

AI在資訊組織的應用發展

陳亞寧

國立臺灣師範大學圖書資訊學研究所

115 年度「圖書館資訊組織進階訓練」研習班
國家圖書館藝文中心 4 樓 421 室

21 August, 2026

大綱

- 自動化主題索引發展的回溯
- 七篇AI在資訊組織的文獻
 - 主題編目
 - 入門型
 - 翻譯
 - 記述編目
 - 實體識別與鏈結資料
- 實徵個案：MARC→實體萃取→實體識別→URI→知識圖譜
- 結語

自動化主題索引：概述／自動化主題索引

自動化主題索引：概述

Golub, K. (2021). Automated subject indexing: An overview. *Cataloging & Classification Quarterly*, 59(8), 702-719, <https://doi.org/10.1080/01639374.2021.2012311>

自動化主題索引

- 面對文件量持續增加，全球圖書館愈來愈多地探索半自動或自動化主題索引方法。
- 這有助於維持書目目標、豐富後設資料，並在不同館藏文件之間建立更多連結，進而改善資訊檢索與取用。

(Golub, 2021, p. 702)

缺乏通行方法／高品質主題取用

缺乏通行方法

- 然而，能在實際作業系統中運作且被普遍接受的自動化方法仍然不足。
- 本文旨在概述自動化主題索引的基本原理、主要方法、在實際圖書館系統中的可能應用、現有運作案例，以及仍需進一步研究的相關挑戰。

(Golub, 2021, p. 702)

高品質主題取用

- 在網路文化資產館藏與跨庫搜尋服務中提供高品質主題取用已被證明相當困難，因而造成不佳的使用者服務。
- 原因可能是在機構預算緊繃的情況下，文件數量快速增加。

(Golub, 2021, p. 702)

自動化建議／作業系統中的應用

自動化建議

- 自動化建議可提供給資訊專業人員、將作品存入典藏庫的作者，或進行標籤與註釋的讀者。
- 除了處理規模與永續問題，自動化主題索引也可用於豐富既有後設資料、建立資訊資源之間更多連結，並提升後設資料一致性。

(Golub, 2021, p. 702)

作業系統中的應用

- 然而，目前仍沒有廣泛可取得、經充分研究且證明能在圖書館作業系統中運作的軟體套件。
- 少數案例出現在專門圖書館，如NASA、美國國家醫學圖書館與美國國家農業圖書館。

(Golub, 2021, p. 703)

背景／指派式標引

背景

- 本文旨在概述自動化主題索引並凸顯最重要發展，以供本領域最佳實務與後續研究參考。

(Golub, 2021, p. 703)

指派式標引

- 本文聚焦於從既有知識組織系統（KOS）進行索引，例如主題標目系統、資訊檢索詞表/索引典或分類系統。
- 本文排除衍生式標引；衍生式標引是從被標引文件中抽取用詞並作為索引詞。

(Golub, 2021, p. 703)

知識組織系統／圖書館與文本文件

知識組織系統

- 另一方面，若 KOS 的使用者介面良好，便可協助處理這些問題，因此在作業資訊系統的主題索引脈絡中最具價值。
- 本文將主題索引用於較廣義的意涵，也包含主題分類。

(Golub, 2021, p. 703)

圖書館與文本文件

- 雖然本文聚焦於圖書館與文本文件脈絡，所描述與討論的許多內容也適用於相關記憶機構與多媒體等其他格式。
- 不過，不同資訊物件仍具有特殊特性，無論人工主題索引或自動化方法都必須納入考量。

(Golub, 2021, p. 704)

基本原理／三個步驟

基本原理

- 此 ISO 將資訊專業人員進行的主題索引定義為包含三個步驟的過程。

(Golub, 2021, p. 704)

三個步驟

- 該 ISO 將資訊專業人員的主題索引定義為三個步驟：判定文件主題內容；進行概念分析以決定應表徵哪些面向；將這些概念或面向轉譯到 KOS。
- 自動化主題索引即為機器式主題索引，其中這三個步驟的人類智識過程由計算語言學與統計等技術取代。

(Golub, 2021, p. 704)

停用詞／斷詞

停用詞

- 一般而言，計算技術會先用於準備待標引文件，此程序稱為前處理。
- 停用詞移除指移除介詞、連接詞、冠詞與限定詞等常見詞，這些詞相較名詞、動詞與形容詞，較少承載文件主題資訊。

(Golub, 2021, p. 704)

斷詞

- **斷詞**是將文本切分為個別詞項（tokens）；這些詞項可保留為單詞（unigrams）、與相鄰詞配對（bigrams），或結合更多詞。
- 因此，前處理會產生由抽取詞項列表構成的文件表徵。

(Golub, 2021, pp. 704 – 705)

詞幹化或詞形還原／詞頻—逆文件頻率

詞幹化或詞形還原

- 詞幹化或詞形還原可處理本質上指涉相同意義但呈現不同形式的詞。
- 另一方面，詞形還原依賴更深層的形態分析，以抽取實際且有意義的詞彙基本形式。

(Golub, 2021, p. 705)

詞頻—逆文件頻率

- 權重依不同啟發式原則計算；常用的是 tf-idf（詞頻—逆文件頻率）公式。
- tf-idf 的啟發式邏輯是：在特定文件中頻繁出現，但整體上少見的詞，更能指示文件主題，因此應給予較高權重。

(Golub, 2021, p. 705)

文件表徵與索引詞／Transformer 語言模型

文件表徵與索引詞

- 同時，目標索引詞也以不同啟發式原則表徵，以便在最後步驟中讓文件表徵與索引詞表徵進行匹配。
- 一旦文件表徵與目標 KOS 中的潛在索引詞匹配，系統便會指派權重以表示該詞相關的信心程度。

(Golub, 2021, p. 705)

Transformer 語言模型

- 上述技術屬於傳統自然語言處理（NLP）方法的典範。
- 在最新深度學習模型（如 Transformer 語言模型）中，步驟並不完全相同。

(Golub, 2021, p. 705)

常見方法／訓練文件

常見方法

- 自動化主題索引最常見的方法常稱為**文本分類**或**文本分類法**，即應用**監督式機器學習演算法**。

(Golub, 2021, p. 706)

訓練文件

- 此方法中，演算法根據已由人工以特定索引詞標引的文件特徵，「學習」目標索引詞的特性；這些文件稱為**訓練文件**。
- 接著演算法會以館藏中的新文件集進行測試，這些文件稱為**測試文件**。

(Golub, 2021, p. 706)

分類器與集成／大量訓練文件

分類器與集成

- 建立監督式機器學習分類器有許多方法，例如支援向量機、人工神經網路、隨機森林、適應性增強、線性模型、樹狀方法，以及近年的Transformer 深度學習模型。
- 此外，也可結合兩種以上不同分類器來做出分類決策。

(Golub, 2021, p. 706)

大量訓練文件

- 監督式機器學習方法(supervised machine learning algorithms)要求每個目標索引詞或類別都具有相對大量的訓練文件。
- 然而，許多文件館藏可用的訓練文件太少，導致分類器難以有良好表現。

(Golub, 2021, p. 706)

大型 KOS／重用智識成果

大型 KOS

- 在這些情況下，可採用半監督式學習。
- 此外，在 DDC 等大型 KOS 中，既有文件在類別間分布高度偏斜，因此通常只有 DDC 前三層能相對良好地計算。

(Golub, 2021, p. 706)

重用智識成果

- 另一種處理訓練文件不足並同時重用建立 KOS 智識成果的方法，是將待標引文件中的詞與 KOS 的目標索引詞進行字串比對。
- 文獻中稱之為詞彙式或字串比對方法，也稱為規則式方法(rule-based approach)。

(Golub, 2021, p. 706)

詞表內建關係／非監督式機器學習方法

詞表內建關係

- 在此方法中，系統探索 KOS 內建關係，並取得有利結果。
- 詞表內建關係對自動化主題索引特別具有潛力，但成功取決於詞表特性與文件館藏。

(Golub, 2021, pp. 706 – 707)

非監督式機器學習方法

- 相關方法還包括文件分群，這是一種非監督式機器學習方法 (unsupervised machine learning approach)，其中目標 KOS，也就是其詞彙與詞間關係，會自動衍生。
- 文件分群可能較適合其他任務，例如組織網路搜尋引擎結果。

(Golub, 2021, p. 707)

複雜的認知與社會現象／黃金標準問題

複雜的認知與社會現象

- 文本是一種複雜的認知與社會現象；文本理解涉及多種知識來源、維持多重推論，並包含個人詮釋。
- 主要挑戰在於處理語意，同時掌握特定時間點理解的完整脈絡，包括領域知識、意識形態、規範、作者與使用者背景，以及使用文本的目的。

(Golub, 2021, p. 712)

黃金標準問題

- 其中一種方法是將自動指派的主題索引詞與既有人工指派詞（作為「黃金標準」）比較，但此方法存在問題。
- 既有後設資料記錄不應作為唯一的「黃金標準」。

(Golub, 2021, p. 712)

半自動解決方案／近期研究議程

半自動解決方案

- 然而，仍存在與更高層次語意理解計算及相關評估議題有關的挑戰，因此對作業資訊系統而言，只能建議採用半自動解決方案。
- 本文旨在概述自動化主題索引，聚焦於從既有 KOS 取用索引詞的指派式標引，以協助改善圖書館資訊系統中的資訊檢索。

(Golub, 2021, p. 713)

近期研究議程

- 由於圖書館似乎很少發表關於其作業系統的學術文章，針對已實作解決方案的調查性文章應列入近期研究議程。
- 未來研究也應在不同脈絡、文件類型、KOS 類型與主題領域中進行測試，以判定何種 KOS 及其特性可支援自動化方法的表現。

(Golub, 2021, p. 713)

