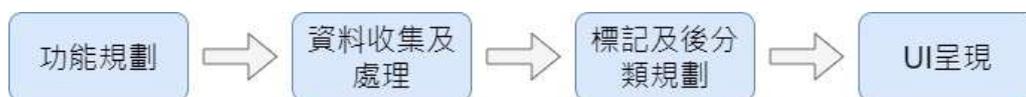


圖 3.1 研究流程圖示

3.2.1 功能規劃

為臺大佛學數位圖書館的書目進行標記，並將標記做為後分類，在書目檢索系統將書目標記及後分類呈現給使用者，輔助使用者做資料篩選。

自動標記的做法選用事先收集相關的專門詞彙並與書目的相關欄位資訊進行

字串比對（新增三種標記類別：佛教宗派、佛教人物、佛教經典），所以必須針對這些類別了解相關的資訊或資料庫，並將資料進行收集及處理。



3.2.2 資料收集及處理

將資料收集及處理分為兩個問題：如何將各個來源的專門詞彙收錄進臺大佛學數位圖書館的資料庫、收錄後的資料如何進行篩選。

(一) 如何將各個來源的專門詞彙收錄進臺大佛學數位圖書館的資料庫

各類別的專門詞彙來源皆不相同，若來源於持續更新的資料庫（如：DILA 人名規範資料庫）代表專門詞彙仍會繼續進行更新，因此不能單單只匯入一次資訊到資料庫中。

針對各類別專門詞彙的資料來源做出不同解析資料的程式，只需更新儲存資料來源的檔案並運行程式，可以使存在於臺大佛學數位圖書館資料中的各類別專門詞彙持續更新。

(二) 收錄後的資料如何進行篩選

收錄後數量眾多的專門詞彙若是不進行任何處理，錯誤的標記將會太多，要讓錯誤的標記數量降低就必須先對資料進行篩選。

針對中文的專門詞彙，UTF-16 編碼長度為 1 的專門詞彙常有錯誤標記的情況發生（例：高、成、亮、進、書...等），因此首先將 UTF-16 編碼長度為 1 的專門詞彙篩選掉。

UTF-16 編碼長度為 2 的中文專門詞彙也有許多較為通用的詞彙（例：實際、傳統、黃金、修行、目錄、文獻...等），但若因為這些詞彙將 UTF-16 編碼長度為 2 的專門詞彙篩選掉又容易漏標有意義的資訊，因為光收錄佛教人物專門詞彙中就有 2 萬多個詞彙都是 UTF-16 編碼長度為 2，為解決這些通用詞彙建立一

個停用詞專用資料表存放這些詞彙，存在於停用詞資料表中的詞彙都不會被拿來做總體的比對標記。



3.2.3 標記及後分類規劃

這部分分為幾個問題：如何進行標記、如何讓標記達到準確、如何將書目標記做為後分類。

(一) 如何進行標記

以收集在資料庫中的專門詞彙與書目的欄位進行字串比對達到自動標記，書目需要比對的欄位要與內文可能有高度相關的欄位，而其中題名、摘要、關鍵詞、目次皆與內文可能有高度相關，因此取這4個欄位來與專門詞彙進行比對，其中關鍵詞較為特殊，因為關進詞本身就應該是一個詞彙，因此不能標記子字串而應該完全配對才標記（例：關鍵詞為”妙法蓮華經”，若專門詞彙中若只有”蓮華”則不應該進行標記），目前標記有三種類別：佛教宗派、佛教人物、佛教經典，與三種類別的專門詞彙做字串比對，比對成功則將字串標記為專門詞彙所屬的類別。

標記的規則也是需要制定（例：標記是否允許重疊），制定的標記規則為標記不允許重疊，且若是開頭位置相同則取最長的標記，若專門詞彙改成”妙法”及”妙法蓮華經”，則此時應該取”妙法蓮華經”標記為”妙法蓮華經”；標記從文本開頭走到文本結尾先標記的為優先，例：題名為”妙法蓮華經”，若專門詞彙有”法蓮”及”蓮華”，則此時應該取”法蓮”標記為”妙法蓮華經”。

標記的規則有了再來是如何儲存標記，做法是以 JSON 儲存，儲存方式為：
{類別:{比對欄位:[標記, 標記, ...]}}，而有了類別及比對欄位後標記儲存方式為 UTF-16 編碼子字串開頭所在的索引位置(0-based indexing)及子字串長度組成的陣列，例：題名為〈天臺智者大師摩訶止觀實踐之道〉（取自釋廣琳，1983年作



品)，專門詞彙有佛教人物的“智者大師”及佛教經典的“摩訶止觀”，自動標記為〈天臺智者大師摩訶止觀實踐之道〉，自動標記儲存為{"person":{"topic":[[2,4]]},"sutra":{"topic":[[6,4]]}}，代表說題名 UTF-16 編碼索引(0-based indexing)為 2 的地方有一個 UTF-16 編碼長度為 4 的佛教人物標記，而題名 UTF-16 編碼索引(0-based indexing)為 6 的地方有一個 UTF-16 編碼長度為 4 的佛教經典標記，之後在以此 JSON 來做 html 上實際的標記。

(二) 如何讓標記達到準確

自動標記後的結果也許仍有許多錯誤或缺漏的標記，為了能夠讓標記越來越準確，因此提供輔助人工檢核，包含單筆書目的人工標記及專門詞彙管理平台，人工標記能夠完善單筆書目的標記，輔助人工檢核也可以將詞彙加入停用詞及專門詞彙的資料表中，隨著專門詞彙資料及停用詞的完善，定時的自動標記更新可以使書目的自動標記越來越準確，人工標記的儲存方式與標記規則皆與自動標記一樣，有了人工標記後的書目，其自動標記更新也不會與人工標記重疊，而以人工刪除的標記在往後書目的自動標記中也不會再次被標記。

因為需要定時的自動標記，因此每次自動標記的執行時間能夠越低越好，對字串比對及資料表更新做優化將有效降低執行時間。

字串比對方面選用 Trie [9] 資料結構來降低執行時間，Trie 又稱為字典樹或前綴樹，Trie 通常由數個字串或數字序列或者其他結構組成一個樹狀結構，一般來說若每一個專門詞彙都需與字串做比對， n 個專門詞彙就必須與字串比對 n 次，而事先將所有專門詞彙編成 Trie 能夠使 n 個專門詞彙只比對 1 次字串就能找出所有符合的標記。

每次自動標記完成後都需將單筆書目的資訊寫回資料庫，使用的指令為 Update Table SET ... WHERE seq=?，以序號做為查找的依據去更新序號符合的書目，因此時間複雜度為 $O(N)$ (N 為書目數量)，若是以序號(seq)為索引鍵建立叢



集索引 [10]-[11]，將能使時間複雜度降為 $O(\log N)$ (N 為書目數量)。

(三) 如何將書目標記做為後分類

因為標記取自不同的欄位，除了將標記本身當作後分類外，也要將標記來源欄位當作後分類，例：題名為〈天臺智者大師**摩訶止觀**實踐之道〉(取自釋廣琳，1983年作品)的書目，因”**摩訶止觀**”為佛教經典，這個書目會歸類到「提及：佛教經典」後分類的”**摩訶止觀**”及”**摩訶止觀**-題名”中，當「提及：佛教經典」後分類篩選”**摩訶止觀**”時搜尋結果中所有欄位標記”**摩訶止觀**”的書目皆會被篩選出來，而當「提及：佛教經典」後分類篩選”**摩訶止觀**-題名”時搜尋結果中題名有標記”**摩訶止觀**”(即”**摩訶止觀**-題名”)的書目才會被篩選出來。

3.2.4 UI 呈現

後分類部分臺大佛學數位圖書館的書目檢索系統本身就有 6 種後分類，這 6 種後分類皆是一層不可展開的勾選框，而這次新增的三種後分類除了有專門詞彙本身，還有標記來源欄位，因此在原先的架構上設計成兩層可展開的結構，第一層為專門詞彙，第二層為專門詞彙標記來源欄位，而標記部分則以文字的背景顏色不同來呈現給使用者(類似螢光筆的效果)。



