

基于大语言模型的交通工程电子档案智能生成与精准检索技术研究

胡琼秀, 徐 凯*

浙江交投高速公路建设管理有限公司 浙江杭州

【摘要】人工智能技术的快速发展,尤其是大语言模型(LLM)在自然语言处理领域的突破,为电子档案的智能生成与精准检索提供了新的技术路径。本文以瑞苍高速公路项目为依托,探讨基于大语言模型的交通工程电子档案智能生成与精准检索技术,通过构建领域知识库、优化语义理解算法,实现档案内容的自动生成、合规性审核及高效检索,从而提升档案管理效率,降低碳排放,为交通工程全生命周期数字化管理提供支撑。研究旨在解决档案标准化、技术适配性及数据安全等核心问题,助力绿色智慧交通体系建设。

【关键词】大语言模型; 交通工程; 电子档案; 智能生成; 精准检索

【基金项目】本文系浙江省档案局科技项目“碳达峰碳中和战略背景下交通建设工程电子档案研究与应用实践(2024-48)”研究成果之一

【收稿日期】2025 年 5 月 17 日

【出刊日期】2025 年 6 月 8 日

【DOI】10.12208/j.jer.20250300

Research on intelligent generation and accurate retrieval technologies for traffic engineering electronic archives based on large language models

Qiongxiu Hu, Kai Xu*

Zhejiang Jiaotou Highway Construction Management Co., Ltd., Hangzhou, Zhejiang

【Abstract】The rapid development of AI, especially the breakthroughs of LLMs in natural language processing, has opened up new technical avenues for the intelligent generation and precise retrieval of electronic archives. This paper, using the Ruicang Highway Project as a basis, explores the intelligent generation and precise retrieval of traffic engineering electronic archives based on LLMs. By building a domain - specific knowledge base and refining semantic understanding algorithms, it aims to automate and ensure compliance in archive content generation while improving retrieval efficiency. These efforts enhance archive management and reduce carbon emissions, thereby supporting the digital management of traffic engineering throughout its entire life cycle. The study focuses on addressing key issues such as archive standardization, technological adaptability, and data security, with the goal of advancing the green and smart transportation system.

【Keywords】Large language models; Traffic engineering; Electronic archives; Intelligent generation; Precise retrieval

在数字中国及“双碳”战略背景下,交通工程电子档案的智能化管理成为推动行业数字化转型的关键环节。传统档案管理依赖纸质与电子“双套制”,效率低下且资源浪费,而《中华人民共和国档案法》的修订为电子档案单套管理提供了法律依据。与此同时,随着交通工程项目规模不断扩大与信息复杂度日益提升,传统档案管理方式在资料生成效率、检索准确率和数据利用深度等方面面临瓶颈。尤其在数字化转型和“智慧交通”推进背景下,海量非结构化文本档案的快速处理

需求日趋迫切^[1]。大语言模型具备强大的自然语言理解与生成能力,为电子档案的智能生成与语义检索提供了全新技术路径,成为推动交通工程档案管理向智能化、自动化迈进的关键支撑^[2]。

1 基于大语言模型的交通工程电子档案与精准检索技术应用优势

基于大语言模型的交通工程电子档案与精准检索技术在理解工程语义、自动化生成文本及提升检索效率方面展现出显著优势。相较传统关键词匹配方法,大

*通讯作者: 徐凯

语言模型能深度解析工程术语与语境关系, 实现对多源异构数据的语义融合与归类, 提升文档生成的逻辑性与一致性^[3]。同时, 其在检索阶段可根据自然语言提问进行智能匹配, 显著提高查准率与查全率, 满足工程人员在设计、施工、运维等环节中对资料快速、准确调用的实际需求。

2 基于大语言模型的交通工程电子档案智能生成与精准检索技术应用难点

2.1 语义理解偏差影响生成质量

交通工程电子档案内容涵盖设计图纸说明、施工日志、监测报告等多类文本, 其专业术语密集、上下文逻辑复杂。大语言模型虽具备强大的语言生成能力, 但在特定工程语境下容易出现语义偏差, 导致自动生成的档案内容出现技术表达不准确或逻辑结构混乱的问题。这种偏差往往源于预训练语料中工程类文本占比较低, 模型对专业表达的建模深度不足, 尤其在涉及规范条文引用或方案细节描述时更易产生误导性内容, 从而影响档案的有效性与可用性。此外, 若模型缺乏对工程任务逻辑链条的理解, 可能会导致生成内容断层、术语误用或前后矛盾^[4], 对后续文档审批、归档与项目复盘产生干扰, 降低智能生成在业务流程中的可信度与可控性。

