

知识图谱视域下高校文书档案著录的智能化实践路径

陈甜甜

广州理工学院

摘要: 本文在分析知识图谱赋能高校文书档案著录四大核心优势的基础上,结合高校档案馆的实践探索,提出“本体构建—实体抽取—知识存储—应用服务”四步走实践路径。研究认为,知识图谱视域下的著录转型不仅是技术工具的升级,更是档案知识组织范式的重构。

关键词: 知识图谱; 高校文书档案; 档案著录; 智能化

DOI: 10.65976/3080-0374.2026.08.014

引言

高校文书档案著录是档案资源与开发利用的基础性工作。传统著录模式以案卷或文件为单位进行条目式描述,存在语义关联薄弱、著录内容粗糙等问题,无法满足现代用户需求。而知识图谱技术以“实体—关系—属性”的三元组结构组织信息,能够将分散的档案数据转化为高度关联的知识网络,为档案著录的智能化转型提供了新的技术视域,将档案馆藏资源利用起来。《“十四五”全国档案事业发展规划》中明确支持知识管理、人工智能等技术在档案信息深加工中的应用^[1],浙江金华^[2]、东北财经大学^[3]、东南大学^[4]等档案馆已实践探索出知识图谱技术的应用价值。本文聚焦知识图谱视域下高校文书档案著录的智能化,系统分析其核心优势,并提出可操作的实践路径。

一、知识图谱赋能高校文书档案著录的核心优势

(一) 从条目描述到实体关联: 著录范式根本转变

传统著录以“件”或“卷”为基本单位,形成的是线性条目列表,用户检索时仅能通过关键词、全文等有限字段进行匹配。这种著录范式本质上仅是对实体档案的“描述”,方便查询利用,告诉用户“有什么文件”,却难以揭示“文件之间有什么关系”。知识图谱则将著录对象,如机构、题名、时间、责任者、文种等均转变成为可独立标识的知识节点,著录工作的核心任务发生根本性转变:不再是“对文件进行描述”,而是“对知识进行建模”。简单来说,学校的一位知名教授的人事档案、其主持的科研项目、教的课程、指导的学生等档案信息,在传统模式下是分属在不同部门、不同类别里的,而知识图谱技术可将这些分散的档案资源关联并可视化地表达出统一的人物知识画像^[5]。

(二) 从关键词匹配到语义查询: 检索实现质的飞跃

高校档案利用者群体多元,需求场景复杂。行政部门可能需要查找一些学校、部门印发政策制度等,学生个人可能需要查找学籍、成绩证明等,教师可能需要查找个人荣誉文件、人事档案等。传统关键词检索方式难以满足这些差异化需求。知识图谱支持的语义查询,能够理解自然语言中的深层意图。例如:用户以“找一下近十年学校关于职称评审的文件”提问,系统能够自动识别“近十年”为时间范围(2017—2026)、“学校”为责任主体(本校)、“职称评审”为事由主题,从图谱中检索出符合条件的文件。用户进一步追问“评审条件有哪些变化”,系统可通过对比不同时期文件中的关键词与条款,自动生成变化对比摘要。这一能力的实现,本质上依赖知识图谱对档案内容语义关系的深度建模。

(三) 从静态固化到动态生长: 知识体系持续进化

传统档案著录工作完成后,相关数据内容便处于固定固化状态。当新增档案资源入库时,往往需要开展重复性人工著录工作,新增档案条目与原有馆藏条目之间也缺少自动化的关联匹配手段。这种静态化、封闭化的管理模式,导致档案知识体系无法依托资源增量持续迭代优化。区别于传统模式,知识图谱具备良好的动态拓展与迭代更新能力。新档案入库时,系统可自动进行实体识别与关系抽取:识别新文件中的机构、人物、事由等实体,并与图谱中的已有节点建立关联。高校文书档案具有明显的“年度周期性”与“事件序列性”。学位评定、职称评审、招生录取、毕业生就业等业务每年循环发生,形成规律性的事件序列。知识图谱能够自动识别同一事件的年度序列(如“2021—2025年度毕业生就业情况”),并关联相关

文件，为学校未来趋势分析、对比研究提供结构化数据支撑。

(四) 从人工标引到人机协同：著录效率的显著提升

高校档案馆普遍面临人手不足的情况，传统著录主要是人工手动著录，一份文件的著录可能需要数十分钟，年度新增档案著录占用大量人力，会导致档案存量积压愈演愈烈。而基于知识图谱的智能著录系统可自动提取档案文件中的实体信息并生成著录草案，著录人员仅需核验修正即可完成。根据鹿城区档案馆实践数据，单件档案著录时间从15分钟缩短至3分钟，效率提升80%。