圖 3.2 後分類範例圖

第四章 功能設計與實作



4.1 架構設計

將架構分為資料處理、自動標記、輔助人工檢核、UI 呈現，定時的資料處理腳本可以更新專門詞彙，輔助人工檢核搭配定時的自動標記越來越準確，最後將標記及後分類呈現給使用者，如圖 4.1。

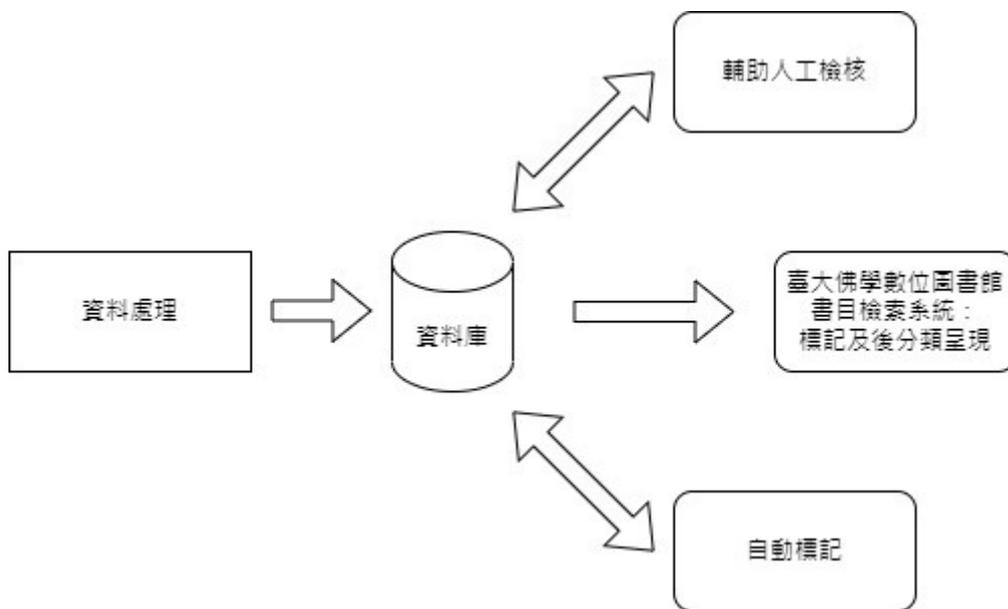


圖 4.1 架構設計圖示

自動標記分為兩個部分：對已收錄的書目進行標記更新、對新匯入的書目進行標記或被更新的舊有書目進行標記更新；輔助人工檢核共有兩個平台：人工標記檢核平台、專門詞彙管理平台；最後標記及後分類呈現則是跟臺大佛學數位圖書館的書目檢索系統做介接。



4.2 資料處理

此次新增的標記類別共三種：佛教宗派、佛教人物、佛教經典，佛教宗派專門詞彙資料取自《佛教圖書分類法(2011年版)》，佛教經人物專門詞彙資料取自 DILA 人名規範資料庫，佛教經典專門詞彙資料取自 CBETA 中華電子佛典協會，需為三種佛學類別及停用詞新增資料表，及為書目的資料表新增欄位存放標記及後分類。

4.2.1 書目資料表欄位新增

按照之前臺大佛學數位圖書館對多重值處理的方式，為存在資料庫中的書目資料表新增 6 個後分類所需欄位，以及存放標記的 3 個欄位及人工檢核時間欄位，共新增 10 個欄位分別是：

- sect，存放「提及的佛教宗派」
- sect_from，存放「提及的佛教宗派」及比對的欄位
- person，存放「提及的佛教人物」
- person_from，存放「提及的佛教人物」及比對的欄位
- sutra，存放「提及的佛教經典」
- sutra_from，存放「提及的佛教經典」及比對的欄位
- tag_auto，存放自動標記的資訊，以 JSON 字串方式儲存，紀錄的方式為 UTF-16 編碼子字串的位置
- tag_arti，存放人工標記的資訊，以 JSON 字串方式儲存，紀錄的方式為 UTF-16 編碼子字串的位置
- tag_remove，存放已刪除的標記資訊，以 JSON 字串方式儲存，紀錄的方式為 UTF-16 編碼子字串的位置
- check_time，存放最近一次的檢核時間



以下取序號為 126449，題名為〈天臺智者大師摩訶止觀實踐之道〉的書目為例，說明比對後的結果（標記紅色為佛教宗派專門詞彙、藍色為佛教人物專門詞彙、綠色為佛教經典專門詞彙）：

題名為〈天臺智者大師摩訶止觀實踐之道〉，其摘要為『天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是非常重要的，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。』，關鍵詞為『天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顛=智者大師=天台大師; 摩訶止觀』，目次為空。（取自釋廣琳，1983年作品），以此筆書目為例，以下為自動標記及人工檢核後的標記和後分類欄位的儲存值：

- sect, "禪宗;Tendaishu;法華宗;Tien-tai Buddhism;華嚴宗;天臺宗;天台宗"
- sect_from, "天臺宗-摘要;法華宗-關鍵詞;Tendaishu-關鍵詞;華嚴宗-摘要;天台宗-關鍵詞;禪宗-摘要;Tien-tai Buddhism-關鍵詞"
- person, "智顛;天台大師;智者大師"
- person_from, "智者大師-題名;智顛-關鍵詞;智者大師-關鍵詞;天台大師-關鍵詞"
- sutra, "摩訶止觀"
- sutra_from, "摩訶止觀-題名;摩訶止觀-關鍵詞"
- tag_auto,


```

      {"sect":{"summary":[[42,2],[45,3]],"keyword":[[0,3]],"person":{"topic":[[2,4]],"keyword":[[37,2],[40,4],[45,4]]},"sutra":{"topic":[[6,4]],"keyword":[[51,4]]}}
      
```
- tag_arti, {"sect":{"keyword":[[4,3],[8,17],[26,9]],"summary":[[0,3],[139,3]]}}
- tag_remove, {"person":{"summary":[["110",2]]}}



| | |
|-------------|--|
| seq | 126449 |
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所创建的佛學之一。它在學 |
| 目次 | |
| sect | 禪宗;Tendaishu;法華宗;Tien-tai Buddhism;華嚴宗;天臺宗;天台宗 |
| sect_from | 天臺宗-摘要;法華宗-關鍵詞;Tendaishu-關鍵詞;華嚴宗-摘要;天台宗-關鍵詞;禪宗-摘... |
| person | 智頌;天台大師;智者大師 |
| person_from | 智者大師-題名;智頌-關鍵詞;智者大師-關鍵詞;天台大師-關鍵詞 |
| sutra | 摩訶止觀 |
| sutra_from | 摩訶止觀-題名;摩訶止觀-關鍵詞 |
| tag_auto | {"sect":{"summary":[[42,2],[45,3]],"keyword":[[0,3]],"person":{"topic":[[2,4]],"keyword... |
| tag_arti | {"sect":{"keyword":[[4,3],[8,17],[26,9]],"summary":[[0,3],[139,3]]}} |
| tag_remove | {"person":{"summary":["110",2]}} |
| 檢核時間 | 2022-08-09 15:17:32 |

表 4.1 書目比對欄位與標記及後分類結果

題名、摘要、關鍵詞、目次，只要比對成功皆會加入後分類，若題名比對到了天台宗，需加入”天台宗”及”天台宗-題名”來確保數量的正確，並且要避免重複加入後分類，若題名及摘要都比對到了天台宗，此時應為”天台宗”及”天台宗-題名;天台宗-摘要”，例：

最終 443138 筆書目中天台宗總共 2463 筆，取自題名共 515 筆，取自摘要共 483 筆，取自關鍵詞共 1659 筆，取自目次共 361 筆。

| | |
|---------|------|
| 天台宗 | 2463 |
| 天台宗-題名 | 515 |
| 天台宗-摘要 | 483 |
| 天台宗-關鍵詞 | 1659 |
| 天台宗-目次 | 361 |

