面向核电生产文件的大模型业务辅助实践研究

胡丹丹，陈莹，杨一奇
福建福清核电有限公司，福建 福州 350300
DOI: 10.61369/TACS.2026070044

摘　　要： 针对核电企业传统生产文档管理系统检索手段单一、缺少语义交互、文件版本异动缺少自动预警、领域知识碎片化等现实难题，本文立足国产化大模型技术开展核电垂类模型定制研究，综合运用多模态 RAG、多级思维链知识图谱 [3][4]、分层语义解析技术，完成系统总体架构、数据架构与七大功能模块设计，打通系统与 ECM、ITASK 现有业务系统的数据对接，依托3万余份运维规程、操作手册完成知识库构建与模型本地化适配。现场试运行数据表明，系统可显著提升文档查阅效率、实现文件全生命周期变更自动提醒，有效防范误用过期文件带来的安全生产隐患。研究梳理了多模态文档识别、小众样本匮乏、知识图谱实时更新等现存技术短板，指明后续迭代优化与场景拓展方向。通过本研究落地国产大模型在核电文档管理领域的工程化应用，形成可复用的核电领域大模型落地范式，能够为智慧核电知识治理与行业数字化转型提供实践参考。

关 键 词： 智慧核电；领域大模型；知识图谱；多模态 RAG；生产文档管理

Research on Large Model-Based Business Support Practices for Nuclear Power Production Documentation

Hu Dandan, Chen Ying, Yang Yiqi
Fujian Fuqing Nuclear Power Co., Ltd., Fuzhou, Fujian 350300

Abstract: To solve the problems of single retrieval mode, insufficient semantic interaction, lack of automatic early warning for document revision and fragmented domain knowledge existing in traditional nuclear power document management system, this paper develops customized vertical large model for nuclear power based on domestic large model technologies. With multi-modal RAG, multi-level chain-of-thought knowledge graph and hierarchical semantic parsing adopted, the overall framework, data structure and seven core functional modules of the auxiliary system are designed. The system realizes data interconnection with ECM and ITASK, and constructs domain knowledge base with more than 30,000 maintenance specifications and operation manuals for model local optimization. Trial operation verifies that the system greatly improves document access efficiency and automatically alerts version changes to avoid safety risks caused by expired documents. This paper analyzes existing technical defects such as low recognition accuracy of non-standard drawings, insufficient rare fault samples and delayed real-time update of knowledge graph, and proposes future research directions. The study realizes the practical engineering application of domestic large model in nuclear document management, forms a replicable implementation paradigm, and provides references for knowledge governance and digital transformation of smart nuclear power industry.

Keywords: smart nuclear power; domain large model; knowledge graph; multi-modal RAG; production document management

引言

（一）研究背景与问题

通用大模型在自然语言理解与智能推理领域发展迅速，在多个民用行业已经形成规模化落地应用 [1][2]，但核电行业受安全管控严格、专业门槛高、业务场景专用性强等因素制约，大模型落地进程相对缓慢，生产文件智能化管理水平偏低 。福清核电长期依托 ECM 文档系统存储各类维修规程、设备操作手册等生产资料，存量文档体量庞大，但原有系统功能偏重于文档归档存储，智能化能力不足，实际使用中暴露出多项痛点：一是仅支持关键词精准匹配检索，无法基于语义进行关联查询，检索效率与精准度较差；二是不具备自然语言问答交互能力，工作人员难以通过口语提问快速提取文档关键知识；三是文件升版、作废等变更信息缺少自动化推送机制，现场作业易

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

误用过期文件，存在安全风险；四是文档之间的引用、关联关系无法自动挖掘，知识零散割裂，难以形成体系化知识资产；五是缺少文档智能精读辅助功能，长篇规程研读耗时久，文档利用效率受限。

（二）研究目标

为解决上述文档管理痛点，依托现有核电文档资源与国产化大模型技术积累，本文开展面向核电生产文件的垂类大模型研发，搭建配套智能业务辅助系统。研究目标包含：基于存量生产文档完成核电领域私有化模型训练，搭建专属知识库；构建语义化检索与智能问答服务，提升资料获取效率；依靠智能关联与消息提醒实现文件全生命周期管控，规避过期规程使用隐患；总结核电大模型落地技术路径，为同行业智能化改造提供参考依据。

（三）论文结构

本文遵循 IMRaD 学术写作框架，第2部分从可行性、系统架构、数据架构、关键技术、系统集成部署展开系统设计论述；第3部分介绍系统落地成果、功能应用与效益测算；第4 部分分析现存技术局限并提出后续优化方向，最后总结全文研究结论。

