
从人工到智能：图书馆书目编目自动化新探索 MDPI Publications

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/39664.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

从人工到智能：图书馆书目编目自动化新探索 MDPI Publications。论文标题：Research on Large Language Model-Based Bibliographic Cataloging Agent in the CNMARC Context

论文链接：<https://www.mdpi.com/2304-6775/14/1/19>

期刊名：Publications

期刊主页：<https://www.mdpi.com/journal/publications>

一、概要

随着图书馆数字化进程的加快，面对激增的文献资源，传统的人工编目方式在效率和成本上面临巨大挑战。为了探索更高效的资源组织路径，中山大与中国人民大学的研究团队在Publications期刊上发表了题为的论文。该研究尝试将大语言模型和智能体技术引入书目数据加工流程，通过构建自动化 workflow 来辅助完成复杂的编目任务，为智慧图书馆的建设提供了新的思路。

二、研究过程与结果

研究团队设计了一套基于大语言模型的自动化编目框架，核心由四个智能体组成：元数据提取智能体（Metadata Extraction Agent, MEA）、描述性编目智能体（Description Cataloging Agent, DCA）、主题分析与标引智能体（Subject Analysis Indexing Agent, SAIA）以及质量控制智能体（Quality Control Agent, QCA）。这一框架模拟馆员实际操作逻辑，将编目任务拆解为相互衔接的环节，实现从原始资料到标准书目数据的全流程转化。

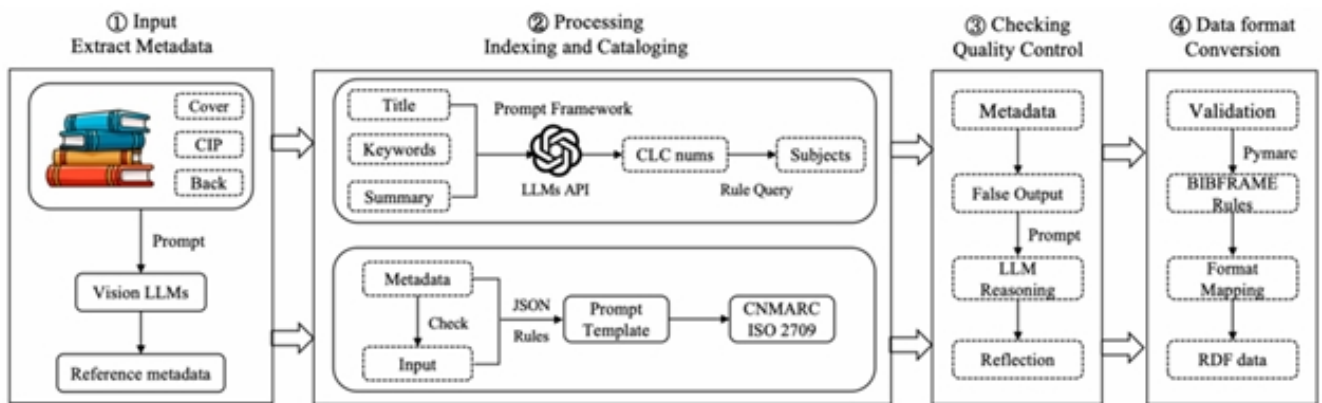


图1 自动编目框架设计

首先，元数据提取智能体负责从封面、版权页等原始资料中识别题名、责任者和出版项等关键要素；随后，描述性编目智能体按照CNMARC规范将这些要素转化为标准字段，确保记录符合行业要求。接下来，主题分析与标引智能体对文献内容进行深度理解，给出分类号和主题标引建议；最后，质量控制智能体对生成结果开展逻辑校验，修正格式偏差或信息遗漏，提升整体可靠性。为检验实际效果，研究团队严格参照《中国图书馆分类法》第五版的标准要求，利用某高校图书馆提供的三万余条真实书目记录进行了大规模实验测试。

表1 多模型元数据识别准确率

Model	Precision	Recall	F1 Score
GLM-4.5V	0.8373	0.8844	0.8602
Doubao-Seed-1.6-vision	0.7700	0.8412	0.8040
qwen3-VL-max	0.8830	0.8962	0.8896
Kimi-K2-Instruct-0905	0.7797	0.8214	0.8000