表 4.2 天台宗比對結果

因此彼此之間的關聯是會有交集的，根據目前臺大佛學數位圖書館取得 solr 統計資料的方式，必須將標記及標記來源欄位都儲存才能不確保出錯。

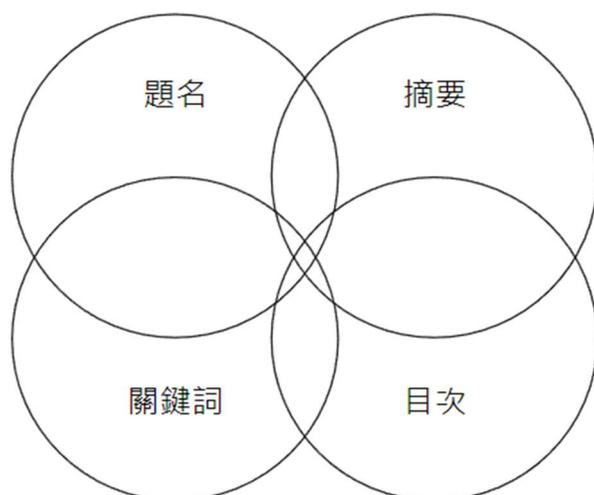


圖 4.2 比對結果交集示意圖

4.2.2 停用詞資料表

收錄的專門詞彙較多，常會有較為通用的詞彙出現，如：傳統、目錄、黃金、修行、樂園、可能...等詞彙，加入停用詞後往後標記更新將不會再將這些詞彙進行自動標記，但人工仍可以標記，以下為資料表的設計：

- seq，表示每筆資料的識別
- id，存放 ID（人工加入以 SW 開頭往上遞增，第 1 筆為 SW00001）
- name，存放停用的詞彙名稱
- version，存放資料來源
- create_time，存放此筆資料的創立時間

4.2.3 佛教宗派專門詞彙

佛教宗派專門詞彙資料取自《佛教圖書分類法(2011 年版)》，其中包含中國佛教宗派及世界各地佛教宗派，其中世界各地佛教宗派包括：西藏佛教宗派、韓國佛教宗派、日本佛教宗派，其中韓國佛教宗派暫未有宗派被收錄其中，以下為資料表的設計：



- seq, 表示每筆資料的識別
- id, 存放分類法的 ID
- name, 存放宗派的名稱或別名
- is_alternative, 識別是否為別名
- language, 存放名稱或別名使用的語言
- name_key, 識別同個宗派不同的名稱或者別名
- remark, 儲存備註
- version, 存放資料來源
- create_time, 存放此筆資料的創立時間

資料中只有 955.2 的”真宗”有收錄別名為”淨土真宗”，但仍保留 is_alternative 的欄位作為區分名稱及別名。

資料匯入資料庫則是先將宗派整理至 excel 中，並且將之匯入，目前只收錄了中文的資料，因此 language 皆為中文，最後匯入共有 45 筆，總共 44 個宗派，中有 1 個宗派有別名。



| | A | B | C | D | E | F | G |
|----|-------|-------|----------------|----------|---|---|---|
| 1 | id | name | is_alternative | language | | | |
| 2 | 810 | 三論宗 | 0 | 中文 | | | |
| 3 | 820 | 天台宗 | 0 | 中文 | | | |
| 4 | 830 | 法相宗 | 0 | 中文 | | | |
| 5 | 840 | 華嚴宗 | 0 | 中文 | | | |
| 6 | 850 | 律宗 | 0 | 中文 | | | |
| 7 | 860 | 禪宗 | 0 | 中文 | | | |
| 8 | 868.1 | 潯仰宗 | 0 | 中文 | | | |
| 9 | 868.2 | 臨濟宗 | 0 | 中文 | | | |
| 10 | 868.3 | 曹洞宗 | 0 | 中文 | | | |
| 11 | 868.4 | 雲門宗 | 0 | 中文 | | | |
| 12 | 868.5 | 法眼宗 | 0 | 中文 | | | |
| 13 | 868.6 | 黃龍宗 | 0 | 中文 | | | |
| 14 | 868.7 | 楊岐宗 | 0 | 中文 | | | |
| 15 | 870 | 淨土宗 | 0 | 中文 | | | |
| 16 | 880 | 密宗 | 0 | 中文 | | | |
| 17 | 891 | 俱舍宗 | 0 | 中文 | | | |
| 18 | 892 | 毘曇宗 | 0 | 中文 | | | |
| 19 | 893 | 成實宗 | 0 | 中文 | | | |
| 20 | 894 | 涅槃宗 | 0 | 中文 | | | |
| 21 | 895 | 攝論宗 | 0 | 中文 | | | |
| 22 | 896 | 地論宗 | 0 | 中文 | | | |
| 23 | 931 | 寧瑪派 | 0 | 中文 | | | |
| 24 | 932 | 噶舉派 | 0 | 中文 | | | |
| 25 | 933 | 薩迦派 | 0 | 中文 | | | |
| 26 | 934 | 格魯派 | 0 | 中文 | | | |
| 27 | 951 | 南都六宗 | 0 | 中文 | | | |
| 28 | 951.1 | 俱舍宗 | 0 | 中文 | | | |
| 29 | 951.2 | 成實宗 | 0 | 中文 | | | |
| 30 | 951.3 | 三論宗 | 0 | 中文 | | | |
| 31 | 951.4 | 法相宗 | 0 | 中文 | | | |
| 32 | 951.5 | 華嚴宗 | 0 | 中文 | | | |
| 33 | 951.6 | 律宗 | 0 | 中文 | | | |
| 34 | 952 | 天台宗 | 0 | 中文 | | | |
| 35 | 953 | 真言宗 | 0 | 中文 | | | |
| 36 | 954 | 禪宗 | 0 | 中文 | | | |
| 37 | 954.1 | 臨濟宗 | 0 | 中文 | | | |
| 38 | 954.2 | 曹洞宗 | 0 | 中文 | | | |
| 39 | 954.3 | 黃檗宗 | 0 | 中文 | | | |
| 40 | 955 | 淨土教 | 0 | 中文 | | | |
| 41 | 955.1 | 淨土宗 | 0 | 中文 | | | |
| 42 | 955.2 | 真宗 | 0 | 中文 | | | |
| 43 | 955.2 | 淨土真宗 | 1 | 中文 | | | |
| 44 | 955.3 | 融通念佛宗 | 0 | 中文 | | | |
| 45 | 955.4 | 時宗 | 0 | 中文 | | | |
| 46 | 956 | 日蓮宗 | 0 | 中文 | | | |

表 4.3 佛教宗派專門詞彙資料

4.2.4 佛教人物專門詞彙

佛教人物專門詞彙資料取自 DILA 人名規範資料庫，為 2022 年 02 月的版本，DILA 有提供 XML 檔案，XML 上各個標籤與網頁上的欄位相對應，XML 檔中會將人物資訊包裹在 person 標籤內，並且每個人物有不同的規範碼 (ID)，而



名稱及別名，包攬在 persName 中，別名會以 alternative 作為區分，並且以 lang 代表語言，如以阿彌陀佛為例：

若以阿彌陀佛為例，其中有：名稱、別名、規範碼 ID No、人名群組 Person Group、朝代 Dynasty、是否為歷史人物 Historical Person、註解（精簡） Notes (simple)、以及出處...等。

阿彌陀佛(a mí tuó fó)

別名 Alternative Names :
 [梵文 (羅馬轉寫)] Amita buddha,Amita,Amitāyus,Amitābha
 [藏文 (羅馬轉寫)] Dpag tu med,Dpag yas
 [正體中文] 阿彌多佛,阿彌多佛,阿彌陀佛,阿彌陀,甘露王如來,無量,無量壽佛,無量壽,阿彌陀摩訶,阿彌多度,無量光佛,無量光,阿彌多婆,阿彌陀囉,阿彌陀婆,無量清淨佛,月巴墨佛,無極尊,彌陀,四聖號

規範碼 ID No : A002843
 人名群組 Person Group : 彌陀三聖;
 舊編碼 : G0845
 朝代 Dynasty : 沒有給定朝代;
 是否為歷史人物 Historical Person :

註解(精簡) Notes(simple) : 為西方極樂世界的教主，觀世音、大勢至兩大菩薩為其脅侍，以極樂淨土實踐接引、教化眾生的偉大悲願，又大乘佛教各宗派普遍接受阿彌陀佛，而淨土宗則以專心信仰阿彌陀佛為其主要特色，而後期的中國佛教徒，多以藥師如來主消災延壽、阿彌陀佛主往生事。（佛光：3680、5125；佛百：3173.2；http://zh.wikipedia.org/wiki/阿彌陀佛，2012.04.23）