大型語言模型與編目：ChatGPT 與 Copilot 如何指派主題標目與索書號

大型語言模型與編目：ChatGPT 與 Copilot 如何指派主題標目與索書號

- Holstrom, C. (2025). Large language models (LLMs) and cataloging: Exploring how ChatGPT and Copilot assign subject headings and call numbers. Proceedings from North American Symposium on Knowledge Organization, 10.
<https://doi.org/10.7152/nasko.v7i1.95648>

文章焦點

- LLM 在需要領域知識的語言與知識密集任務中展現能力。
- 作者因此檢驗通用 LLM 是否能支援主題編目任務。

(Holstrom, 2025, p. 1)

測試系統

測試系統

- 研究比較 ChatGPT 與 Copilot。
- 測試範圍涵蓋 LCSH、Sears、LCC、DDC、Cutter 號與 MARC 欄位。

(Holstrom, 2025, p. 1)

主要發現／建議角色

主要發現

- LLM 有自動編目潛力，但仍有多種嚴重缺點。
- 問題包括不夠明確、使用非權威詞彙、標目組合與 MARC 格式錯誤。

(Holstrom, 2025, p. 1)

建議角色

- 目前 LLM 較適合做輔助與教學工具。
- 導入時仍須注意環境傷害、去技能化、智識竊用與偏見問題。

(Holstrom, 2025, p. 1)

知識組織系統的複雜性／主題標目建構

知識組織系統的複雜性

- KOS 用於表徵、彙整／聚合與支援資訊檢索。
- 每套 KOS 都包含複雜規則、詞彙、語法與記號系統。

(Holstrom, 2025, p. 1)

主題標目建構

- LCSH 的建構需要判斷資源「關於什麼」(aboutness)。
- 還必須將主題轉換為權威標目，並組合出精確的合成標目。

(Holstrom, 2025, p. 1)

編目員專業／文獻回顧

編目員專業

- 編目員需要長期學習與實務經驗。
- 人類專業仍是處理複雜 KOS 的黃金標準。

(Holstrom, 2025, p. 1)

文獻回顧

- 本研究置於 LLM 與生成式 AI 應用於編目的新興討論之中。

(Holstrom, 2025, p. 2)

自動化後設資料背景／LLM 的前提

自動化後設資料背景

- 為節省成本，圖書館長期探索自動與半自動後設資料指派。
- 過去技術包括字串比對、自然語言處理與監督式機器學習。

(Holstrom, 2025, p. 2)

LLM 的前提

- LLM 可產生符合程式語言複雜語法的程式碼。
- 作者據此推測其可能學會索引語言與控制詞彙的語法模式。

(Holstrom, 2025, p. 2)

研究問題／早期 MARC 研究

研究問題

- 核心問題是商用 LLM 是否能學會複雜 KOS 規則。
- 研究也希望分類並解釋 LLM 成功與失敗的原因。

(Holstrom, 2025, p. 2)

早期 MARC 研究

- 先前研究測試 ChatGPT 產生 MARC 記錄。
- 作者指出其記錄品質仍低於人類編目員，且有欄位遺漏與權威控制問題。

(Holstrom, 2025, p. 2)

研究設計／既有研究的提醒

研究設計

- 研究以四類資源測試 ChatGPT 與 Copilot 的主題編目效能。

(Holstrom, 2025, p. 3)

既有研究的提醒

- 既有研究強調生成式 AI 不是編目問題的簡單解方。
- 主題標目指派仍是更深層、更困難的挑戰。

(Holstrom, 2025, p. 3)

相關 LCSH 測試／最接近的比較研究

相關 LCSH 測試

- Chow 等測試電子學位論文的 LCSH 指派。
- ChatGPT 的 MARC 格式較佳，但有效 LCSH 比例偏低。

(Holstrom, 2025, p. 3)

最接近的比較研究

- Dobreski 與 Hastings 進行三種 LLM 的主題編目測試。
- 其發現與本研究相呼應：分類號常過於寬泛或錯誤。

(Holstrom, 2025, p. 3)

四類資源／工具與取用

四類資源

- 四類資源可區分 LLM 可能依賴的資料來源。
- 研究試圖分辨模型是理解內容，還是利用既有後設資料。

(Holstrom, 2025, p. 3)

工具與取用

- 研究使用免費公開版，不做額外訓練。
- 測試工具為 ChatGPT 與 Copilot。

(Holstrom, 2025, p. 4)

提示任務

提示任務

- 每筆資源都要求模型產生主題標目、分類號、Cutter 號與 MARC 欄位。
- 涵蓋 LCSH、Sears、LCC 與 DDC等四種的40筆測試對象(4 x 40 x 2 = 320)。

(Holstrom, 2025, p. 4)

來源：Holstrom (2025, 第 4 頁)

結果與討論／提示範例

結果與討論

- 本節彙整分析 320 組提示與回應。

(Holstrom, 2025, p. 4)

提示範例

- 提示將模型設定為「有經驗的圖書館編目員」。
- 並要求輸出主題標目、分類號、Cutter 號與 MARC。

(Holstrom, 2025, p. 4)

回應語料／表面正確性

回應語料

- 研究總共取得 320 份回應。
- 評估重點包括各種 KOS 與 MARC 語法、準確性、明確性與詳盡性、倫理議題及與資深編目人員間的差距。

(Holstrom, 2025, p. 4)

表面正確性

- Copilot 的回應乍看之下似乎相關且完整。
- 作者指出這種表面正確性也是 LLM 常見特徵。

(Holstrom, 2025, p. 4)

未授權標目／LCC 範例

未授權標目

- 進一步檢查後發現多個標目為非權威標目或不準確。
- Copilot 對前組合標目的複雜規則掌握不足。

(Holstrom, 2025, p. 4)

LCC 範例

- Copilot 在 LCC 與 MARC 050 範例中表現較成功。
- 其索書號與 LC 記錄相符，且 MARC 欄位與分欄正確。

(Holstrom, 2025, p. 5)

複製既有編目成果／LCSH 困難

複製既有編目成果

- 正確結果可能來自直接取用人類既有編目成果。
- 這引發知識竊用與依賴人類資料的問題。

(Holstrom, 2025, p. 6)

LCSH 困難

- 兩種模型在四類資源的 LCSH 指派皆有困難。
- 錯誤涵蓋不準確、非權威標目、過寬或過窄、複分錯用與缺乏詳盡性。

(Holstrom, 2025, p. 6)

Copilot LCSH 表格／ChatGPT LCSH 表格

Copilot LCSH 表格

- 表 1 彙整 Copilot 在四類資源中的主題編目錯誤。
- 非權威標目是最顯著的問題，尤其在虛構資源與全文文章中。

(Holstrom, 2025, p. 7)

ChatGPT LCSH 表格

- 表 2 彙整 ChatGPT 的主題編目錯誤。
- ChatGPT 的非權威標目數量低於 Copilot，但仍不可忽視。

(Holstrom, 2025, p. 7)

Table 1. Summary of the types of subject cataloging errors produced by Copilot while cataloging well-known resources, obscure resources, fabricated resources, and full-text articles.

	Well-Known Resources	Obscure Resources	Fabricated Resources	Full-Text Articles
Total subject headings	92	94	83	91
Inaccurate	6	8	-	3
Non-authorized	27	33	41	46
Overly broad	6	7	-	3
Overly specific	3	6	-	8
Overuse or mis-use of subdivisions	18	27	32	26
Redundancy	4	3	4	5
Lack of exhaustivity	2	1	-	4

(Holstrom, 2025, p. 7)

Table 2. Summary of the types of subject cataloging errors produced by ChatGPT while cataloging well-known resources, obscure resources, fabricated resources, and full-text articles.

	Well-Known Resources	Obscure Resources	Fabricated Resources	Full-Text Articles
Total subject headings	87	91	72	90
Inaccurate	3	5	-	4
Non-authorized	22	23	19	24
Overly broad	7	2	-	5
Overly specific	2	1	-	4
Overuse or mis-use of subdivisions	14	19	12	10
Redundancy	3	5	2	2
Lack of exhaustivity	3	3	-	6

(Holstrom, 2025, p. 7)

錯誤率／知名圖書

錯誤率

- 知名與冷門資源的回應沒有完全符合 LC 黃金標準。
- 非權威詞彙比例高，代表模型輸出仍不可靠。

(Holstrom, 2025, p. 8)

知名圖書

- 在知名書籍上，兩種模型表現相對較好。
- 但這可能反映它們能存取或再現既有 LC 記錄。

(Holstrom, 2025, p. 8)

虛構資源／詳盡性的正面發現

虛構資源

- 研究以虛構書名測試模型在沒有既有記錄時的能力。
- 兩種模型會辨識資源為假想或虛構，仍提供建議標目。

(Holstrom, 2025, p. 8)

詳盡性的正面發現

- LLM 有時能為冷門資源提出較完整的概念線索。
- 即使標目屬非權威標目，仍可能啟發人類編目員補充主題面向。

(Holstrom, 2025, p. 9)

Sears、LCC、DDC 與 MARC/LCSH 總結

Sears、LCC、DDC 與 MARC

- 相較於主題標目，模型在分類號與部分 MARC 輸出上較成功。

(Holstrom, 2025, p. 10)

LCSH 總結

- 整體而言，兩種模型的 LCSH 指派表現不佳。
- Copilot 的「成功」尤其可能來自複製既有記錄。

(Holstrom, 2025, p. 9)

Sears List 行為／授權標目的細節

Sears List 行為

- Sears List 較小且規則不同。
- 兩種模型常把 Sears 與 LCSH 提示視為同類問題，甚至回覆相同標目。

(Holstrom, 2025, p. 9)

授權標目的細節

- Sears List 對「權威標目」的概念較具在地化彈性。
- 模型沒有辨識或處理這一差異。

(Holstrom, 2025, p. 9)

分類號較成功／虛構分類範例

分類號較成功

- LCC 與 DDC 分類號的成功率高於主題標目。
- 成功可能只是因為模型能檢索到現成答案。

(Holstrom, 2025, p. 10)