二、案例分析：东北财经大学档案馆的智能化实践^[3]

(一) 实践背景与动因

东北财经大学档案馆距今已成立七十多年，馆藏资源丰富，但档案资料难梳理、难利用，尤其是关联信息难挖掘。虽然学校档案馆此前已完成校史文献、历史照片、科研成果等数万份档案的数字化转化，积累了海量基础数据，但传统研究模式依赖人工逐份检索、手动关联，既难以高效激活数据价值，也无法系统提炼数据背后的规律，更难以以为办学决策提供精准的历史参考。

在此困境下，2025年东北财经大学档案馆与学校智慧校园建设中心协同合作，以人工智能技术为驱动，对馆藏档案资源进行深度挖掘与数据价值转化，初步构建起校史知识图谱。

(二) 知识图谱构建流程

东北财经大学的校史知识图谱构建遵循“多源异构数据采集—智能解析与实体抽取—知识存储与图谱构建—智能应用与决策支持”的技术路线，其核心架构如表1所示。

(三) 应用成效分析

东北财经大学档案馆利用知识图谱与AI技术，系统分析了学校七十余年的发展历程趋势、学科演变及人员流动等宏观问题，形成了涵盖“发展目标与定位迭代之路”“不同发展阶段的共同点”“不同阶段

发展重点演变”等核心议题的研究报告。这一成果不仅揭示了学校发展的内在规律与历史经验，更实现了档案工作从“被动查找资料”向“主动挖掘数据价值”、从“单点资料收集”向“系统化数据研判”的根本性转变，使档案工作真正从资料沉淀的“末端”走向了辅助决策的“前端”。

(四) 经验与启示

东北财经大学档案馆以校史研究为切入点，利用其实体类型稳定、关联关系明确等特点作为知识图谱建设的“试验田”，利用人工智能技术作为打通“数据—研究—决策”链路的关键抓手，通过AI技术与档案管理的深度融合，跨部门、跨学科的创新实践，证明高校档案数字化建设大有可为。

三、高校文书档案著录智能化转型的实践路径

基于上述理论分析与案例验证，本文提出知识图谱视域下高校文书档案著录智能化转型的“四步走”实践路径（如图1所示）。

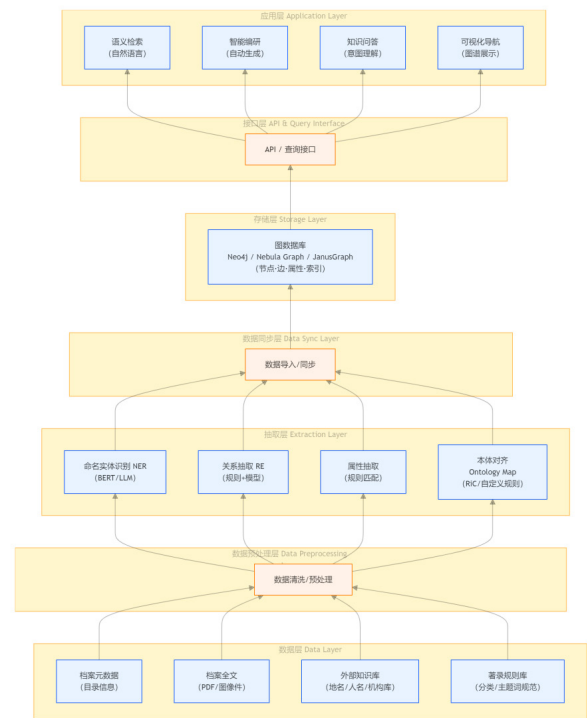


图1 知识图谱驱动的档案著录技术架构图

表1 东北财经大学校史知识图谱核心架构表

层级	功能	技术实现
数据层	多源异构档案汇聚	校史文献、历史照片、科研成果等数万份数字化档案
解析层	智能化档案解析	PaddleOCR（竖版手写体/扫描件/低光照图片识别）
抽取层	实体与关系抽取	Qwen系列模型（Turbo与Plus版本）语义理解与推理
存储层	知识图谱存储	Neo4j图数据库
应用层	智能分析与决策支持	机构演变分析、学科趋势研判、人员流动分析

(一) 本体构建: 建立高校文书档案领域知识模型
本体是知识图谱的“骨架”, 决定了知识组织的逻辑框架。高校文书档案本体的构建, 需要在档案学元数据传统与知识工程语义要求之间寻求平衡。