SameAs :

- Wikidata :
 - https://www.wikidata.org/wiki/Q236242

Occurs in :

- CBETA T50n2059_p0358c26 梁高僧傳: 釋慧遠傳 { 阿彌陀 }
- CBETA T50n2059_p0362b17 梁高僧傳: 釋僧演傳 { 彌陀 }
- CBETA T50n2059_p0362b21 梁高僧傳: 釋僧演傳 { 無量壽佛 }

圖 4.3 DILA 人名規範資料庫查詢結果

在此只用到規範碼 ID No、名稱、別名，因此 XML 要解析的標籤為 person，及 person 中的 persName，如圖 4.4。



```

<person xml:id="A002843" ana="notHistorical">
  <persName xml:lang="zho-Hant">阿彌陀佛</persName>
  <persName xml:lang="san-Latn">Amita buddha</persName>
  <persName xml:lang="bod-Latn">Dpag tu med</persName>
  <persName type="alternative" xml:lang="bod-Latn">Dpag yas</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌多佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿弭跢佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿弭鞞佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌陀</persName>
  <persName type="alternative" xml:lang="san-Latn">Amita</persName>
  <persName type="alternative" xml:lang="zho-Hant">甘露王如來</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量</persName>
  <persName type="alternative" xml:lang="san-Latn">Amitāyus</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量壽佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量壽</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌陀庾斯</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌多度</persName>
  <persName type="alternative" xml:lang="san-Latn">Amitābha</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量光佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量光</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌多婆</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌鞞嚩</persName>
  <persName type="alternative" xml:lang="zho-Hant">阿彌陀婆</persName>
  <persName type="alternative" xml:lang="zho-Hant">無量清淨佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">月巴墨佛</persName>
  <persName type="alternative" xml:lang="zho-Hant">無極尊</persName>
  <persName type="alternative" xml:lang="zho-Hant">彌陀</persName>
  <persName type="alternative" xml:lang="zho-Hant">四聖號</persName>

```

圖 4.4 DILA 人名規範資料庫 xml

以下為資料表的設計：

- seq，表示每筆資料的識別
- id，存放規範碼 ID No（人工加入以 PS 開頭往上遞增，第 1 筆為 PS00001）
- name，存放人物的名稱或別名
- is_alternative，識別是否為別名
- language，存放名稱或別名使用的語言
- name_key，識別同個人物不同的名稱或者別名
- remark，儲存備註



- version，存放資料來源
- create_time，存放此筆資料的創立時間

若是有多個不同語言的名稱則以中文為 name_key=0 的資料，以下以阿彌陀佛為例：

以阿彌陀佛為例的話名稱共 3 種不同語言，並且有其他 24 個別名，因此 name_key 由 0 至 26，表 4.4 為資料庫中的儲存。

| | | | | | | | | |
|-------|---------|--------------|---|----------|----|------|------|-------------------------|
| 95034 | A002843 | 阿彌陀佛 | 0 | zho-Hant | 0 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95035 | A002843 | Amita buddha | 0 | san-Latn | 1 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95036 | A002843 | Dpag tu med | 0 | bod-Latn | 2 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95037 | A002843 | Dpag yas | 1 | bod-Latn | 3 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95038 | A002843 | 阿彌多佛 | 1 | zho-Hant | 4 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95039 | A002843 | 阿彌跢佛 | 1 | zho-Hant | 5 | NULL | DILA | 2022-08-04 14:08:31.190 |
| 95040 | A002843 | 阿彌鞞佛 | 1 | zho-Hant | 6 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95041 | A002843 | 阿彌陀 | 1 | zho-Hant | 7 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95042 | A002843 | Amita | 1 | san-Latn | 8 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95043 | A002843 | 甘露王如來 | 1 | zho-Hant | 9 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95044 | A002843 | 無量 | 1 | zho-Hant | 10 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95045 | A002843 | Amit ā yus | 1 | san-Latn | 11 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95046 | A002843 | 無量壽佛 | 1 | zho-Hant | 12 | NULL | DILA | 2022-08-04 14:08:31.193 |
| 95047 | A002843 | 無量壽 | 1 | zho-Hant | 13 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95048 | A002843 | 阿彌陀廐斯 | 1 | zho-Hant | 14 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95049 | A002843 | 阿彌多廐 | 1 | zho-Hant | 15 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95050 | A002843 | Amit ā bha | 1 | san-Latn | 16 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95051 | A002843 | 無量光佛 | 1 | zho-Hant | 17 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95052 | A002843 | 無量光 | 1 | zho-Hant | 18 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95053 | A002843 | 阿彌多婆 | 1 | zho-Hant | 19 | NULL | DILA | 2022-08-04 14:08:31.197 |
| 95054 | A002843 | 阿彌鞞蟠 | 1 | zho-Hant | 20 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95055 | A002843 | 阿彌陀婆 | 1 | zho-Hant | 21 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95056 | A002843 | 無量清淨佛 | 1 | zho-Hant | 22 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95057 | A002843 | 月巴墨佛 | 1 | zho-Hant | 23 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95058 | A002843 | 無極尊 | 1 | zho-Hant | 24 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95059 | A002843 | 彌陀 | 1 | zho-Hant | 25 | NULL | DILA | 2022-08-04 14:08:31.200 |
| 95060 | A002843 | 四聖號 | 1 | zho-Hant | 26 | NULL | DILA | 2022-08-04 14:08:31.200 |

表 4.4 佛教人物專門詞彙資料於資料庫

資料匯入資料庫則是先刪除舊有來源為 DILA 人名規範資料庫的專門詞彙，並重新解析 XML 匯入專門詞彙，資料庫中共 86773 筆資料，其中共 44667 個人物，共 16611 個人物有別名。



4.2.5 佛教經典專門詞彙

臺大佛學數位圖書館資料庫中已有自己的佛教經典資料表，其中來源包括：CBETA、ACIP、BDK America、SAT DB、VRI、高雄淨宗學會...等，其中本次只取來源為 CBETA 的資料，因此以下舉例將以 CBETA 為例。

以下為資料表的設計：

- seq，表示每筆資料的識別
- lan，存放經典名稱使用的語言
- category，存放類別（例如 CBETA 的 T 為大正新脩大藏經，X 為卍新纂大日本續藏經選錄）
- version，存放資料來源（例如 CBETA 或 SAT DB）
- sutrano，存放經典的編號（例如 CBETA 的長阿含經（22 卷）為 T 類別的第 1 本，因此為 T0001，七佛經（1 卷）為 T 類別的第 2 本，因此為 T0002，人工加入以 SU 開頭往上遞增，第 1 筆為 SU00001）
- filename，存放經典的檔案名稱（例如 CBETA 的長阿含經（22 卷）為 T 類別的第 1 冊第 1 本因此為 T01n0001，七佛經（1 卷）為 T01n0002）
- books，存放經典收錄在第幾冊（例如 CBETA 的長阿含經（22 卷）收錄在第 1 冊，因此為 01，雜阿含經（50 卷）收錄在地 2 冊，因此為 02）
- part，存放經典收錄在哪一個部（例如 CBETA 的長阿含經（22 卷）收錄在阿含部，六度集經（8 卷）收錄在本緣部）
- sutraname，存放經典的名稱
- volumes，表示有幾卷（例：長阿含經（22 卷）即總共 22 卷）