表2 多模型描述性编目准确率

Model	Precision	Recall	F1
DeepSeek-V3.2 Exp	0.9804	0.9796	0.9797
Kimi-K2-Instruct	0.6861	0.5016	0.5212
Doubao-Seed-1.6	0.6313	0.7072	0.6671
qwen3-max	0.8391	0.8689	0.8444

表3 多模型《中图法》标引准确率

CLC Hierarchy	Precision	Recall	F1 Score	Sample Size
1	0.8663	0.8140	0.8393	26,355
2	0.7667	0.6992	0.7314	26,355
3	0.6858	0.6173	0.6498	26,267
4	0.5799	0.5257	0.5515	25,770
5	0.5094	0.4586	0.4827	22,731
6	0.4591	0.4014	0.4283	16,582
7	0.4486	0.3557	0.3968	9364
8	0.4848	0.3473	0.4047	5052
9	0.5010	0.3458	0.4092	2104
10	0.5193	0.3439	0.4138	880

同时，各选一本中文图书和一本英文图书作为实例验证，考察该编目框架在实际编目场景中的表现。结果显示，该框架在元数据提取和主题分析环节均达到较高准确率，初步实现了高效的自动化转化。

表4 基于CNMARC的中英文图书编目实例（ISO 2709格式）

00764 a2200217 4500001000800000010003200008100004100040101000800081102001500089105001900104200006000123210004000183215002800223225002900251330014600280606004700426690001600473701001600489702001900505801002200524 2025001 -a978-7-308-24237-0-dCNY88.00 -a20250101d2025 em y0chiy0121 ea 0 -achi -aCN-b330100 -aay z 000yy 1 -a太平无相-e侍王府如是说-f赵威,刁常宇著 -a杭州-c浙江大学出版社-d2025 -c图-d710mm×1000mm 1/16 -a走向平衡系列丛书 -a本书以太平天国侍王府为研究对象,深入探讨其作为古建筑与文物的历史价值、保护现状与未来研究方向。0 -a太平天国-x侍王府-x古建筑-x研究 -aTU-092.2-v5 0 -a赵威-4著 0 -a刁常宇-4著 0-aCN-bZJU-c20250125
00725nam0a2200217 45000100028000001000035000281010008000631020011000711050014000822000086000096205001900182210005000201215001300251225002400264330010500288606002700393690012004207010210043280102600453# \$a978-1-449-37143-2\$dUS \$39.99# \$a20260301d2013 em y0engy0103 ba#0 \$aeng# \$aUS\$bCA# \$ay z 000yy#1 1\$aLearning SPARQL\$eQuerying and Updating with SPARQL 1.1\$fBob DuCharme# \$aSecond Edition# \$aSebastopol, CA\$cO'Reilly Media, Inc.\$d2013# \$c图\$d24cm# \$aA Nutshell Handbook# \$a本书通过大量实例讲解如何使用SPARQL 1.1对RDF数据集进行检索、操作及数据更新,涵盖关联数据与语义网核心概念。#0 0\$aSPARQL\$x程序语言\$x教材# \$aTP312\$5v5#0 0\$aDuCharme\$bBob\$4著#0 0\$aCN\$bGemini\$c20260301#&

三、结论与前景

该研究证明了自动化编目在提升图书馆工作效率方面的潜力，尤其是在处理标准化程度较高的书目数据时能减轻人工负担。尽管目前的系统在逻辑上更倾向于预设好的单智能体线性执行，而非具有复杂信息流和交互特征的多智能体系统，但其展现出的性能已足以作为未来智能化编目系统的实验基准。未来研究将致力于提升模型在处理疑难文献时的精细度，文献复分等复杂问题，并结合检索增强生成、大模型微调进一步优化流程以适应多样化的文献加工需求。

Publications 期刊介绍

主编：Andrew Kirby, Arizona State University, USA

期刊主题涵盖学术出版，学术交流以及学术文化的各个方面；涉及学术交流理论与实践、开放获取模式、学术文献数据库、出版伦理道德、

版权、数字化出版及其他创新技术、同行评审、学术评估及其影响等热门研究话题。现已被ESCI (Web of Science)、Scopus、DOAJ、CNKI等重要数据库收录。

2024 Impact Factor 2.5 2024 CiteScore 10.0 Time to First Decision 27.4 Days Acceptance to Publication 4.8 Days

来源：Publications

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发