虛構分類範例

- 對虛構書名，兩模型可能產生一致的分類。
- 但MARC 欄位仍可能錯用，顯示格式控制不足。

(Holstrom, 2025, p. 10)

MARC 不一致／分類號比對

MARC 不一致

- 相較於分類號本身，MARC 欄位格式不一致性更常見。
- 每類資源皆產生 20 個分類結果。

(Holstrom, 2025, p. 10)

分類號比對

- 可比對案例中，90% 的 LCC/DDC 分類與 LC 相符。
- 不相符時，模型有時給出合理的替代分類。

(Holstrom, 2025, p. 10)

編目員注意事項／LLM 執行編目工作的意涵

編目員注意事項

- 即使分類號看似成功，也不能直接信任。
- 編目員需警覺：模型可能只是複製人類工作，且介面呈現容易造成信任錯覺。

(Holstrom, 2025, p. 11)

LLM 執行編目工作的意涵

- 兩種模型有潛力，但不是專為編目設計，且缺點眾多。

(Holstrom, 2025, p. 12)

Copilot 分類與 MARC 錯誤

Copilot 分類與 MARC 錯誤

- 表 3 顯示 Copilot 在 MARC 語法與 Cutter 號上仍有錯誤。
- 全文文章與虛構資源中的 MARC 語法錯誤尤其明顯。

(Holstrom, 2025, p. 11)

來源：Holstrom (2025, 第 11 頁)

Table 3. Summary of the types of cataloging errors produced by Copilot while assigning LCC and DDC class number and corresponding MARC fields for well-known resources, obscure resources, fabricated resources, and full-text articles.

	Well-Known Resources	Obscure Resources	Fabricated Resources	Full-Text Articles
Total classes	20	20	20	20
Matches Library of Congress catalog	19	18	-	-
Inaccurate	0	0	-	3
Incorrect mapping of class name to notation	0	0	0	0
MARC syntax incorrect	4	3	5	5
Incorrect application of the Cutter Table*	1	3	0	1

* Note that Cutter numbers do not always follow the Cutter Table perfectly, so this study considered reasonable matches to be correct applications of the Cutter Table.

(Holstrom, 2025, p. 11)

ChatGPT 分類與 MARC 錯誤

ChatGPT 分類與 MARC 錯誤

- 表 4 顯示 ChatGPT 的 MARC 錯誤較少。
- 但 Cutter 號應用在各類資源中仍不穩定。

(Holstrom, 2025, p. 12)

來源：Holstrom (2025, 第 12 頁)

Table 4. Summary of the types of cataloging errors produced by ChatGPT while assigning LCC and DDC class number and corresponding MARC fields for well-known resources, obscure resources, fabricated resources, and full-text articles.

	Well-Known Resources	Obscure Resources	Fabricated Resources	Full-Text Articles
Total classes	20	20	20	20
Matches Library of Congress catalog	20	15	-	-
Inaccurate	0	0	-	1
Incorrect mapping of class name to notation	0	0	0	0
MARC syntax incorrect	2	0	2	3
Incorrect application of the Cutter Table*	2	2	4	2

(Holstrom, 2025, p. 12)

MARC 與 Cutter 細節

MARC 與 Cutter 細節

- 兩種模型能將分類名稱與分類號對應。
- MARC 050 與 082 的語法錯誤率仍高。

(Holstrom, 2025, p. 11)

來源：Holstrom (2025, 第 11 頁)

審核負擔

審核負擔

- 即使只有一個錯誤，也會破壞編目記錄的可信度。
- 若必須全面審核與修正，LLM 省時價值就大幅降低。

(Holstrom, 2025, p. 12)

來源：Holstrom (2025, 第 12 頁)

輔助與教學工具／結論與未來研究

輔助與教學工具

- 目前 LLM 適合協助人類，而非自動取代人類。
- 它們可提供可能被忽略的標目或分類線索，供編目員批判性使用。

(Holstrom, 2025, p. 13)

結論與未來研究

- ChatGPT 與 Copilot 尚不足以取代人類編目員，但可作為支援 KOS 與主題編目的輔助工具。

(Holstrom, 2025, p. 13)

AI 聊天機器人與主題編目／執行摘要

AI 聊天機器人與主題編目

- Dobreski, B. & Hastings, C. (2025). AI chatbots and subject cataloging: A performance test. Library Resources & Technical Services, 69(2).
<https://doi.org/10.5860/lrts.69n2.8440>

執行摘要

- 本研究檢驗免費 AI 聊天機器人能否指派主題標目與分類號。
- 在 DDC、LCC 與 LCSH 任務中，整體表現普遍不佳。
- 研究結果提醒：主題編目仍需人類專業判斷與監督。

(Dobreski & Hastings, 2025, p. 1)

核心研究問題

核心研究問題

- 圖書館日益關注 AI 工具在編目流程中的應用。
- 核心問題是：易用的聊天機器人能否降低主題編目所需的時間與專業門檻。
- 本文重視實證測試，而非只討論 AI 的一般潛力。

(Dobreski & Hastings, 2025, pp. 1 – 2)

主題編目為何困難

主題編目為何困難

- 主題編目需要分析資源的「關於什麼」(aboutness)。
- 編目員必須將主題分析對應到 DDC、LCC 與 LCSH 等形式系統。
- 此工作耗時，因為這些系統需要專門訓練與實務經驗。

(Dobreski & Hastings, 2025, p. 2)

測試工具與任務

測試工具與任務

- 研究比較 ChatGPT 3.5、Gemini 1.0 與 Copilot 2024.5 built。
- 每個聊天機器人都被要求指派 DDC、LCC 與 LCSH。
- 研究旨在記錄特定時間點，免費工具的實際能力。

(Dobreski & Hastings, 2025, p. 2)

文章結構／圖書館中的 AI 興趣

文章結構

- 文獻回顧說明 AI 在圖書館與編目中的位置。
- 研究方法以編目教材為基礎，設計可重複的能力測試。
- 結果與討論分析錯誤數量、錯誤類型及其實務意涵。

(Dobreski & Hastings, 2025, pp. 2 – 3)

圖書館中的 AI 興趣

- 圖書館正探索 AI 在參考服務、探索介面、館藏管理與編目中的用途。本文將焦點到主題編目的實際表現。

(Dobreski & Hastings, 2025, pp. 1 – 2)

廣泛的 AI 實驗

廣泛的 AI 實驗

- 近期圖書館 AI 案例包括參考服務聊天機器人、圖書推薦搜尋、教學用 AI 人物與檔案轉譯。
- 案例多樣性顯示專業社群對 AI 服務有高度興趣。
- 然而，實際導入研究少，仍著重於概念性討論。

(Dobreski & Hastings, 2025, p. 2)

編目領域興趣上升

編目領域興趣上升

- 編目員已開始探索開源與廠商支援的 AI 工具。
- 通用 LLM 聊天機器人具可取用性，且擅長處理文字，因此受到關注。
- 本文將這種可取用性視為有必要測試的實務原因。

(Dobreski & Hastings, 2025, pp. 1 – 3)

編目員使用情形

編目員使用情形

- 一項針對 26 所大學的調查發現，編目流程中已有 AI 實驗。
- 受訪者提到使用 Bard/Gemini、Bing/Copilot 與 ChatGPT。
- 不少受訪者仍認為這些工具並未提高生產力。

(Dobreski & Hastings, 2025, p. 3)

品質與可靠性疑慮

品質與可靠性疑慮

- AI 工具可能支援後設資料工作，但輸出品質仍是主要疑慮。
- 過去編目測試發現，AI 產生的記錄看似合理，卻含有虛構或錯誤資訊。
- 當輸出看起來像權威後設資料時，此風險更為險要。

(Dobreski & Hastings, 2025, pp. 1 – 3)

先前 MARC 與 BIBFRAME 測試

先前 MARC 與 BIBFRAME 測試

- Breeding 測試 ChatGPT 產生 MARC 與 BIBFRAME 記錄，發現嚴重不正確。
- Brzustowicz 的結果較正面，但仍建議由專業編目員審核。
- 既有文獻共同指向編目員監督的必要性。

(Dobreski & Hastings, 2025, p. 3)

先前主題編目測試

先前主題編目測試

- 奧克拉荷馬州立大學的小型測試發現，ChatGPT 在 DDC 較弱，但在 LCSH 較可靠。
- 另一研究以學位論文題名與摘要提示 ChatGPT 產生主題標題。
- 這些研究顯示潛力，但也呈現有效性問題。

(Dobreski & Hastings, 2025, p. 3)

研究缺口

研究缺口

- 先前研究尚未充分比較多個聊天機器人在多種主題編目任務上的表現。
- 作者希望建立可記錄、可重複的方法，以便未來版本再測。
- 研究缺口在於實證表現證據，而非一般性的 AI 倡議。

(Dobreski & Hastings, 2025, pp. 3 – 4)

文章貢獻

文章貢獻

- 本研究為三個常用免費聊天機器人提供效能基準。
- 研究記錄 98 個提示與 294 個聊天機器人回應。
- 其價值不只在分數，也在錯誤類型的質性分析。

(Dobreski & Hastings, 2025, p. 2)

可重複的效能測試／測試來源

可重複的效能測試

- 研究以主題編目教材練習題、簡單提示與明確評分規則，檢驗聊天機器人的表現。

(Dobreski & Hastings, 2025, pp. 3 – 5)

測試來源

- 測試取自 Broughton 《Essential Classification》 第二版。
- 該書用於初學者主題編目教學。
- 其練習以描述性圖書資訊為基礎，適合用來測試基本指派能力。

(Dobreski & Hastings, 2025, pp. 3 – 4)