一、研究方法与系统设计

（一）项目可行性分析

从技术、资源、落地条件三方面完成可行性论证。在技术层面，国内国产化大模型产业链逐步完善，中核体系已有多款核电自研垂类大模型落地案例，相关 RAG、知识图谱工程化技术成熟，能够支撑本系统研发；在资源层面，ECM 系统沉淀海量高质量核电生产文档，可满足大模型训练与知识库建设的数据需求，参研团队拥有多年核电信息化建设经验，人才储备充足；在落地条件层面，核电现有信息化底座完善，具备系统对接、内网部署的软硬件基础，可支撑平台平稳落地。

（二）系统架构设计

1. 总体架构

系统采用分层模块化设计思路[10]，自下而上分为基础设施层、模型服务层、业务服务层、应用服务层，架构设计兼顾国产化自主可控、横向扩展与数据安全，适配核电内网涉密部署环境。基础设施层提供硬件算力与底层系统支撑；模型服务层封装大模型、向量检索、知识图谱等 AI 基础能力；业务服务层实现各类智能化业务逻辑；应用服务层面向终端用户提供可视化操作入口。

2. 功能架构

系统划分七大核心功能模块，模块间数据互通、能力联动：

（1）大模型数据预处理和训练模块：搭建自动化文档处理链路，实现多格式文档解析、元数据抽取、文件向量化处理，支撑存量文档批量治理与新增文档增量入库、领域模型微调；

（2）智能搜索模块：结合传统数据库查询与语义向量检索技术[5]构建多层次的混合检索架构，确保搜索结果的全面性与精准度。

（3）智能推荐模块：结合岗位属性、用户浏览行为与文档语义相似度搭建混合推荐引擎，实现资料个性化推送；

（4）智能问答模块：构建基于 RAG 的知识问答功能[3]，为用户提供精准、可追溯的问答服务。

（5）智能关联模块：以领域知识图谱为基础，自动挖掘文档引用、版本依赖等内在关联，实现知识链路可视化查询；

（6）智能理解模块：依托分层语义解析算法，自动对长篇规程生成分级摘要与标准化大纲，简化文档阅读成本；

（7）智能提醒模块：基于事件驱动规则引擎，实时抓取文件变更数据，定向推送升版、作废预警信息。

3. 数据架构

按照数据全生命周期流转逻辑，数据架构分为六层：数据生产层汇聚 ECM 系统存量文件、部门云盘资料、业务系统台账等多源异构数据；数据传输层依靠接口与消息组件完成跨系统数据交互；数据存储层采用混合存储方案，结构化元数据存入 PolarDB，全文检索依托 Elasticsearch，向量数据、图谱数据、原始文件分别存储于向量库、图数据库与对象存储；数据加工层集成 OCR、多模态 RAG、核电专用大模型，完成数据清洗、结构化与知识抽取；数据应用层封装标准化服务接口；应用场景层面向检修、文档管理等岗位落地智能化应用。

（三）关键技术

1. 面向生产文档的知识图谱构建技术

以全量核电运维与操作文档为数据源，结合多级思维链推理与核电知识工程方法[6]，按照实体识别、关系挖掘、属性补全三级抽取策略完成知识结构化，搭建适配核电维修业务的专属知识图谱，为系统智能问答、知识关联、版本溯源提供底层知识支撑。

2. 多模态 RAG 向量库构建与检索技术

针对核电文档包含 PDF、扫描件、图表等多模态文件的特点，搭建统一向量空间的多模态 RAG 架构[7][8]，实现图文异构数据统一向量化存储与一体化检索，弥补传统纯文本 RAG 无法兼容非标图纸、老旧扫描文档的缺陷。

3. 多级语义解析文档智能处理技术

构建篇章 – 段落 – 句子三级递进语义解析模型[9]，依托大模型语义理解能力实现规程文档自动摘要、大纲生成，提升厚重技术文档的阅读与知识提炼效率。

（四）系统集成与国产化部署

系统采用 REST API + 消息中间件完成和 ECM、ITASK 等业务系统对接，实时同步文档变更与业务数据。软硬件部署落实自主可控要求，硬件选用国产算力服务器，操作系统、数据库等基础软件选用国产化产品，整套系统部署在核电内部涉密局域

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

网，满足行业信息安全管理规范。

二、系统落地结果与效益评估

（一）项目落地成果

研究形成两项核心成果：一是一套全国产化核电生产文件智能辅助系统软件，包含七大功能模块前后端完整程序；二是集成多模态 RAG、混合检索、知识抽取能力的核电领域智能底层引擎，基于本地3万余份生产文档完成领域模型训练优化。项目已完成超3万份维修规程、操作手册入库，建成核电维修专属知识库。