2.2 数据结构复杂制约模型匹配

交通工程项目在全生命周期中会积累大量结构化与非结构化数据, 包括 BIM 模型参数、传感器数据、图文报告、合同文件等。这些数据类型之间的逻辑关联性强, 格式多样, 对语言模型提出更高要求。在生成或检索过程中, 大语言模型难以充分解析表格、图像、图纸与文本之间的隐性关系, 造成信息片段化提取或误关联的问题。数据结构复杂性还增加了预处理与标签标注的技术成本, 进一步限制了模型在真实业务流程中的适配效果。同时, 不同系统或单位对档案结构定义不一, 缺乏统一的格式规范, 也使得模型在处理多源信息时难以建立标准化的信息调用路径, 这对生成结果的准确性与检索响应的一致性构成挑战, 亟需配套的数据治理机制加以支撑。

2.3 术语标准不一削弱检索精度

交通工程行业中不同地区、单位或项目对同一工程概念的术语表述存在差异, 导致模型在语义检索时容易产生歧义匹配。例如同一设备或工艺名称在不同资料中存在缩写、别称、翻译等表述差异, 增加了语义统一建模的难度。若缺乏标准化术语库的辅助支持, 大语言模型对查询请求的解析容易偏离用户意图, 影响

检索结果的准确性与覆盖面, 进而削弱技术在高精度资料查找与工程知识管理方面的应用价值。特别是在跨项目调阅、历史案例对比或区域协同建设中, 不统一的术语体系会使检索范围局限, 错失关键信息, 降低模型对工程语言适配能力的泛化性和工程知识图谱构建的完整度, 不利于形成统一的语义资源体系。

2.4 隐私安全风险限制应用范围

交通工程档案常涉及招投标信息、项目预算、设计方案与施工细节等敏感内容, 其数据合规性与信息安全保障至关重要。大语言模型在训练或调用过程中若处理未经脱敏的档案数据, 可能引发隐私泄露或知识产权纠纷等风险。同时, 模型推理过程的不可解释性也使得其输出结果难以满足审计与安全合规要求。这些问题使得相关单位在实际部署中面临较高的数据治理与技术监管成本, 从而限制了技术在关键业务场景中的大规模推广。在国有工程、涉密项目或法律约束严格的环境中, 使用开放式或云端部署的模型更易产生信任障碍, 用户对数据控制权、存储路径及访问记录的安全敏感性较强, 进一步制约了模型与核心业务系统的融合深度与部署范围^[5]。

3 基于大语言模型的交通工程电子档案智能生成与精准检索技术应用策略

3.1 优化语料体系增强语义理解

为提升大语言模型在交通工程领域的语义理解能力, 应系统构建高质量、专业化的语料体系。教师可通过精选项目档案、技术规范、标准手册等文本资源, 建立覆盖设计、施工、监理、养护等全流程的语料库, 并结合语义标注技术提炼术语及逻辑关系, 增强模型对专业语境的适配性。同时, 引导学生掌握领域语言特征与表达逻辑, 使其在训练过程中强化模型对工程文本的识别与还原能力, 从源头提升内容生成的准确性与专业度^[6]。

比如, 在“城市快速路智能化改造”项目中, 工作人员需要针对工程文本中频繁出现的结构加固、沥青摊铺温度控制、交通组织优化等专业术语, 系统构建符合本项目特征的专属语料库。他们将施工组织设计、监理日志、图纸说明等文本资料按工程阶段进行分类整理, 并结合人工标注的方式提取关键技术表达与逻辑结构, 建立包含词汇表、语义标签和上下位关系的专业术语知识图谱。该语料体系不仅用于微调语言模型, 还作为校验工具保障生成文本的术语准确性与逻辑一致性。在此基础上, 模型能在生成日报、进度报告或质量记录时, 自动复用符合工程标准的表达, 减少人工修订

成本,同时提升档案内容的专业性和语境契合度,显著增强信息生成环节的智能化水平。

3.2 规范数据格式统一结构标准

构建统一的数据结构标准是提升模型生成与检索质量的重要前提。教师在教学与实践环节中应注重引导学生理解交通工程数据的层级逻辑与字段属性,通过制定规范的数据模板与格式转换规则,实现文本、图表、图像等多种资料的一体化表达。统一结构不仅有利于模型的高效调用与语义映射,也为后续的信息挖掘与知识图谱构建提供坚实基础。通过强化数据规范意识,可有效降低输入混乱对模型训练与任务执行的干扰风险^[7]。

比如,在“跨江特大桥建设信息化集成”项目中,工作人员需要处理来源多样、格式各异的工程数据,如结构荷载监测数据、材料验收单、施工进度条目、地质勘察报告等。为确保大语言模型能够有效解析这些数据并支持后续生成任务,项目组制定统一的档案数据结构规范,包括字段命名、编码标准、表单布局与嵌入式标签体系,并将纸质扫描件数字化后根据模板进行结构映射。此外,通过建立一体化的数据接口系统,实现文档、图纸、照片与监测曲线之间的逻辑对接,使得模型在调用信息生成月报、专项汇总或风险提示时具备高度的结构清晰性与内容完整性。标准化工作不仅提升了数据输入质量,也为模型在实际工程应用中的稳定运行打下基础。