選擇的系統

選擇的系統

- 研究聚焦 DDC、LCC 與 LCSH，因為它們在圖書館編目中具有重要地位。
- 題目選自涵蓋這些分類系統的章節。
- 研究避免姓名 Cutter 號或特殊表格等過窄任務。

(Dobreski & Hastings, 2025, p. 4)

聊天機器人與版本

聊天機器人與版本

- 測試工具為 ChatGPT、Gemini 與 Copilot。
- 作者使用 2024 年 5 月可用的免費版本：ChatGPT 3.5、Gemini 1.0 與 Copilot build 2024.5。
- 此選擇反映不受預算限制即可取得的工具。

(Dobreski & Hastings, 2025, p. 4)

提示設計／題目組成

提示設計

- 每一題練習都被改寫為單一提示詞。
- 作者只做少量修改，並避免複雜提示工程。
- 這些提示詞刻意保持簡單，便於重複測試。

(Dobreski & Hastings, 2025, p. 4)

題目組成

- 相同的 98 個問題分別詢問三個工具。
- 題組包含 23 題 DDC、25 題 LCC 與 50 題 LCSH。
- 三個工具合計產生並保存 294 份回應。

(Dobreski & Hastings, 2025, p. 4)

評分類別

評分類別

- 回應與教材答案鍵進行比較。
- 依任務不同，答案被標為正確、可接受、接近、錯誤或拒答。
- 在 LCSH 任務中，若回應含多個標題，研究者選取最接近者評分。

(Dobreski & Hastings, 2025, pp. 4 – 5)

驗證來源

驗證來源

- 即使是錯誤答案，研究者也檢查其是否為有效號碼或標題。
- 驗證來源包括 WebDewey、Classification Web、編目手冊與 OCLC WorldShare Record Manager。
- 這使研究能區分「存在但不適當」與「幻覺式不存在」的答案。

(Dobreski & Hastings, 2025, p. 5)

研究者專業／杜威十進分類法

研究者專業

- 研究者具專業編目經驗，且有十年以上編目與分類教學經驗。
- 此專業很重要，因為部分答案需判斷是否接近或可接受。
- 方法結合答案鍵比對與專業驗證。

(Dobreski & Hastings, 2025, p. 5)

杜威十進分類法

- DDC 結果顯示三個聊天機器人分數都偏低，常見錯誤為號碼過廣或主題錯置。

(Dobreski & Hastings, 2025, pp. 5 – 6)

DDC 整體表現

DDC 整體表現

- 三個工具在 23 題 DDC 問題上表現皆不佳。
- 最終成績為 ChatGPT 26%、Copilot 9%、Gemini 17%。
- 即使表現最佳的工具，也未接近可接受的可靠性。

(Dobreski & Hastings, 2025, p. 5)

	ChatGPT	Copilot	Gemini
Correct	5	1	4
Acceptable alternative	1	1	0
Incorrect	17	21	19
Refusal	0	0	0
Final grade	26%	9%	17%
No. of valid DDCs	16	14	16
Percentage valid	70%	61%	70%

(Dobreski & Hastings, 2025, p. 5)

有效性與正確性

有效性與正確性

- 許多回傳的 DDC 號碼確實存在，但並不適合題名。
- ChatGPT 與 Gemini 的有效 DDC 比率皆為 70%，Copilot 為 61%。
- 號碼存在並不等於主題判斷正確。

(Dobreski & Hastings, 2025, p. 5)

主要 DDC 錯誤類型

主要 DDC 錯誤類型

- DDC 最常見錯誤是指派過於寬泛的號碼。
- 另一常見錯誤是為指派號碼給錯誤主題。
- 部分回應也使用不存在的號碼，或不能使用DDC Table的類號，或未遵循號碼順序規則。

(Dobreski & Hastings, 2025, pp. 5 – 6)

	ChatGPT (<i>n</i> = 17)	Copilot (<i>n</i> = 21)	Gemini (<i>n</i> = 19)
Incorrect topic	5	6	8
DDC number does not exist	2	2	0
DDC number too general	7	12	8
DDC number too specific	1	0	1
Did not follow order rules	2	1	2

(Dobreski & Hastings, 2025, p. 6)

DDC 錯配例子

DDC 錯配例子

- 關於大教堂的書被指派一般建築號，而非預期的大教堂號。
- 死刑書目的題名被指派成人身法律犯罪的號碼。
- 這些錯誤顯示問題在主題分析，而不只是格式不熟。

(Dobreski & Hastings, 2025, pp. 5 – 6)

DDC 小結

DDC 小結

- 免費聊天機器人未展現可靠的 DDC 號碼指派能力。
- 編目員仍須驗證號碼是否存在、意義是否正確、是否夠具體，以及是否符合規則。
- 這種驗證負擔限制了節省時間的效益。

(Dobreski & Hastings, 2025, pp. 5 – 6)

美國國會圖書館分類法／LCC 整體表現

美國國會圖書館分類法

- LCC 表現同樣不佳。Gemini 特別吃力，而不存在的 LCC 號碼成為重要錯誤類型。

(Dobreski & Hastings, 2025, pp. 6 – 7)

LCC 整體表現

- LCC 測試包含 25 題。
- 最終成績為 ChatGPT 20%、Copilot 24%、Gemini 4%。
- Gemini 僅提供一個正確答案，且有兩題拒答。

(Dobreski & Hastings, 2025, p. 6)

	ChatGPT	Copilot	Gemini
Correct	2	5	1
Close	2	0	0
Acceptable alternative	1	1	0
Incorrect	20	19	22
Refusal	0	0	2
Final grade	20%	24%	4%
Number of valid LCCs	13	13	3
Percentage valid	52%	52%	13%

LCC

(Dobreski & Hastings, 2025, p. 6)

有效 LCC 號碼稀少

有效 LCC 號碼稀少

- ChatGPT 與 Copilot 各有 52% 的回應提供有效 LCC 號碼。
- Gemini 的有效 LCC 比率僅 13%。
- Gemini 更常產生不存在的 LCC 號碼。

(Dobreski & Hastings, 2025, p. 6)

主要 LCC 錯誤類型

主要 LCC 錯誤類型

- LCC 最常見錯誤是為指派號碼給錯誤主題。
- 其他錯誤包括過於一般的號碼、只給主類號，以及提供號碼範圍。
- 這些錯誤顯示工具難以掌握 LCC 結構與主題精確性。

(Dobreski & Hastings, 2025, pp. 6 – 7)

	ChatGPT (<i>n</i> = 20)	Copilot (<i>n</i> = 19)	Gemini (<i>n</i> = 22)
Incorrect topic	6	12	11
LCC number does not exist	2	0	8
LCC number too general	6	5	1
LCC number too specific	2	0	0
Provided main class only	3	2	2
Provided number range only	1	0	0

(Dobreski & Hastings, 2025, p. 6)

LCC 錯配例子／LCC 小結

LCC 錯配例子

- 針對古羅馬職業的書，Gemini 建議了埃及史的分類號。（張冠李戴）
- 部分工具只回傳如「BV」這樣的主類號。
- ChatGPT 曾回傳一段號碼範圍，而非單一分類號。

(Dobreski & Hastings, 2025, pp. 6 – 7)

LCC 小結

- 這些工具並不適合可靠地指派 LCC。
- 由於 LCC 含有空缺與未指派範圍，幻覺號碼較易被發現，但仍具問題。
- 評估結果仍需人類 LCC 專業知識判別。

(Dobreski & Hastings, 2025, pp. 6 – 7)

美國國會圖書館主題標題表／LCSH 整體表現

美國國會圖書館主題標題表

- 主題標題指派相對較有潛力，尤其是 ChatGPT；但整體仍未達可依賴的編目品質。

(Dobreski & Hastings, 2025, pp. 7 – 8)

LCSH 整體表現

- LCSH 測試包含 50 題。
- 最終成績為 ChatGPT 54%、Copilot 10%、Gemini 26%。
- ChatGPT 是所有工具與任務組合中分數最高者，但仍未達及格水準。

(Dobreski & Hastings, 2025, p. 7)

	ChatGPT	Copilot	Gemini
Correct	21	4	11
Close	5	1	1
Acceptable alternative	1	0	1
Incorrect	23	45	33
Refusal	0	0	4
Final grade	54%	10%	26%
Number of valid LCSHs	38	18	24
Percentage valid	76%	36%	48%

(Dobreski & Hastings, 2025, p. 7)

有效標題與相對潛力

有效標題與相對潛力

- ChatGPT 在 76% 的題目中至少提供一個有效 LCSH。
- Copilot 與 Gemini 在同一指標上都低於 50%。
- 這支持一種看法：聊天機器人可能較適用於主題標目，而非分類號。

(Dobreski & Hastings, 2025, p. 7)

主要 LCSH 錯誤類型

主要 LCSH 錯誤類型

- Copilot 與 Gemini 常產生不存在的 LCSH 標題。
- ChatGPT 最常遺漏必要的複分。
- 其他錯誤包括標題過廣、過窄，或地理複分形式錯誤。

(Dobreski & Hastings, 2025, pp. 7 – 8)

	ChatGPT (<i>n</i> = 23)	Copilot (<i>n</i> = 45)	Gemini (<i>n</i> = 33)
Incorrect topic	1	1	2
LCSH does not exist	5	26	19
LCSH too broad	2	2	3
LCSH too narrow	3	1	8
LCSH lacking subdivision	11	15	0
Incorrect geographic subdivision	1	0	1

(Dobreski & Hastings, 2025, p. 8)

每次回應多個標題

每次回應多個標題

- ChatGPT 在 50 個 LCSH 提示中共回傳 298 個標題，平均每書 6 個。
- 所有 ChatGPT 提供的標題中，只有 63% 為有效標題。
- 更多建議可能有助於使用者，但也提高篩選需求。

(Dobreski & Hastings, 2025, p. 8)