（二）功能应用情况

系统在福清核电内网上线试运行，七大模块全部投入实际业务使用，各模块功能及应用场景如表1所示。试运行期间系统运行稳定，各项功能匹配现场使用需求。

表 1 功能模块实现情况

功能模块	核心能力	应用场景
大模型数据预处理和训练	多格式文档自动解析、元数据批量抽取、表格结构化、文档增量向量化、领域模型增量微调	ECM 存量文档批量治理、新上传文档自动化入库处理
智能搜索	关键词 + 语义 + 图谱混合检索、多格式跨文档检索、结果智能重排序	运维人员现场快速查阅规程、故障排查查找技术依据
智能问答	自然语言提问、答案溯源、关联问题推荐、规程条款精准解答	新员工规程学习、检修操作要点快速咨询
智能推荐	岗位画像推荐、浏览关联推荐、检索结果配套推荐	日常办公主动获取同类参考资料
智能关联	知识图谱可视化、三级关联穿透查询、文档引用链路追溯	梳理规程引用关系、排查文件版本依赖
智能理解	长文档分层摘要、标准化自动大纲生成	快速通读厚重规程、提炼文件核心内容
智能提醒	文件升版 / 作废全链路联动提醒、订阅领域变更推送	规避现场误用过期作废规程

实测数据显示：人工检索单份资料平均耗时 5~10min 且需要人工筛选内容，系统检索平均耗时缩短至 10s 以内，文档获取效率提升 90% 以上；依靠变更自动提醒功能，从管理源头规避误用过期文件带来的安全隐患。

（三）效益分析

1. 经济效益

系统大幅压缩运维人员资料查阅工时，节约人力成本；借助文件变更预警，减少因规程版本错误引发的检修返工、设备整改等额外支出；依托知识关联与智能推荐实现历史维修经验复用，有效降低同类故障检修成本。

2. 社会效益

本研究实现国产大模型在核电生产文档场景落地，为国内同类型核电企业智能化建设提供实践范本；推动核电知识管理由传统数字化归档转向智能化知识应用，加速行业数字化升级；通过提质增效助力核电清洁能源稳定供应，支撑国家双碳战略落地。

三、讨论与结论

（一）现存技术局限性

受历史文档质量与项目数据样本约束，系统仍存在三点不足：老旧手绘图纸、模糊扫描非标文档的多模态识别精度不足，非标图表向量化效果有待优化；训练数据以常规运维规程为主，特殊试验、罕见故障案例样本偏少，冷门业务问答准确率仍有提升空间；当前知识图谱以阶段性批量更新为主，全自动化实时实体抽取机制仍需持续迭代。

（二）结论

本文紧扣智慧核电数字化建设发展趋势，针对传统核电文档管理系统现存短板，依托国产化大模型技术，融合多模态 RAG、多级思维链知识图谱、分层语义解析关键技术，完成核电生产文件智能辅助系统的架构设计与功能开发，实现系统与 ECM、ITASK 系统对接并完成三万余份生产文档知识库建设。试运行结果证明，系统有效提升文档检索效率，实现文件版本变更自动化预警，规避安全生产风险，经济与社会效益显著。

本研究完成国产垂类大模型在核电生产文档领域的工程化落地，形成可复制的核电行业大模型落地范式，填补了国内核电生产文档智能化大模型应用的落地空白，为智慧核电知识治理与行业数字化转型提供技术与工程参考。

参考文献

[1] 刘知远，孙茂松，林衍凯，等 . 大模型技术的行业应用与挑战 [J]. 中国科学：信息科学，2024，54(1)：1-28.

[2]Lewis P，Perez E，Piktus A，et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Advances in Neural Information Processing Systems，2020，33：9459-9474.

[3] 王昊奋，漆桂林，陈华钧 . 知识图谱：方法、实践与应用 [M]. 北京：电子工业出版社，2019.

[4] 中国信息通信研究院 . 大模型技术及应用发展报告 (2024)[R]. 北京：中国信通院，2024.

[5]Johnson J，Douze M，J é gou H. Billion-scale similarity search with GPUs[J]. IEEE Transactions on Big Data，2021，7(3)：535-547.

[6]Chen W，Yang Z，Xue L，et al. Multimodal retrieval-augmented generation：a survey[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing，2024：4562-4580.

[7] 李航 . 统计学习方法 (第3版)[M]. 北京：清华大学出版社，2024.

[8]Zhang T，Kishore V，Wu F，et al. BERTScore：evaluating text generation with BERT[C]//International Conference on Learning Representations，2020.

[9] 中国电子技术标准化研究院 . 人工智能标准化白皮书 (2023版)[R]. 北京：中国电子技术标准化研究院，2023.

[10]Johnson J，Douze M，J é gou H. Billion-scale similarity search with GPUs[J]. IEEE Transactions on Big Data，2021，7(3)：535-547.

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.