- author，存放作者
- filepath，存放此經典儲存的 URL

以 CBETA 的大正新脩大藏經中的阿含部 T01 的前十冊為例，在資料庫中儲存如表 4.5。

| seq | lan | category | version | sutrano | filename | books | part | sutraname | volumes | author | filepath |
|-----|------|----------|---------|---------|----------|----------|------|-----------------|---------|--------------|---|
| 1 | 1193 | zh | T | CBETA | T0001 | T01n0001 | 01 | 阿含部 長阿含經 | 22 | 後秦 佛陀耶舍共竺佛念譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 2 | 1194 | zh | T | CBETA | T0002 | T01n0002 | 01 | 阿含部 七佛經 | 1 | 宋 法天譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 3 | 1195 | zh | T | CBETA | T0003 | T01n0003 | 01 | 阿含部 毘婆尸佛經 | 2 | 宋 法天譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 4 | 1196 | zh | T | CBETA | T0004 | T01n0004 | 01 | 阿含部 七佛父母姓字經 | 1 | 失譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 5 | 1197 | zh | T | CBETA | T0005 | T01n0005 | 01 | 阿含部 佛說泥洹經 | 2 | 西晉 白法祖譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 6 | 1198 | zh | T | CBETA | T0006 | T01n0006 | 01 | 阿含部 般泥洹經 | 2 | 失譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 7 | 1199 | zh | T | CBETA | T0007 | T01n0007 | 01 | 阿含部 大般涅槃經 | 3 | 東晉 法顯譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 8 | 1200 | zh | T | CBETA | T0008 | T01n0008 | 01 | 阿含部 大堅固婆羅門緣起經 | 2 | 宋 施護等譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 9 | 1201 | zh | T | CBETA | T0009 | T01n0009 | 01 | 阿含部 人仙經 | 1 | 宋 法賢譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |
| 10 | 1202 | zh | T | CBETA | T0010 | T01n0010 | 01 | 阿含部 白衣金幢二樂羅門緣起經 | 3 | 宋 施護等譯 | http://buddhism.lib.ntu.edu.tw/FULLTEXT/sutra/T/... |

表 4.5 佛教經典專門詞彙資料於資料庫

資料庫中 version 為”CBETA”的共 4515 個詞彙。

4.2.6 資料處理實作

每一次進行總體的更新標記時，需先檢查書目欄位、停用詞資料表及專門詞彙的資料表是否存在，刪除原有的專門詞彙，重新解析並再匯入更新後的專門詞彙（因為原有專門詞彙的來源資料庫如 DILA 會持續更新，隨著資料庫的更新也必須更新收錄的專門詞彙），圖 4.5 為整個流程，因為每個專門詞彙的儲存方式不一樣（佛教宗派詞組為 excel，佛教人物詞組為 XML），因此整個解析的程式需要分開寫（圖中藍色部分）。

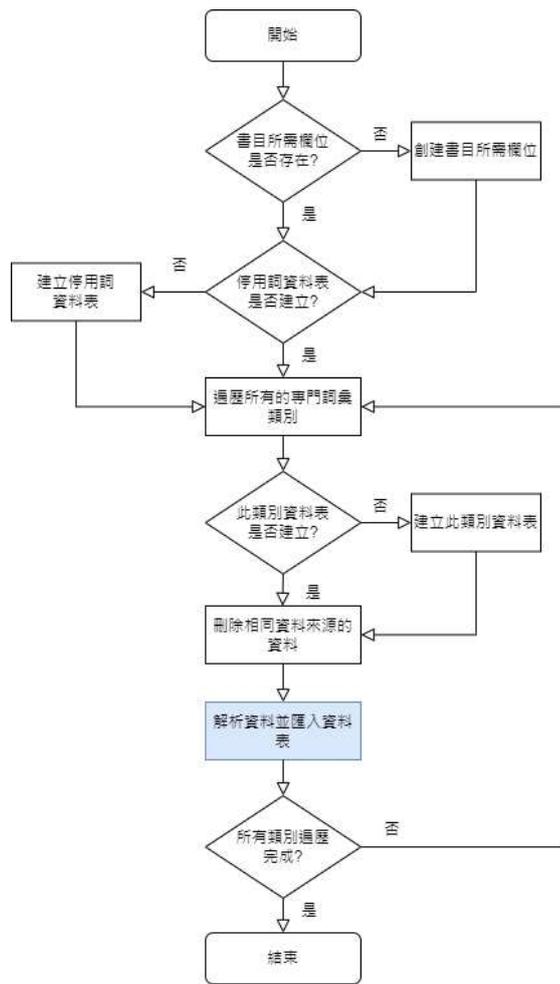


圖 4.5 資料匯入流程

4.2.7 小結

各類別專門詞彙長度為 1 的資料較容易產生錯誤標記的情況，因此將長度為 1 的專門詞彙篩選掉，統整為表 4.6。

| 類別 | 匯入資料數量 | 篩選後數量 | 篩選後重名數量 |
|----------|--------|-------|---------|
| 佛教宗派專門詞彙 | 45 | 45 | 11 |
| 佛教人物專門詞彙 | 86773 | 85769 | 6911 |
| 佛教經典專門詞彙 | 4514 | 4513 | 220 |

表 4.6 各類別資料處理統計



4.3 自動標記實作

整體可以分為兩個部分：對已收錄的書目進行標記更新、對新匯入的書目進行標記或被更新的舊有書目進行標記更新。

(一) 對已收錄的書目進行標記更新

為已收錄的書目進行標記更新並且更新後分類欄位，而人工標記的優先度大於自動標記（若是自動標記與人工標記重疊到，則不會取自動標記的結果），且比對到的標記會去已刪除的標記中尋找，若存在已刪除中的標記中則不會被標記，若此次自動標記結果與目前儲存的自動標記結果不同則會重製人工檢核時間，圖 4.6 為整個流程。

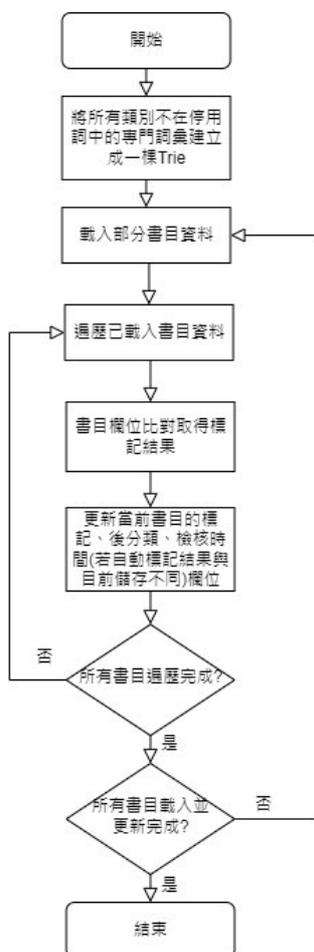
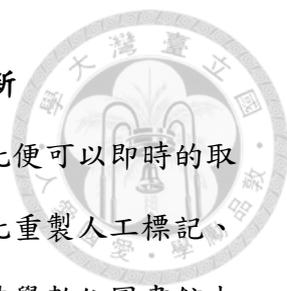


圖 4.6 已收錄的書目標記流程



(二) 對新匯入的書目進行標記或被更新的舊有書目進行標記更新

對於新匯入的書目或更新舊有書目欄位做即時的比對，如此便可以即時的取得標記更新，並且因為是新匯入的書目或者被更新的書目，因此重製人工標記、已刪除的標記、人工檢核時間，對於匯入書目及更新書目臺大佛學數位圖書館本身就有一個流程了，在不破壞其本身的流程下將此流程插入臺大佛學數位圖書館匯入書目的流程中，圖 4.7 為整個流程。

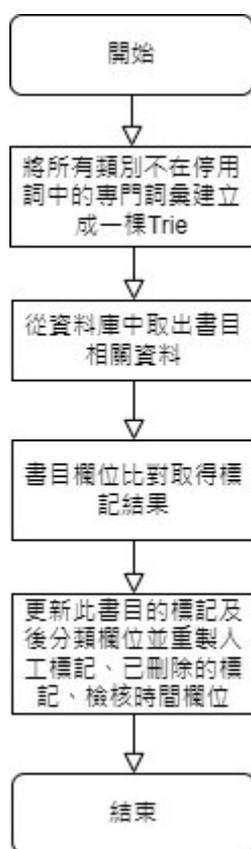


圖 4.7 新匯入或更新之書目比對流程

(三) 小結

對於如果要加入新的專門詞彙類別（例：地名），只需在各個流程的變數中新增變數，但專門詞彙的解析程式仍需要重新實作，或直接在資料庫中建立資料表並匯入專門詞彙。

因專門詞彙的來源會進行持續更新，因此應定時排程更新專門詞彙及進行標記更新。



4.4 輔助人工檢核

為了提高自動標記的正確性，採取輔助人工檢核的方法。輔助人工檢核，整體可以分為兩個部分：人工標記檢核平台、專門詞彙管理平台，人工標記檢核平台進行單筆書目的人工標記檢核，專門詞彙管理平台進行停用詞以及各個專門詞彙的管理。