LCSH 小結

LCSH 小結

- LCSH 指派是本研究中最有潛力的任務，尤其對 ChatGPT 而言。
- 然而，沒有任何工具能產生可直接用於編目的可靠標題。
- 權威控制與編目員審查仍不可或缺。

(Dobreski & Hastings, 2025, pp. 7 – 9)

實務意涵／分類任務仍不適用

實務意涵

- 討論將結果轉化為對編目實務的提醒：AI 輸出可能看似合理，但驗證仍高度耗費人力。

(Dobreski & Hastings, 2025, pp. 9 – 11)

分類任務仍不適用

- 三個工具在 DDC 與 LCC 測試中皆表現不足。
- 錯誤常涉及過廣號碼，或完全不同的主題號碼。
- 部分解釋看似合理，因此更需要仔細檢查。

(Dobreski & Hastings, 2025, p. 9)

主題標題仍不足

主題標題仍不足

- ChatGPT 在 LCSH 上比在分類號上更有潛力。
- 即便如此，54% 的成績表示它不能在無編目員監督下被信任。
- 本研究並不支持目前由聊天機器人獨立進行編目。

(Dobreski & Hastings, 2025, pp. 9 – 10)

編目員驗證清單

編目員驗證清單

- 編目員應檢查建議的號碼或標題是否存在。
- 他們也應確認其意義是否與聊天機器人聲稱一致。
- 此外，還需判斷是否有更精確號碼或更完整複分標題。

(Dobreski & Hastings, 2025, p. 10)

提示工程與流程成本

提示工程與流程成本

- 本研究使用簡單提示詞，因此更強的提示工程可能改善結果。
- 然而，反覆提示與驗證可能抵消預期的效率收益。
- 資訊專業人員的 AI 素養必須包括評估聊天機器人編目結果的能力。

(Dobreski & Hastings, 2025, pp. 10 – 11)

未來研究方向／主要結論

未來研究方向

- 此測試可用於未來版本的同一批工具。
- 後續研究可比較免費與付費工具，或測試初學編目員搭配聊天機器人的表現。
- FAST 或 MeSH 等其他詞彙系統也可能產生不同結果。

(Dobreski & Hastings, 2025, p. 11)

主要結論

- 三個免費聊天機器人都不足以勝任 DDC、LCC 或 LCSH 指派。
- 觀察到的錯誤類型指出編目員實驗 AI 時應注意之處。
- 主題編目仍需人力、系統熟悉度與專業判斷。

(Dobreski & Hastings, 2025, pp. 11 – 12)

負責任使用訊息／編目、詮釋資料與生成式人工智慧

負責任使用訊息

- AI 工具未來可能協助編目員，尤其是在主題標題指派上。
- 負責任導入需要持續測試，而非預設工具可靠。
- 本文的基準測試提供了追蹤進展的方法。

(Dobreski & Hastings, 2025, pp. 11 – 12)

執行摘要

編目、後設資料與生成式人工智慧

Roncaglia, G. (2026).
Cataloguing, metadata, and
generative AI: Early experiences
and future perspectives. JLIS. It
17(1), 106-27.
<https://doi.org/10.36253/jlis.it-693>.

- 生成式 AI 能協助編目，但後設資料標準要求嚴謹性。
- 文章以「形式化架構」與「機率式生成」的張力作為核心。
- 有用的 AI 必須與 LOD、LRM 及負責任治理相互銜接。

(Roncaglia, 2026, p. 106)

核心研究問題／作者如何使用 AI

核心研究問題

- LLM 能否支援書目描述，而不破壞形式化控制？
- 問題不只是自動化，而是可信的語意表徵。
- 答案取決於資料模型、驗證程序與機構目標。

(Roncaglia, 2026, pp. 106-107)

作者如何使用 AI

- 作者將多個 LLM 作為研究寫作的「陪練」。
- AI 協助主題建議、文獻探索與薄弱處檢查。
- 最終論證、責任與作者性 (authorship) 仍明確屬於人類作者。

(Roncaglia, 2026, pp. 107-108)

為何詮釋資料對 AI 很困難／三種策略

為何後設資料對 AI 很困難

- 書目後設資料受到規則、粒度與語意限制。
- LLM 流暢但具機率性，可能產生貌似合理的虛構書目。
- 核心挑戰是把生成結果根植於權威結構。

(Roncaglia, 2026, pp. 106, 113)

三種策略

- 專門化 AI：將模型限制於受控且標準化的資料中。
- RAG：結合 LLM 彈性與權威知識庫。
- 下一代 LLM：讓神諭學會更像建築師般運作。

(Roncaglia, 2026, pp. 113-114)

專門化 AI／自動索引根源

專門化 AI

- 以受控資料訓練的系統，通常比當代生成式 AI 更接近傳統機器學習。

(Roncaglia, 2026, p. 114)

自動索引根源

- 早期自動索引延伸索引彙編與統計文本分析。
- 《托馬斯索引》是計算/數位人文領域的重要先例。
- 初期樂觀程度超過早期演算法的實際能力。

(Roncaglia, 2026, p. 115)

圖書館專家系統

圖書館專家系統

- 1980 至 1990 年代出現的是**聚焦式專家系統**，而非通用 AI。
- IRIS(Intelligent Reference Information System)專案協助使用者選擇索引並存取網路資源。
- 這些系統以**規則為基礎**，且**限於特定領域**。

(Roncaglia, 2026, p. 115)

資料庫選擇系統／機器學習時期

資料庫選擇系統

- 伊利諾大學開發專家系統，以模擬參考館員選擇資料庫。
- 該系統使用控制詞彙與資料庫特徵。
- 此例顯示結構化知識能引導使用者探索。

(Roncaglia, 2026, pp. 115-116)

機器學習時期

- 2009 至 2018 年間，圖書館 AI 文獻明顯成長。
- 中國在發表量與引用趨勢中逐漸突出。
- Transformer 之前的 ML 已出現相關問題，後來由 LLM 的問題更大。

(Roncaglia, 2026, p. 116)

OCR 與手寫辨識／AI 素養與倫理

OCR 與手寫辨識

- 神經網路對 OCR、手寫辨識與困難文本相當有用。
- Transkribus 等工具展現輔助轉錄工作流程。
- 這類應用較適合受限辨識任務，而非開放式生成。

(Roncaglia, 2026, pp. 116-117)

AI 素養與倫理

- IFLA 聲明預見圖書館在AI 素養的重要性。
- 它強調個人化、偏見、倫理與批判意識。
- 部分分析因早於 LLM 爆發而已顯陳舊。

(Roncaglia, 2026, p. 117)

案例：Annif、SAGE、OCLC、BL

案例：Annif、SAGE、OCLC、BL

- Annif 以開源 ML 工具支援自動主題標引。
- SAGE 使用 NLP 與 ML 豐富 Europeana 的 RDF 資料。
- OCLC 與大英圖書館將 ML 用於複本偵測與語言識別（Languid:Language Identification Project）。

(Roncaglia, 2026, pp. 117-118)

美國國會圖書館 ECD／專門化 AI 的限制

美國國會圖書館 ECD

- 探索計算描述專案測試 ML 於大規模書目記錄。
- 模型從數位書籍中建議 LCSH 主題標目與 NAF 作者名稱。
- 人工審核仍是品質與問責的必要條件。

(Roncaglia, 2026, p. 118)

專門化 AI 的限制

- 受控語料可降低，但無法消除幻覺或錯誤分類。
- 人類參與審核可能昂貴且使工作異化。
- 封閉式專門化犧牲自然語言 LLM 的彈性。

(Roncaglia, 2026, p. 118)

檢索增強生成／RAG 工作流程

檢索增強生成

- 讓建築師的受控資料與神諭的自然語言彈性協作。

(Roncaglia, 2026, p. 119)

RAG 工作流程

- 檢索階段：從可信外部知識庫找出相關證據。
- 生成階段：結合查詢與證據，以流暢語言回答。
- 模型因此被植基，不只依賴內部訓練語料。

(Roncaglia, 2026, p. 119)

RAG 用於參考服務

RAG 用於參考服務

- RAG 很適合參考聊天機器人與自然語言目錄存取。
- 例子包括 Alphabetica 的 Alphy 與歐洲議會檔案的 Archibot。
- 它可個人化語言、語域、摘要與多模態回應。

(Roncaglia, 2026, p. 119)

RAG 用於後設資料

RAG 用於後設資料

- 早期僅用 LLM 的後設資料實驗很快遇到**幻覺問題**。
- RAG 可結合 OCR、LLM 與控制索引典產生後設資料。
- OCUL/Scholars Portal 政府文件專案展現此方向。

(Roncaglia, 2026, p. 119)

商業系統與不透明性／下一代 LLM

商業系統與不透明性

- AI 功能已出現在探索、參考與後設資料編輯平台中。
- 商業供應商常未充分揭露模型、資料或架構。
- 圖書館採用 AI 中介工具前需要透明資訊。

(Roncaglia, 2026, p. 120)

下一代 LLM

- 新興能力可能讓生成模型更能遵守形式化後設資料限制。

(Roncaglia, 2026, p. 120)

擴展脈絡視窗／多模態與世界模型

擴展脈絡視窗

- 更大的脈絡視窗讓模型一次考量更多文件。
- 附加標準與語料可作為局部事實的依據材料。
- 限制仍在：大型館藏、動態更新與殘餘幻覺。

(Roncaglia, 2026, p. 120)

多模態與世界模型

- 多模態系統整合文字、影像、聲音、影片與其他訊號。
- 文章追問：標準化資料能否也成為一種「模態」。
- 這類模型或能更好區分自由表達與受限後設資料。

(Roncaglia, 2026, p. 120)

推理模型／深度研究與代理能力

推理模型

- 推理模型會在最終答案前產生中介推理過程。
- 它們可被訓練朝向精確、結構與驗證。
- 文章想像以「文獻推理」的後設資料任務。

(Roncaglia, 2026, p. 121)