3.3 融合多模信息提升检索精度

针对交通工程档案多模态特征,应探索融合文本、图像、表格等多种信息形式的策略,以提升检索响应的语义深度与结果匹配的精度。教师可引导学生学习多模态建模原理与跨模态对齐方法,将图纸注释、设备图像与文本说明建立关联通道,使大语言模型能够理解各模态间的语义映射关系。通过搭建具备融合能力的智能系统,有助于模型在处理复杂查询时实现语义完整响应,进而提升系统在工程资料检索过程中的实用性能。

比如,在“地铁车站建设多模档案系统”项目中,工作人员需要构建一个支持文本、图像与结构数据融合检索的智能平台,以满足施工单位对关键构件、节点方案或历史变更记录的快速定位需求。该项目通过采集现场照片、图纸文件与施工日志信息,采用多模态学习算法训练语言模型对不同模态内容进行语义映射统一,将图像中的构件标签与文本中的技术说明建立逻辑链接。在实际操作中,用户可以输入自然语言查询,

如“站厅层梁体裂缝处理方案”,系统能够同时呈现相关图纸节点、裂缝图像及其处理记录文本,确保查询结果具备高语义关联性 & 决策参考价值。该系统显著提升了档案检索的准确度与响应效率,突破了传统文本检索在复杂场景下的适用性限制。

3.4 构建安全机制保障数据隐私

在推动大语言模型落地应用的过程中,必须同步构建覆盖数据输入、处理与输出全流程的安全机制。教师应指导学生掌握数据脱敏、访问控制、模型审计等关键技术手段,强化数据合规处理意识,防止敏感信息在生成或检索过程中被泄露。同时,应在教学中渗透模型可信性与可解释性原则,引导未来工程人员在技术部署中设定权限边界与应用规范,确保智能档案系统在保障隐私安全的前提下稳定运行,服务工程管理实践。

比如,在“智能高速公路全生命周期档案管理”项目中^[8],工作人员需要在确保数据可用的同时构建全面的数据安全保护体系。项目部署前期,团队对涉及项目招标、造价控制、工艺方案等敏感内容进行全面审查,并采用实体识别与信息脱敏技术对关键信息字段进行加密与替代处理。此外,系统设有严格的权限分级管理,只有经过授权认证的用户才能访问特定档案内容,防止非授权人员操作数据。在模型调用过程中,系统启用日志追踪与动态审计机制,记录用户查询内容及模型响应情况,确保过程可回溯、可监管。同时,该项目采用本地化部署方案,确保所有训练数据与推理过程均在安全环境内运行,杜绝数据泄露风险,保障了档案系统在实际运维过程中的数据主权与信息安全,实现智能化应用与隐私合规的平衡。

4 结语

综上所述,本文主要研究了基于大语言模型的交通工程电子档案智能生成与精准检索技术的应用背景、关键优势、现实难点与应对策略。通过分析语义理解、数据结构、术语标准与安全保障等方面的问题,提出了优化语料、统一格式、多模融合与隐私机制构建等路径,旨在为交通工程档案管理的智能化转型提供技术支撑与实践参考。研究表明,大语言模型在工程信息处理中的潜力显著,但其应用需紧密结合工程实际,强化技术适配与系统安全保障。

参考文献

- [1] 王希,胡翔,冀秀娟,等.城市轨道交通工程档案信息化建设影响因素与对策分析——以江苏省五家地铁集团为例[J].档案与建设,2024,(01):101-104.

- [2] 林颖.城市轨道交通工程竣工档案接收的探索与创新——以杭州市城市建设档案馆为例[J].兰台内外,2024,(02):25-27.
- [3] 陈雯,袁月,王俊淘,等.城市轨道交通工程中电子文件元数据归档的难点及对策分析[J].档案管理,2024,(01):101-104.
- [4] 郝丽,顾杰隽,李滢,等.深中通道电子档案全域管理体系与成套技术应用研究[J].兰台世界,2023,(12):69-72.
- [5] 强冬梅,景红,王俊淘,等.我国城市轨道交通工程档案管理标准建设的现状调查及改进策略[J].山西档案,2023,(03):176-181.
- [6] 王世兴.基于知识图谱的轨道交通工程档案数据化研究[D].山西大学,2023.
- [7] 黄天麒.基于 BIM 平台的建设项目电子文档管理研究[D].四川大学,2023.
- [8] 张春景.高速公路电子档案管理系统功能研究[J].兰台世界,2021,(S1):65-66.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>