(一) 人工標記檢核平台

對於自動標記的結果，可以從臺大佛學數位圖書館後台進入人工標記檢核平台，對單筆書目進行操作。平台能做的操作有：加入新的標記、將標記移除、將詞彙加入停用詞、將詞彙加入專門詞彙清單，操作並保存後會即時進行該筆書目的自動標記比對，若標記有變動會即時反饋給檢核人員，圖 4.8 為整個流程。

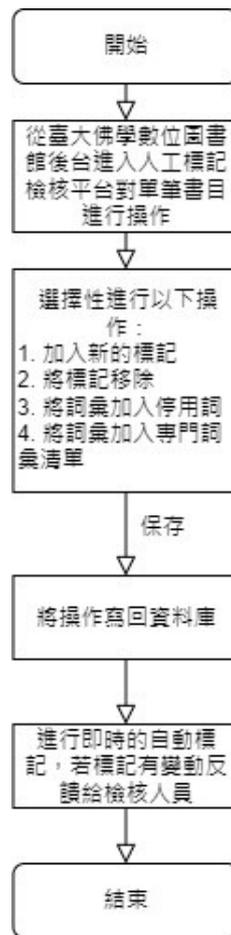


圖 4.8 人工檢核平台操作流程

(二) 專門詞彙管理平台

提供專門詞彙管理平台去編輯各類別的專門詞彙及停用詞，因專門詞彙的數量較多無法一次載入前端，因此提供兩種查詢方式：以 500 為單位的範圍查詢（將人工加入的詞彙皆放在最後面）、以部分符合的關鍵詞查詢，並提供以下操作：將非人工加入的專門詞彙加入停用詞、將人工加入的專門詞彙或停用詞刪除、人工新增專門詞彙或停用詞，圖 4.9 為整個流程。

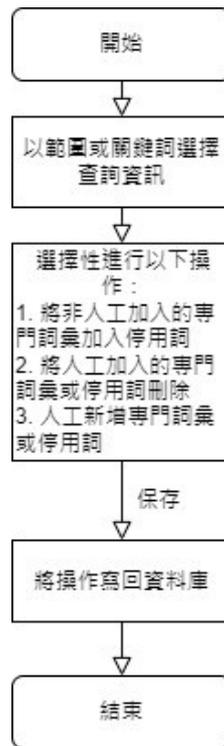


圖 4.9 專門詞彙管理平台操作流程

(三) 小結

檢核平台是為自動標記輔以人工作業方式，可以檢核標記也可進行人工標記，並且搭配專門詞彙管理平台可以更加完善各類別的專門詞彙及停用詞，使得不斷更新後的自動標記成果能夠越來越準確。



第五章 UI 功能設計與範例演示

5.1 輔助人工檢核範例演示

分別演示人工檢核平台及專門詞彙管理平台操作流程。

(一) 人工檢核平台

從臺大佛學數位圖書館後台可以進入人工檢核平台，查看單筆書目目前的標記（標記紅色為佛教宗派專門詞彙、藍色為佛教人物專門詞彙、綠色為佛教經典專門詞彙），以序號為 126449 的書目為例。

| 全部類別 | |
|------|--|
| 全部類別 | 佛教宗派 佛教人物 佛教經典 未檢核 |
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是重要，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顓=智者大師=天台大師; 摩訶止觀 |
| 目次 | |

圖 5.1 人工檢核平台操作範例 1

點選當前的標記可以刪除，而勾選加入 Stopword 後再點刪除，則是將此詞彙加入停用詞當中並且刪除標記。

| 全部類別 | |
|------|--|
| 全部類別 | 佛教宗派 佛教人物 佛教經典 未檢核 |
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是重要，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Budd; 佛教人物: 實際 <input checked="" type="checkbox"/> 加入 Stopword <input type="button" value="刪除"/> 天台大師; 摩訶止觀 |
| 目次 | |

圖 5.2 人工檢核平台操作範例 2



選取文字能夠進行標記，而勾選加入 list 則是將此詞彙加入專門詞彙的資料表當中，而關鍵字的標記規則跟自動標記一樣，若是選取的範圍不是完全比對則無法標記（這個例子選取”法華”不會產生標記選項）。

全部類別 佛教宗派 佛教人物 佛教經典 未檢核

| | |
|-----|--|
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是重要，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顓=智者大師=天台大師; 摩訶止觀 |
| 目次 | |

加入list資訊
Stopwords : 實際

保存

圖 5.3 人工檢核平台操作範例 3

右邊是要加入專門詞彙或者停用詞的詞彙，點選垃圾桶的圖示能夠刪除加入停用詞或專門詞彙的操作，操作完成後必須按下方的保存才會將結果寫回資料庫當中。

全部類別 佛教宗派 佛教人物 佛教經典 未檢核

| | |
|-----|--|
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是重要，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顓=智者大師=天台大師; 摩訶止觀 |
| 目次 | |

加入list資訊

- 佛教宗派 : 天臺宗
- 佛教宗派 : 法華宗
- 佛教宗派 : Tien-tai Buddhism
- 佛教宗派 : Tendaishu
- Stopwords : 實際

保存

圖 5.4 人工檢核平台操作範例 4



保存後會先將當前書目的標記資訊更新，再把需要加入的停用詞及專門詞彙寫入資料庫，最後針對該筆書目再進行一次自動標記，若標記有變動則會反饋給檢核人員。

保存成功, 標記有變動!

全部類別 佛教宗派 確定

| | |
|-----|---|
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是非常重要的，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顛=智者大師=天台大師; 摩訶止觀 |
| 目次 | |

保存

加入list資訊

佛教宗派 : 天臺宗

佛教宗派 : 法華宗

佛教宗派 : Tien-tai Buddhism

佛教宗派 : Tendaishu

Stopwords : 實際

圖 5.5 人工檢核平台操作範例 5

因為”天臺宗”被加入了專門詞彙當中，因此在摘要中未被標記的”天臺宗”也被標記起來，而已刪除的標記也不會再次被標記，右上角為最近一次的保存時間。

全部類別 佛教宗派 佛教人物 佛教經典 2022-08-10 15:51:37 保存

| | |
|-----|---|
| 題名 | 天臺智者大師摩訶止觀實踐之道 |
| 摘要 | 天臺宗是中國佛教八大宗之一，也是中國所創建的佛學之一。它在學術上的地位是非常重要的，與禪宗、華嚴宗並駕齊驅，尤具有優越之地位。故歷代學者均喜以天臺著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於天臺宗之認識，更為實際而重要。 |
| 關鍵詞 | 天台宗=法華宗=Tien-tai Buddhism=Tendaishu; 智顛=智者大師=天台大師; 摩訶止觀 |
| 目次 | |

保存

加入list資訊

圖 5.6 人工檢核平台操作範例 6



(二) 專門詞彙管理平台

專門詞彙管理平台用來管理總體的專門詞彙表單，左上角會先顯示各類別專門詞彙及停用詞的總數量及人工數量，因為數量太多無法一次載入，因此提供兩種方式查詢：以 500 為單位的範圍查詢（將人工加入的詞彙皆放在最後面）、以部分符合的關鍵詞查詢。