深度研究與代理能力

- 深度研究會比較多個來源並評估可靠性。
- 代理式系統可在目錄與資料庫間執行一連串動作。
- 這可能在不建置專屬 RAG 架構下接近 RAG 效果。

(Roncaglia, 2026, p. 121)

本地小／中型模型／倫理與制度議程

本地小／中型模型

- 小型與中型語言模型可透過微調在本地運行。
- 本地使用可保護讀者敏感資料與受著作權保護內容。
- 它們依賴萃取或篩選，未真正消除大型模型成本。

(Roncaglia, 2026, pp. 121-122)

倫理與制度議程

- 圖書館必須處理偏見、透明度、AI 素養與問責。
- AI 採用應依公共利益導向的圖書館價值來評估。
- 重視隱私的應用使本地模型格外具吸引力。

(Roncaglia, 2026, pp. 117, 122)

最後重點／生成式 AI 強化編目

最後重點

- 不要把前沿 AI 勉強套入概念上過時的 MARC 中心模型。
- 有前景的路徑是讓 AI 對齊 (align) LOD、RDF、LRM 與語意粒度。
- 未來任務不是取代編目員，而是重新設計可信後設資料流程。

(Roncaglia, 2026, p. 122)

Wade-Giles 轉換為拼音

生成式 AI 強化編目：Wade-Giles 轉換為拼音

- Sun, L. (2025). Enhancing cataloging with generative AI: Converting Wade-Giles to Pinyin. *Cataloging & Classification Quarterly*, 63(4), 267 – 283.
<https://doi.org/10.1080/01639374.2025.2508956>

生成式 AI 工具／三種工具評估

生成式 AI 工具

- 生成式 AI 正透過提升效率與擴大多語資源取用來改變圖書館編目。
- 本研究聚焦於 MARC 記錄中 Wade-Giles 題名自動轉換為拼音。

(Sun, 2025, p. 267)

三種工具評估

- 研究評估 ChatGPT、Copilot 與 Gemini。
- 主要挑戰包括輸出不一致、準確率波動，以及資料前處理需求。

(Sun, 2025, p. 267)

前景與限制／圖書館編目的多面向

前景與限制

- AI 導入編目流程具有前景。
- 但仍需持續精進，才能達到可擴展的效能最佳化。

(Sun, 2025, p. 267)

圖書館編目的多面向

- 生成式 AI 已逐漸融入圖書館編目的多個層面。
- 作用在於提升書目記錄的**效率、準確性與完整性**。

(Sun, 2025, p. 267)

目錄項目與近用／多語資源

目錄項目與近用

- AI 可用於增補目錄項目。
- 依據內容分析，系統可建議相關作品、版本、譯本與連結主題。

(Sun, 2025, p. 267)

多語資源

- 生成式 AI 對目錄翻譯與多語檢索具有價值。
- 它能擴大與最佳化館藏的可及範圍。

(Sun, 2025, p. 267)

Wade-Giles 題名／編目中的 AI

Wade-Giles 題名

- 研究以主要研究型圖書館中的 MARC 題名轉換為案例。
- 面對超過 14,000 筆 Wade-Giles 題名，AI 被視為可行方案。

(Sun, 2025, p. 267)

編目中的 AI

- AI 因能協助與改善圖書館編目而受到重視。
- 相關討論聚焦於應用、工具、挑戰與未來發展。

(Sun, 2025, p. 268)

概念性討論／ChatGPT 與 MARC 記錄

概念性討論

- 既有討論多偏概念或前瞻，較少深入分析編目技術。
- 議題包括 NLP、大型訓練模型、自動化後設資料生成、主題標目生成、錯誤偵測、使用者體驗與目錄紀豐富化等。

(Sun, 2025, p. 268)

ChatGPT 與 MARC 記錄

- Brzustowicz 的研究是少數探討 ChatGPT 生成 MARC 記錄的嘗試。
- 其實驗顯示 AI 可能簡化記錄建立流程並提升效率。

(Sun, 2025, p. 268)

批評與限制／翻譯功能

批評與限制

- 該研究因方法透明度、MARC 正確判斷與實務性受到批評。
- 另一項疑慮是其僅分析單一記錄。

(Sun, 2025, p. 268)

翻譯功能

- 調查顯示，翻譯功能是編目使用 AI 的最常見理由。
- 準確性、一致性與可靠性仍是持續的關切重點。

(Sun, 2025, p. 268)

分類與書目著錄／圖書館系統供應商

分類與書目著錄

- 既有研究探討 AI 在書目著錄中的效率與一致性。
- 另有研究指出，在 1,000 筆訓練樣本下，AI 分類準確率超過 70%。

(Sun, 2025, p. 269)

圖書館系統供應商

- 圖書館供應商率先將 AI 整合進編目流程。
- OCLC、Ex Libris 與 OverDrive 等已發展自動化後設資料生成、探索工具改善與工作流程優化。

(Sun, 2025, p. 269)

Alma AI 後設資料助理／Wade-Giles 與拼音

Alma AI 後設資料助理

- Alma 在 Metadata Editor 中加入 AI Metadata Assistant。
- 該工具能分析館藏資源並建議相關後設資料，協助編目員提高效率。

(Sun, 2025, p. 269)

Wade-Giles 與拼音

- Wade-Giles 與大陸拼音都是以拉丁字母表示華語的羅馬化系統。
- 大陸拼音在準確性與清晰度上較具優勢，並成為國際通用系統。

(Sun, 2025, p. 269)

大規模轉換／轉換任務的複雜性

大規模轉換

- LC 於 2000 - 2001 年完成大規模 Wade-Giles 轉拼音。
- 轉換涵蓋超過 15,900 筆權威檔與 3,200,000 筆書目記錄。

(Sun, 2025, p. 270)

轉換任務的複雜性

- 任務複雜性涉及姓名、地理轄區通稱，以及多語混合羅馬化。
- 流程同時使用程式自動化與人工介入。

(Sun, 2025, p. 270)

未轉換記錄／未轉換簡略記錄

未轉換記錄

- 許多學術圖書館仍保有可轉換，但尚未更新的記錄。
- 因機構通常未公布數量，未轉換記錄的規模仍不明確。

(Sun, 2025, p. 270)

未轉換簡略記錄

- 作者所在圖書館曾接收 OCLC 的大陸拼音覆蓋記錄。
- 部分簡編記錄缺少 OCLC 號，仍以 Wade-Giles 形式留在資料庫。

(Sun, 2025, p. 271)

Alma 分析功能／使用者不可發現

Alma 分析功能

- 2018 年導入 Alma 後，館員才得以透過分析功能找出總量。
- 最終發現約有 14,000 筆這類記錄。

(Sun, 2025, p. 271)

使用者不可發現

- 因 Wade-Giles 已不再使用且記錄缺少漢字，**使用者幾乎無法發現這些資源**。
- 轉換為大陸拼音因此成為必要任務。

(Sun, 2025, p. 271)

人工轉換過慢／自動化選項

人工轉換過慢

- 自 2020 年起投入人力進行手動轉換。
- 截至研究時僅更新 2,000 筆；若維持此速度需約 28 年。

(Sun, 2025, p. 271)

自動化選項

- 自動化有多種途徑，例如 Python 程式。
- 機構風險管理政策限制了程式方案，且轉換必須離線進行。

(Sun, 2025, p. 271)

生成式 AI 機會／三種 AI 工具

生成式 AI 機會

- 生成式 AI 的出現提供另一種可能性。
- AI 可處理大量資料與多語內容，有助於縮短轉譯時間並降低人力依賴。

(Sun, 2025, p. 271)

三種 AI 工具

- 研究測試 ChatGPT、Copilot 與 Gemini。
- 選擇原因是三者被視為市場中的主要工具。

(Sun, 2025, p. 272)

工具特性／既有評估

工具特性

- Copilot 定位為支援任務與生產力的 AI 夥伴。
- Gemini 則強調創意生成與內容產生。

(Sun, 2025, p. 272)

既有評估

- 已有許多研究檢視這些工具與其他生成技術。
- 圖資領域的比較分析指出三種工具能迭代產生相對一致的回應。

(Sun, 2025, p. 272)

付費版本／MARC 記錄與題名

付費版本

- 作者預測試顯示免費與付費版使用相同轉譯機制。
- 研究使用付費版，以利用檔案上傳與複製貼上功能。

(Sun, 2025, p. 272)

MARC 記錄與題名

- 研究透過 Alma 擷取 Wade-Giles 編目項目的 MARC 記錄。
- 由於作者名太少，研究決定只轉換題名。

(Sun, 2025, p. 272)

測試樣本／自訂字典

測試樣本

- 研究隨機抽取 1,000 筆 Wade-Giles 題名。
- 每種工具先以預設轉譯規則進行測試。

(Sun, 2025, p. 273)

自訂字典

- 第二階段上傳自訂字典並要求 AI 遵循。
- 字典包含 Wade-Giles 音節及對應的大陸拼音。

(Sun, 2025, p. 273)

工作階段客製化／針對性修正

工作階段客製化

- 此設定旨在讓每次工作階段的轉譯規則一致。
- 提示語採簡潔直接形式，以減少混淆並提高可重製性。

(Sun, 2025, p. 273)

針對性修正

- 即使使用字典，各工具仍會產生不同錯誤。
- 研究透過多次修正來盡量降低不準確性。

(Sun, 2025, p. 273)

不準確程度不一／檔案上傳與複製貼上

不準確程度不一

- 測試顯示，不準確與錯誤受到工具及輸入輸出方式影響。
- ChatGPT 在轉換測試中的分數意外偏低。

(Sun, 2025, p. 273)

檔案上傳與複製貼上

- 資料提供方式會影響準確率。
- 檔案上傳與將題名貼入文字框會得到不同結果。

(Sun, 2025, p. 273)

上傳檔案準確率／自訂字典效果

上傳檔案準確率

- 在檔案上傳模式下，Gemini 與 Copilot 準確率高於 ChatGPT。
- ChatGPT 在內建模型、自訂字典與針對性修正後皆約低於 60%。

(Sun, 2025, p. 274)

自訂字典效果

- 研究要求 AI 優先使用自訂字典。
- 然而結果未顯著改善。

(Sun, 2025, p. 274)

Table 1. Accuracy rates of conversions by AI tools with uploaded files.