在实体类型设计上, 应覆盖高校文书档案的核心要素, 除“题名”“责任者”“文号”“成文日期”等通用的著录项外, 还应结合高校特点增设“组织机构”(院系、部处、研究中心)、“人员类别”(教职工、学生、校友)、“教学活动”(招生、培养、学位授予)、“科研项目”(立项、结题、成果转化)等特色实体类型。在关系类型设计上, 应涵盖档案实体间的典型关联: 隶属关系、签发关系、承办关系、引用关系、时间序列及因果关联等。关系设计的丰富程度, 直接影响知识图谱的语义表达能力。

(二) 实体与关系抽取: 从档案文本到知识三元组
实体与关系抽取是知识图谱构建的核心环节, 其目标是将非结构化或半结构化的档案文本转化为结构化的知识三元组。

在技术路线上, 可采用“规则+模型”的混合策略。对于结构较为规范的档案元数据(如文号、成文日期、责任者等), 可通过规则匹配直接提取; 对于非结构化的档案全文(如会议纪要、领导讲话、工作总结等), 则需要借助自然语言处理技术进行实体识别与关系抽取。可借鉴东北财经大学 Qwen 系列模型, 对校史档案中的实体及关系进行深度抽取与语义增强。

(三) 知识存储与图谱构建: 搭建可计算的语义网络

实体与关系抽取工序完成后, 需搭建科学合理的知识图谱系统。其不仅需具备基本的存储架构, 还需兼具查询检索与逻辑推理能力。从技术应用现状来看, 图数据库是承载档案知识图谱建设的主流技术载体。以 Neo4j、Nebula Graph 为代表的图数据库, 依托节点与边的拓扑结构完成知识要素存储, 可高效适配多跳关联检索、深层路径挖掘等复杂需求, 这与档案资源关联化管理要求完美契合。东北财经大学选用的是 Neo4j 作为底层存储工具, 完成了校史档案知识图谱的搭建。

(四) 应用服务与迭代优化: 从技术验证到价值落地

知识图谱的最终价值体现在应用服务中。基于知识图谱, 可开发语义检索、智能问答、可视化导航、知识编研等多种应用。

在语义检索层面, 可实现自然语言化的查询交互

模式, 能够结合上下文语境开展语义推理, 实现多模态智能检索应用, 联动整合馆藏内具有内在关联的所有档案资源。在智能编研场景中, 依托语义解析与知识图谱融合技术, 平台可自动挖掘全宗年度档案中的核心人物、关键事件、重要时间节点等关键信息要素, 快速完成全宗指南编撰、机构沿革梳理、年度大事记整理等基础性编研工作。面向知识咨询服务场景, 用户可通过智能问答机器人开展自主查询, 依托用户意图识别、图谱知识检索、内容智能整合的完整逻辑链路, 形成一体化、闭环式的档案知识服务模式。

此外, 知识图谱的建设应建立用户使用反馈机制, 利用查询日志、点击流等数据持续优化实体识别模型与关系权重, 形成“使用—反馈—迭代”的良性循环。

四、结语

本文以知识图谱技术为着力点, 探索高校文书档案著录的智能化转型。在理论层面, 阐明了知识图谱在实体关联、语义查询、动态生长、人机协同等方面的突出优势。在实践层面, 以东北财经大学档案馆为典型案例, 其利用 OCR 模型、Qwen 系列模型、Neo4j 图数据库等技术, 完成了校史知识图谱构建, 充分证明知识图谱技术的可行性, 进而提出“本体构建—实体抽取—知识存储—应用服务”四步走实践框架。研究认为, 著录转型不仅是技术工具的升级, 更是档案知识组织范式的重构。

参考文献:

- [1] 中华人民共和国国家档案局. 中办国办印发《“十四五”全国档案事业发展规划》[EB/OL]. (2021-06-09)[2025-09-04]. <https://www.saac.gov.cn/daj/yaow/202106/899650c1b1ec4c0e9ad3c2ca7310eca4.shtml>.
- [2] 中华人民共和国国家档案局. 浙江金华 利用档案知识图谱深度开发红色资源 [EB/OL]. (2025-05-23)[2026-01-04]. <https://www.saac.gov.cn/daj/c100206/202505/e2883b2b5f494e87a79bd5b0cdc90190.shtml>.
- [3] 东北财经大学. 数据驱动, 让档案工作走向“前端” [EB/OL]. (2025-11-07)[2026-03-04]. https://www.dufe.edu.cn/content_95039.html.
- [4] 孙婷婷, 赵爽. 数据要素化驱动的高校融媒体档案管理转型研究——基于东南大学的困境解析与实践进路 [J]. 档案与建设, 2025(7):107-113.
- [5] 舒志梅. 数字人文背景下的档案知识图谱构建研究 [J]. 山西档案, 2020(2):53-60.