圖 5.7 專門詞彙管理平台操作範例 1

查詢後可以用勾選框對查詢結果進行篩選，之後對查詢結果進行操作：將非人工加入的專門詞彙加入停用詞、將人工加入的專門詞彙或停用詞刪除。



圖 5.8 專門詞彙管理平台操作範例 2



右上角的新增按鈕能夠以人工方式將詞彙加入專門詞彙類別或者停用詞當中。

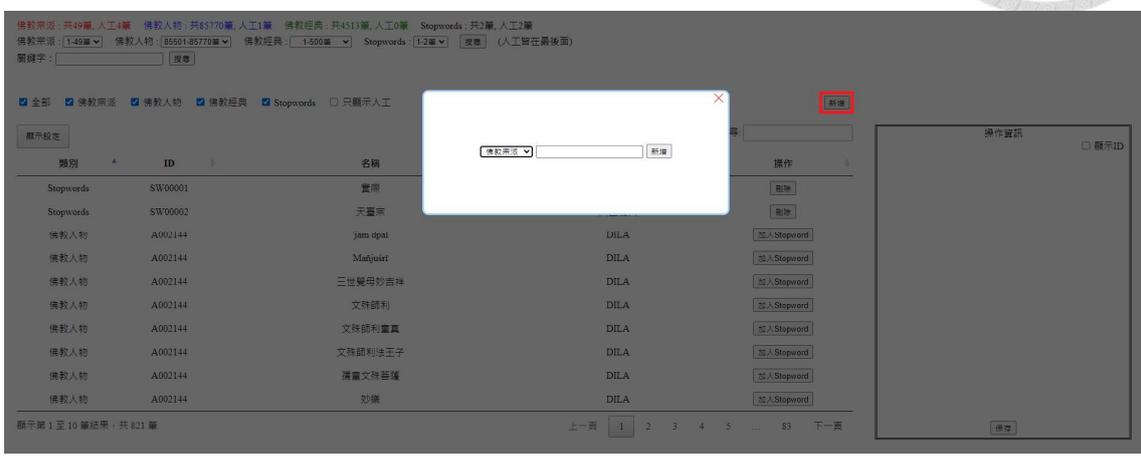


圖 5.9 專門詞彙管理平台操作範例 3

操作資訊框中可以點選右邊的垃圾桶圖示刪除操作資訊，右上角可以顯示目前操作資訊的 ID，操作完成後仍須按下方保存才能儲存操作資訊，會回饋給使用者操作資訊是否保存成功。



圖 5.10 專門詞彙管理平台操作範例 4



5.2 書目後分類及標記呈現

有關書目後分類及標記的呈現，我們採用範例演示。使用臺大佛學數位圖書館的書目檢索系統在題名部分查找”戒律”，圖 5.11 為搜尋結果，有 6 種既有後分類，而後分類顯示方式為一層含有多個可選的勾選框，只有出版年份可以選擇一個年份的範圍做篩選。

題名(書名) 戒律 查詢 清除

檢索策略: (戒律.ti)

634 筆查詢結果

書目匯出 1 / 32 GO

相關經律論

- 佛說目連問戒律中五百經重事經
- 佛說目連問戒律中五百經重事經 [全文]
- 佛說目連問戒律中五百經重事
- 戒律傳來記
- 佛說目連問戒律中五百經重事

顯示更多

DILA 佛學術語字辭典

- 戒律
- 嚴淨戒律
- 戒律主義者

顯示更多

相關詞

載入中...

1 1915 - 2021 篩選

2021 (5)

2020 (2)

2019 (3)

2018 (4)

2017 (23)

篩選 顯示更多

資料類型

出處題名

關鍵詞

著者

- 釋普圓 (24)
- 土橋秀高 (19)
- 石田補堂 (14)
- 釋靈嚴 (12)
- 平川彰 (11)
- 釋獻如 (11)

篩選 顯示更多

語言

- 中文 (344)
- 日文 (286)

1 “律分五部”與中古佛教對戒律史的知識史建構 [全文]

王磊

《中國哲學史》, n.4 (2015), 37 - 45+86

點閱次數: 55

2 〈現代禪的戒律觀〉對談與回應

陳美華

《世界宗教學刊》, n.3 (2004.06), 83 - 85

點閱次數: 265

3 《中阿含經》之戒律觀

鄧先銘

《中國佛教》, v.32 n.9-10 (1988.10), 10 - 17

點閱次數: 74

4 《涅槃經》戒律思想之探討

釋文慧

《慧苑佛學年刊》, 創刊號 (1982), 197 - 210

點閱次數: 129

5 《慧光基誌》所見北朝戒律學的展開 [全文]

陳志遠 (著)

《藝林園地佛學學刊》, v.3 n.1 《宗教石刻研究與玄奘研究》專刊 (2020.04), 1 - 33

點閱次數: 15

6 「佛光新戒條」與戒律的現代化轉型 [全文]

夏德美 (著)

圖 5.11 書目後分類及標記呈現範例 1

新增的「提及：佛教宗派」、「提及：佛教人物」、「提及：佛教經典」後分類，皆是設計成兩層結構，第一層為標籤名稱，第二層為比對到的來源欄位。

查詢結果: 戒律

檢索策略: (戒律 1)

634 筆查詢結果

1 書目匯出

1 32 GO

相關經律論

佛說目連問戒律中五百經重華經
佛說目連問戒律中五百經重華經
佛說目連問戒律中五百經重華經
戒律傳承記
佛說目連問戒律中五百經重華經

顯示更多

DILA 佛學術語字彙典

戒律
戒律戒律
戒律三義書

顯示更多

相關詞

載入中...

載入中...

提及:佛教宗派

律宗 (23) ▾

華名 (6)

禪宗 (9)

顯密經 (12)

目次 (1)

禪宗 (15) ▾

天台宗 (7) ▾

曹洞宗 (5) ▾

真宗 (5) ▾

法華宗 (4) ▾

篩選 顯示更多

提及:佛教人物

提及:佛教經典

1 “律分五部”與中古佛教對戒律史的知識史建構 [全文]
王義
《中國哲學史》, n.4 (2015), 37 - 45+86
點閱次數: 55

2 〈現代禪的戒律觀〉對談與回應
陳典華
《世界宗教學刊》, n.3 (2004.06), 83 - 85
點閱次數: 265

3 《中阿含經》之戒律觀
鄧先銘
《中國佛教》, v.32 n.9-10 (1988.10), 10 - 17
點閱次數: 74

4 《涅槃經》戒律思想之探討
釋文慧
《華梵佛學年刊》, 創刊號 (1982), 197 - 210
點閱次數: 129

5 《慧光基誌》所見北朝戒律學的展開 [全文]
陳志遠 (著)
《華林國際佛學學刊》, v.3 n.1 《宗教石刻研究與玄奘研究》專刊 (2020.04), 1 - 33
點閱次數: 15

6 “佛光新戒條”與戒律的現代化轉型 [全文]
夏德典 (著)
《人間佛教學報 藝文》, n.31 (2021.01), 64 - 79

圖 5.12 書目後分類及標記呈現範例 2

點選列會觸發勾選框選取，而點選三角形將會展開第二層，若勾選第一層的標籤第二層將一併被選起（第二層全選第一層也將被選起），取消第一層勾選將取消第二層全部勾選，若取消第二層勾選則第一層也一併被取消。

查詢結果: 戒律

檢索策略: (戒律 1)

634 筆查詢結果

1 書目匯出

1 32 GO

相關經律論

佛說目連問戒律中五百經重華經
佛說目連問戒律中五百經重華經
佛說目連問戒律中五百經重華經
戒律傳承記
佛說目連問戒律中五百經重華經

顯示更多

DILA 佛學術語字彙典

戒律
戒律戒律
戒律三義書

顯示更多

相關詞

載入中...

載入中...