	Built-in Model	Custom Dictionary	Targeted Refinement
ChatGPT	≤59%	≤57%	≤60%
Copilot	≤70%	≤70%	≤72%
Gemini	≤83%	≤81%	≤83%

Table 2. Accuracy rates of conversions by AI tools with copying-pasting titles.

	Built-in Model	Custom Dictionary	Targeted Refinement
ChatGPT	≤80%	≤79%	≤81%
Copilot	≤86%	≤86%	≤87%
Gemini	≤87%	≤87%	≤88%

(Sun, 2025, pp. 274–275)

輸出格式效果／螢幕顯示結果

輸出格式效果

- 輸出格式對準確率有顯著影響。
- 若要求產生下載檔，準確率可能降至個位數甚至零。

(Sun, 2025, p. 274)

螢幕顯示結果

- 若要求直接在螢幕顯示結果，準確率明顯較高。
- 差異可能來自 AI 的技術限制與工作流程。

(Sun, 2025, p. 274)

複製貼上準確率／複製貼上限制

複製貼上準確率

- 與檔案上傳相比，複製貼上原始題名使所有工具準確率提升。
- 表 2 彙整複製貼上模式的準確率。

(Sun, 2025, p. 275)

複製貼上限制

- 複製貼上法有明顯限制。
- 每次只能處理有限題名；本研究設定為 100 筆。

(Sun, 2025, p. 275)

輸出不一致／策略未能完全解決

輸出不一致

- 輸出不一致是整個測試過程的主要挑戰。
- ChatGPT 的準確率在不同測試間波動尤其顯著。

(Sun, 2025, p. 275)

策略未能完全解決

- 研究嘗試更換檔案格式、分批處理與改變輸出格式。
- 這些策略都未能完全解決不一致問題。

(Sun, 2025, p. 276)

效能退化／Token 限制與不可預測行為

效能退化

- 重複測試時不一致更頻繁。
- 作者將此類現象連結到 AI 的效能退化或模型衰退。

(Sun, 2025, p. 276)

Token 限制與不可預測行為

- 輸入與輸出 token 限制會導致不可預測行為。
- 可能出現截斷、跳過題名或不一致轉譯。

(Sun, 2025, p. 276)

Gemini 與 Copilot 的脈絡理解／工具特定範例

Gemini 與 Copilot 的脈絡理解

- Gemini 與 Copilot 雖不完美，但整體優於 ChatGPT。
- 兩者較能依脈絡與意義調整轉寫。

(Sun, 2025, p. 276)

工具特定範例

- Gemini 與 Copilot 能將 han yu 視為「漢語」整體詞組。
- ChatGPT 則較嚴格依字典或字面意義，可能造成不自然結果。

(Sun, 2025, p. 277)

平衡取徑／從書目記錄之文本到資料

平衡取徑

- 本研究強調生成式 AI 在轉寫任務中的轉型潛力。
- 不一致輸出顯示仍需結合人工監督與 AI 流程。
- 未來應精進模型並推動圖書館與開發者合作。

(Sun, 2025, p. 280)

ChatGPT生成MARC21書目紀錄之 可行性研究

黃文彥^a 陳亞寧^{b*}

摘要

本文旨在探討 ChatGPT 自動生成 MARC21 書目紀錄的可行性研究，採取正確率、錯誤率、新增率、缺漏率與相似度等多元指標，以評量 ChatGPT 自動生成書目紀錄的效益。本文自國家圖書館全國圖書書目資訊網 (National Bibliographic Information Network, 簡稱 NBINet) 書目資料庫選取 55 筆預行編目資料為研究樣本，以包含預行編目資料的提示語要求 ChatGPT-4o 生成對應的 MARC21 書目格式紀錄後，再以 NBINet MARC21 紀錄為基準，評估 ChatGPT-4o 之自動化後設資料生成的能力。結果顯示 ChatGPT 生成書目紀錄之正確率為 98.4%、錯誤率為 1.6%，而相較於 NBINet 書目紀錄，ChatGPT 生成書目紀錄之新增率為 102% 與缺漏率為 27%。在相似率方面，委由 ChatGPT 以五等級李克特表評判 ChatGPT 生成紀錄與預行編目資料間之平均相似度可達 3.39。除此之外，本文也發現 ChatGPT 具有 MARC21 書目資料之自動化格式對照與轉換、新增、修正與補充等能力外，也提出錯誤的類型 (如資料之錯誤與重複)。

關鍵詞：人工智慧、ChatGPT、資訊組織、自動化後設資料生成、
自動化編目

[https://doi.org/10.6120/JoEMLS.202511_62\(3\).0057.RS.AM](https://doi.org/10.6120/JoEMLS.202511_62(3).0057.RS.AM)

從書目記錄之文本到資料

從書目記錄之文本到資料

- Zaragoza, T., Nicolas, Y., & Le Provost, A. (2022). From text to data inside bibliographic records: Entity recognition and entity linking of contributors and their roles from statements of responsibility. IFLA WLIC2022 Satellite Conference on Artificial Intelligence, Jul 2022, Galway, Ireland.
<https://repository.ifla.org/items/65edd46c-f668-4a73-b829-e488ff99b8a9>

。

責任敘述中的貢獻者與角色，透過實體辨識與實體連結轉化為結構化資料

Sudoc 是法國高等教育聯合目錄，由 Abes 經營。

Sudoc 是法國高等教育聯合目錄，
由 Abes 經營。

(Zaragoza, Nicolas, & Le
Provost, 2022, p. 1)

作為擁有一千五百萬筆記錄的大型
資料庫，Sudoc 的資料品質問題會
影響使用者經驗與維護工作。

(Zaragoza, Nicolas, & Le
Provost, 2022, p. 1)

資料品質問題多樣，包括不正確、模糊、分類錯置、重複、不一致或缺漏。

資料品質問題多樣，包括不正確、模糊、分類錯置、重複、不一致或缺漏。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 1)

有些資料並非真正缺失，而是隱藏在書目記錄內部的文字之中。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 1)

貢獻者姓名與角色會由文件轉錄到 MARC 描述欄位，也就是責任敘述。／研究背景

貢獻者姓名與角色會由文件轉錄到 MARC 描述欄位，也就是責任敘述。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 1)

研究背景

- 此流程是 Abes 首次採用機器學習，並建立由館員參與其中的一般化方法。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 2)

圖書館目錄是歷史悠久、規模龐大且高度結構化的資料庫，由具有強烈品質倫理的專業人員建立與維護。

圖書館目錄是歷史悠久、規模龐大且高度結構化的資料庫，由具有強烈品質倫理的專業人員建立與維護。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 2)

在鏈結資料網路中，較差資料污染良好資料始終是一項持續風險。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 2)

不完整的書目記錄／英文摘要

不完整的書目記錄

- 本文描述從責任敘述文字中抽取貢獻者與角色之結構化資料的努力。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 2)

英文摘要

- 研究目標是自動產生或修正檢索點，包含姓氏、名字、功能關係代碼 (relator, 著作方式)，以及可選的 IdRef 識別碼。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 2)

B200\$f 與 B200\$g 欄位依據文件上所見內容記錄責任敘述。

B200\$f 與 B200\$g 欄位依據文件上所見內容記錄責任敘述。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

責任敘述轉錄文件原文，僅提供最低限度結構化：每一角色或功能對應一個責任敘述。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

這些高度結構化資料來自編目員對文件的智識分析，而不是題名頁的轉錄。

這些高度結構化資料來自編目員對文件的智識分析，而不是題名頁的轉錄。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

每一行稱為一個檢索點，並指向一位唯一的人。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

人類可輕易數出（i.e., 識別）責任敘述中有六位不同人物，並與五個檢索點比較。

人類可輕易數出（i.e., 識別）責任敘述中有六位不同人物，並與五個檢索點比較。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

然而，自動得出這個結論並不容易。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

三十萬個作者檢索點缺少關係代碼，這正是研究欲填補的缺口。／
人物與角色實體抽取

三十萬個作者檢索點缺少關係代碼，
這正是研究欲填補的缺口。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 3)

人物與角色實體抽取

- 名稱實體辨識(Name Entity Recognition, NER)是自然語言處理在大量非結構化文字上抽取實體的應用。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

使用的 fr_core_news_lg 模型以 7200 篇人工標註法文維基百科文章與 3000 多句法文語句訓練。

使用的 fr_core_news_lg 模型以 7200 篇人工標註法文維基百科文章與 3000 多句法文語句訓練。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

該模型可抽取人物、組織與地點，但不能抽取角色。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

Prodigy 是一套標註工具，可讓研究團隊更快速、容易地標註大量資料。

Prodigy 是一套標註工具，可讓研究團隊更快速、容易地標註大量資料。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

使用 Prodigy 標註一萬筆書目記錄後，研究團隊可重新訓練 spaCy 模型，以改善人物抽取並開始抽取角色。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

新模型可用精確率(precision)、召回率(recall)與準確率(accuracy)三項指標評估。

新模型可用精確率(precision)、召回率(recall)與準確率(accuracy)三項指標評估。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 4)

重新訓練後，模型能偵測 99% 的人物與角色實體，且其預測有 99% 為正確。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 5)

研究已可使用 NER 模型分析責任敘述，偵測其中提及的人物與表示其角色的關鍵詞。

研究已可使用 NER 模型分析責任敘述，偵測其中提及的人物與表示其角色的關鍵詞。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