提及:佛教宗派

律宗 (23) ▾

華名 (6)

禪宗 (9)

顯密經 (12)

目次 (1)

禪宗 (15) ▾

華名 (5)

禪宗 (6)

顯密經 (5)

天台宗 (7) ▾

曹洞宗 (5) ▾

真宗 (5) ▾

法華宗 (4) ▾

篩選 顯示更多

提及:佛教人物

提及:佛教經典

1 “律分五部”與中古佛教對戒律史的知識史建構 [全文]
王義
《中國哲學史》, n.4 (2015), 37 - 45+86
點閱次數: 55

2 〈現代禪的戒律觀〉對談與回應
陳典華
《世界宗教學刊》, n.3 (2004.06), 83 - 85
點閱次數: 265

3 《中阿含經》之戒律觀
鄧先銘
《中國佛教》, v.32 n.9-10 (1988.10), 10 - 17
點閱次數: 74

4 《涅槃經》戒律思想之探討
釋文慧
《華梵佛學年刊》, 創刊號 (1982), 197 - 210
點閱次數: 129

5 《慧光基誌》所見北朝戒律學的展開 [全文]
陳志遠 (著)
《華林國際佛學學刊》, v.3 n.1 《宗教石刻研究與玄奘研究》專刊 (2020.04), 1 - 33
點閱次數: 15

6 “佛光新戒條”與戒律的現代化轉型 [全文]
夏德典 (著)
《人間佛教學報 藝文》, n.31 (2021.01), 64 - 79

圖 5.13 書目後分類及標記呈現範例 3



每種類別只會顯示最多 6 筆的標籤，多出來的標籤則要點選旁邊的顯示更多，以離左邊勾選框的偏移長度區分第一及第二層，名稱旁邊的勾選框會全選全部選項，而原本的選項及顯示更多面板的選項將是同步的。



圖 5.14 書目後分類及標記呈現範例 4

選取資料類型的”期刊論文”、「提及：佛教宗派」的”律宗-關鍵詞”及”律宗-摘要”、「提及：佛教人物」”道宣-關鍵詞”，已選的篩選標籤將顯示在左上角並且依不同類別分開顯示，而這樣的篩選結果會是，資料類型（”期刊論文”）& 「提及：佛教宗派」（”律宗-關鍵詞”|”律宗-摘要”）& 「提及：佛教人物」（“道宣-關鍵詞”），”&”為交集、”|”為聯集。



圖 5.15 書目後分類及標記呈現範例 5

點選單筆書目可以看大致資訊，所有類別的標記都用黃色底色標記起來，而查詢詞則用紅字表示。



圖 5.16 書目後分類及標記呈現範例 6

單筆書目檢索頁面中若有查詢詞也是用紅字表示，所有類別的標記也是用黃底來表示。



進階查詢 · 查詢歷史 · 使用說明

題名(書名) ▾

閱讀本文

作者或出版者未授權
無法提供閱讀

加值服務

WorldCat

META-CAT+
即時跨館聯合查詢

國立臺灣大學圖書館
National Taiwan University Library

佛教圖書館
館藏聯合目錄

NDDS

書目管理

書目匯出

天臺智者大師摩訶止觀實踐之道

作者 釋真琳

出版日期 1983

出版地 香港, 中國 [Hong Kong, China]

資料類型 博碩士論文=Thesis and Dissertation

使用語言 中文=Chinese

學位類別 碩士

校院名稱 香港珠海大學

系所名稱 中國文學研究所

指導教授 涂公遂

畢業年度 1983

關鍵詞 **天臺宗=法華宗=Tien-tai Buddhism=Tendaishu; 智者大師=天台大師; 摩訶止觀**

摘要 **天臺宗**是中國佛教八大宗之一，也是中國所創建的佛學之一，它在學術上的地位是**非常重要**，與**禪宗**、**華嚴宗**並駕齊驅，尤具有優越之地位。故歷代學者均喜以**天臺**著述為研究對象，然在研究記載中，一般學者的研究，皆側重在理論上的檢討，而忽略了實踐的重要部份。因此，本人的研究，著重在實踐之道，這對於**天臺宗**之認識，更為實際而重要。

點開次數 247

建檔日期 2005.09.23

更新日期 2015.10.15

圖 5.17 書目後分類及標記呈現範例 7

5.3 本章回顧

本章演示如何操作人工檢核平台及專門詞彙管理平台，書目檢索系統演示 UI 的功能及後分類與標記的呈現。

第六章 結論與未來工作



6.1 研究結論

本研究採取自動標記合併人工檢核方式，為台大佛學圖書館所收錄的書目資料，新增佛教相關專門詞彙標記，並整合於台大佛學數位圖書館的書目檢索系統中，除了提供專門詞彙標記的後分類外，也於書目資訊中的突顯查詢詞及比對到的專門詞彙；此外，也為人工檢核方式，提供操作的介面，以利台大佛學數位圖書館的後端作業。

在本研究中，自動標記所做的方法因為收集的專門詞彙仍有許多較為通用的詞彙，因此仍需搭配停用詞才能使得標記較為準確，而對於沒收錄進資料庫的詞彙亦能通過人工檢核加入專門詞彙資料表中，因此即使無法對於所有的書目都進行人工檢核，亦能讓其餘書目的標記也更加準確。

與臺大佛學數位圖書館原有的後分類不同的地方在於，這種模式新增的後分類皆不是 metadata 既有欄位的後分類，而是經處理過後所產生的後分類，因為提前對書目做出了標記，並為使用者的檢索結果提供後分類，讓使用者不需在大量資料中自行做特定名詞的篩選。這種做法將可以使得篩選變得更為方便也可以通過不同的後分類標籤找尋到一些關聯的資訊（例：哪些人物或宗派常常一起出現...等）。

後分類前端及 UI 部分因為臺大佛學數位圖書館已有自己的後分類架構，而此次所做的後分類是兩層可展開的方式，因此與原本的架構稍有不同，需要在不影響原本架構上做更改，設計上具有彈性且貼近原先架構使維護跟更新上較為方便，若往後有嘗試新增的標記類別亦可以沿用相同架構。

本研究為臺大佛學數位圖書館創建的自動標記的程序及方法，日後如有需求新增新的詞彙類別，皆可援用。



6.2 未來工作

本研究雖然整體架構已經完成，但仍有些可以延伸思考的地方，人名資料取自 DILA 人名規範資料庫，但資料表欄位我只設計了最基本的幾個欄位，而 DILA 人名規範資料庫的資料提供許多額外的資訊，也許可以將資料表的欄位做的完整並且可以有些其他應用，佛教宗派欄位也可以將欄位設計得更加完整（例如是屬於哪個國家的宗派）。

目前在前端已經有後分類篩選了，在前端後分類的專門詞彙上做一些額外的操作，可以給使用者更詳細的人名資訊或者連結到 DILA 人名規範資料庫做進一步的參考查詢，經典也可以連結到可閱讀的頁面，若是使用者不了解這個專門詞彙的話也可以帶給使用者一些線索。

利用標記的統計資訊去給使用者一些推薦，或給使用者一些提示，讓使用者能夠找到更多的聯繫，並將一些資料統計視覺化做成一些圖表資訊，讓使用者能夠一目了然目前查詢資料的分布。

嘗試加入新的類別進入當前架構並觀察標記結果，也可以嘗試不同的後分類策略或不同的標記方式並比較與目前的標記方式及後分類差異。

參考文獻



- [1] 國立臺灣大學數位人文研究中心。(1995)。國立臺灣大學佛學數位圖書館。國立臺灣大學。2022年6月檢自：http://doi.airiti.com/LandingPage/NTURCDH/10.6681/NTURCDH.DB_DL_MBS/Collection
- [2] 廖昱甄(2022)。個人化書目管理應用工具——以臺大佛學數位圖書館為例（未出版碩士論文）。國立臺灣大學，臺北市。
- [3] 香光尼眾佛學院圖書館(2011)。佛教圖書分類法 2011 年版（初版）。嘉義市：香光書鄉。
- [4] 藍文欽(2011)。《佛教圖書分類法(2011 年版)》評介。佛教圖書館館刊，53，48-58。
- [5] DILA 人名規範資料庫。2022 年 6 月檢自：<https://authority.dila.edu.tw/person/>
- [6] CBETA 中華電子佛典協會。2022 年 6 月檢自：<https://cbeta.org/>
- [7] Markus 半自動標記平台。2022 年 6 月檢自：<https://dh.chinese-empires.eu/markus/beta/index.html>
- [8] 何浩洋(2014)。MARKUS：中文古籍文本半自動標記平台。第五屆數位典藏與數位人文國際研討會, pp.117-137, 2014.
- [9] Maabar, Maha (17 November 2014). Trie Data Structure. Retrieved June, 2022, from <https://bioinformatics.cvr.ac.uk/trie-data-structure/>
- [10] Microsoft. Clustered and Nonclustered Indexes Described. Retrieved June, 2022, from [https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/ms190457\(v=sql.110\)](https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2012/ms190457(v=sql.110))
- [11] Bayer, R., & McCreight, E. (2002). Organization and maintenance of large ordered indexes. In Software pioneers (pp. 245-262). Springer, Berlin, Heidelberg.