角色關鍵詞連結

- 下一步是將角色關鍵詞連結到控制的角色代碼。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

UNIMARC 約有 155 種不同的關係代碼。

UNIMARC 約有 155 種不同的關係代碼。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

對人類而言，在電影情境中將 réalisateur 連結至 300 電影導演相當容易。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

若缺乏足夠脈絡，消除歧義會變得困難。／因此，研究改採文本分類方法。

若缺乏足夠脈絡，消除歧義會變得困難。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

因此，研究改採文本分類方法。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

文本分類是一種機器學習技術，透過詞頻與預先定義特徵，將開放式文本分派到預定類別。

文本分類是一種機器學習技術，透過詞頻與預先定義特徵，將開放式文本分派到預定類別。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

人類工作限於選擇特徵、抽取資料、建立訓練集，以及選擇最適合問題的演算法。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

研究提供的特徵包括角色關鍵詞、文件內容類型、責任敘述位置與文件類型。

研究提供的特徵包括角色關鍵詞、文件內容類型、責任敘述位置與文件類型。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

研究以 NER 模型從責任敘述中抽取角色關鍵詞與人名。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 6)

研究比較人物實體與檢索點中的姓名字串，並用 Levenshtein 距離處理縮寫、複合姓名與可能拼字錯誤。

研究比較人物實體與檢索點中的姓名字串，並用 Levenshtein 距離處理縮寫、複合姓名與可能拼字錯誤。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 7)

由於訓練集必須盡可能準確，研究必須忽略無法排除的歧義案例。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 7)

研究能為十三種最常見功能各建立一千筆訓練資料。

研究能為十三種最常見功能各建立一千筆訓練資料。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 8)

為避免建立有偏的預測模型，研究必須建立平衡訓練集。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 8)

K 最近鄰方法／預測新點類別時，KNN 會尋找其三個最近鄰並採多數決。

K 最近鄰方法

- 研究先實作 KNN(K-Nearest Neighbour)模型作為可行性初測，再嘗試多種演算法並選擇最適合者。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 8)

預測新點類別時，KNN 會尋找其三個最近鄰並採多數決。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 9)

在訓練與測試資料有限的情況下，結果顯示整體預測模型尚稱良好。

在訓練與測試資料有限的情況下，
結果顯示整體預測模型尚稱良好。

(Zaragoza, Nicolas, & Le
Provost, 2022, p. 10)

最大的落差出現在 070 作者類別
(070 Author)的預測欄位。

(Zaragoza, Nicolas, & Le
Provost, 2022, p. 10)

Label	Precision	Recall	F1-Score	Support
005	0,99	0,97	0,98	224
065	0,93	0,78	0,85	211
070	0,44	0,91	0,6	196
100	0,96	0,81	0,88	217
230	0,99	0,98	0,98	179
300	0,92	0,98	0,95	191
340	0,95	0,61	0,74	201
365	0,97	0,9	0,93	181
440	0,96	0,78	0,86	197
651	0,81	0,6	0,69	58
690	0,94	0,96	0,95	206
727	0,95	0,99	0,97	207
730	0,99	0,89	0,94	205

Figure: Confusion Matrix for KNN model, K=6 , 13k training

(Zaragoza, Nicolas, & Le Provost, 2022, p. 9)

英文摘要／取0.7以下

英文摘要

- 研究注意到兩個專門化「集合箱」：350 雕版師(Engraver)與 673 研究團隊主任(Research team head)。無法識別相關的角色-040 Artist, 560 Originator(原來的PI), 695 Scientific advisor。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 12)

Label	Precision	Recall	F1-Score	Support
003	0,67	0,01	0,03	114
005	0,9	0,99	0,94	192
010	0,84	0,69	0,76	146
040	0,15	0,24	0,18	169
065	0,81	0,78	0,79	185
070	0,37	0,9	0,53	194
080	0,84	0,81	0,83	199
100	0,74	0,74	0,74	196

取0.7以下

(Zaragoza, Nicolas, & Le Provost, 2022, pp. 10-11)

180	0,96	0,93	0,95	222
205	0,5	0,78	0,61	218
212	0,82	0,8	0,81	175
220	0,76	0,59	0,66	136
230	0,91	0,94	0,93	203
273	0,47	0,44	0,45	213
300	0,92	0,95	0,94	198
340	0,67	0,35	0,46	204
350	0,62	0,86	0,72	196
365	0,96	0,88	0,92	199
410	0,93	0,87	0,9	190
440	0,83	0,62	0,71	213
460	0,76	0,64	0,69	181
470	0,71	0,37	0,49	147
550	0,92	0,88	0,9	132
555	0,38	0,08	0,14	173
595	0,53	0,25	0,34	201
600	0,97	0,72	0,83	199
651	0,61	0,47	0,53	185
673	0,21	0,86	0,34	217
690	0,97	0,86	0,91	222
710	0,88	0,59	0,7	192
727	0,85	0,08	0,16	198
730	0,99	0,88	0,93	210
956	0,88	0,51	0,65	192
958	0,82	0,48	0,61	205

後續工作／研究計畫大幅增加可用於文本分類模型訓練的資料量。

後續工作

- 後續工作首先希望最佳化並改善既有成果。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 12)

研究計畫大幅增加可用於文本分類模型訓練的資料量。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 12)

另一種可行解方是針對不同需求建立不同模型。

另一種可行解方是針對不同需求建立不同模型。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 12)

研究可只採納「足夠確定」的預測，忽略「不確定」的結果。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 13)

結論

結論

- 本專案目標是訓練模型分析責任敘述文字，自動產生缺漏的貢獻者檢索點，或為既有檢索點補上缺漏功能關係代碼。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 13)

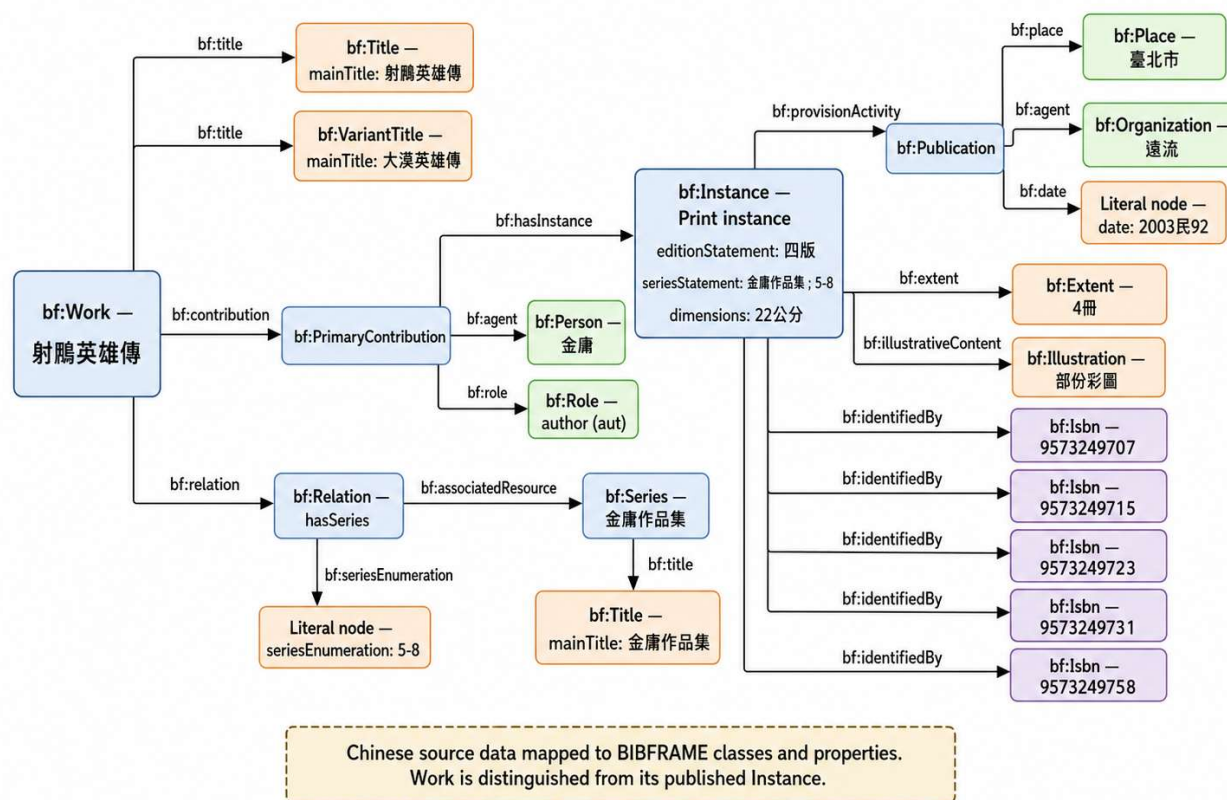
藉由有效率的標註與資料準備工具，並讓館員參與其中，圖書館可同時在資料品質與人力資源發展上取得實質進展。

(Zaragoza, Nicolas, & Le Provost, 2022, p. 13)

BIBFRAME知識圖譜

BIBFRAME Entity-Relationship Graph for 射鵬英雄傳

Based on the BIBFRAME Work-Instance model



結語

- 品質與正確性仍是最高原則
- 雖然AI可以提昇效率，卻陷入正確性與效率間之拔河拉距
- AI一般使用原則是
 - 問對問題
 - 轉化人類問題為AI解譯的問題
 - 提示詞／語工程（prompt engineering）愈形重要
 - 選對適用的AI工具平臺
 - 使用環境的個人化設立
- 個人隱私與資安仍然很重要，尤其是資料
- 思辨（critical thinking）仍是使用AI的最高指導原則

備註

- 本份簡報係使用ChatGPT 5.5 & 5.6 Sol之High版本生成後，經個人逐一交叉訂正。
- 使用期間：2026-07-10至2026-07-14。

盡力而為，敬請指教！

若有任何疏漏，還望見諒。