DOI: 10.19666/j.rlfd.202510063

基于知识图谱和检索增强生成的电站设备故障知识问答

成泽林，徐靖权，陈东超，邓杰文，司和勇
（东北电力大学能源与动力工程学院，吉林 吉林 132012）

［摘 要］【目的】为了解决电站设备故障知识管理中存在的知识碎片化、语义不一致及检索效率低下等问题，提出一种融合知识图谱与检索增强生成技术的知识智能问答方法，以实现对故障知识的有效整合、语义统一与高效精准查询。【方法】该方法基于失效模式及影响分析表，通过本体建模对故障领域知识进行结构化语义表达，将非结构化文本转化为符合本体约束的结构化三元组，进而构建出可以进行语义推理的故障知识图谱，并同步生成图节点与文本段的向量化索引，形成“事实关系+语义嵌入”的双重知识基底。在问答阶段，通过执行图谱关系检索与向量语义检索，将双通道结果融合为大语言模型的增强上下文，从而生成更加科学的问答结果。【结果】在 20 个涵盖多类问答场景的典型问题上，对比了纯大语言模型、整表文本注入、向量检索增强以及所提出方法的问答结果。实验结果表明，所提出的方法在信息覆盖率和语义适配度上均表现最优，综合得分达 91.2，性能优于对比方法。【结论】所提出的知识问答方法有效融合了知识图谱的精确关系推理与向量检索的语义泛化能力，从而在保证较低响应时延的同时，实现了准确性、覆盖度与可解释性的全方位提升，为电站设备故障知识管理与应用提供了有效的解决方案。

［关 键 词］电站设备；故障知识问答；知识图谱；检索增强生成；大语言模型

［引用本文格式］成泽林，徐靖权，陈东超，等. 基于知识图谱和检索增强生成的电站设备故障知识问答[J]. 热力发电, 2026, 55(7): 158. Cheng Z L, Xu J Q, Chen D C, et al. Fault knowledge question answering for power plant equipment based on knowledge graph and retrieval-augmented generation[J]. Thermal Power Generation, 2026, 55(7): 158.

Fault knowledge question answering for power plant equipment based on knowledge graph and retrieval-augmented generation

Cheng Zelin, Xu Jingquan, Chen Dongchao, Deng Jiewen, Si Heyong
（School of Energy and Power Engineering, Northeast Electric Power University, Jilin 132012, China)

Abstract: [Objective] Knowledge pertaining to power station equipment failures is typically dispersed across various reports, maintenance records, and expert experience. This fragmentation, coupled with inconsistencies in semantic description, severely hinders effective knowledge sharing and reuse. Conventional retrieval methods, which rely primarily on keyword matching, often fail to capture deeper semantic relationships, leading to inefficient queries, so knowledge graph and retrieval-augmented generation (KG-RAG) techniques are proposed. The method aims to consolidate fault knowledge, unify its semantics, and enable efficient yet precise querying, with the goal of enhancing the intelligence of power station operation and maintenance. [Methods] Using the failure mode and effects analysis tables as the core knowledge source, a formal domain ontology was first constructed through modeling. This ontology structurally represents key entities, such as equipment, failure modes, symptoms, and causes, along with their attributes and relationships. Based on this framework, unstructured text was converted into structured triples that comply with ontological constraints, progressively building a fault knowledge graph capable

收 稿 日 期：2025-10-28 修 回 日 期：2025-12-16 接 受 日 期：2025-12-19
基 金 项 目：吉林省科技发展计划项目（20230101339JC）
Supported by: Scientific and Technological Development Program of Jilin Province (20230101339JC)
第一作者简介：成泽林（2000），男，硕士研究生，主要研究方向为基于知识驱动的旋转机械振动故障诊断，czl18035456717@163.com。
通信作者简介：陈东超（1985），男，博士，副教授，主要研究方向为电站设备智能运维，cdc@neepu.edu.cn。

of semantic reasoning. Simultaneously, embedding models were employed to generate vector representations for nodes in the graph and their associated textual descriptions. A dual knowledge base was thus established, indexed for both graph structure and semantic vectors, effectively combining “factual relationships” with “semantic embeddings”. During the question-answering phase, structured retrieval is first performed using the knowledge graph to accurately locate relevant equipment and fault chains. If the graph-based results are insufficient, a semantic retrieval channel is activated, leveraging vector similarity to recall related text passages. The outputs from both retrieval channels are then deduplicated, ranked, and fused into a coherent, information-rich context. This enhanced context is fed into a large language model, which reasons over it to generate final answers. The responses not only present factual information but also incorporate explanations and relevant suggestions, yielding scientifically grounded and comprehensive results. [Results] To evaluate the proposed method, an experiment was conducted using 20 typical fault-related questions spanning multiple scenarios, including factual queries, cause inference, and handling recommendations. The performance of four approaches was compared: a standalone large language model, full-table text injection, vector-only retrieval augmentation, and the KG-RAG method. The evaluation metrics covered information coverage, semantic relevance, and response time. Experimental results indicated that the proposed method outperformed the others in both information coverage and semantic relevance, achieving an overall score of 91.2. This demonstrates that the hybrid mechanism, which integrates knowledge graph retrieval with semantic vector search, provides more comprehensive and accurate support for knowledge-intensive question-answering tasks. [Conclusion] The KG-RAG-based question-answering method presented in this study effectively combines the precise relational reasoning of knowledge graphs with the semantic generalization capability of vector retrieval. Low response latency is maintained while accuracy, coverage, and explainability in the generated answers are improved. A practical and effective solution for managing and applying fault knowledge in power station contexts is thus offered. Future work could focus on dynamic knowledge updating, multi-modal fault knowledge integration, and cross-domain knowledge transfer, all of which would further enhance the method’s practicality and generalizability.

Key words: power-plant equipment; fault-knowledge question answering; knowledge graph; retrieval-augmented generation; large language model

电站汽轮机、发电机、水泵、风机等关键设备普遍具备结构复杂、运行工况多变、故障类型多样等特点[1]。长期的运行与维护实践过程中，积累了大量与设备故障相关的领域知识，这些知识蕴含着丰富的故障因果逻辑与处置经验，是实现设备故障诊断与健康管理的重要基础[2]。目前，故障知识依附于专家长期积累的经验与口头传承，或者蕴含在运维技术手册、检修报告、故障案例中，缺乏统一的组织与语义规范，呈现出高度分散与异构化的特征[3-4]。不同来源的知识在术语表达、因果逻辑、结构形式等方面存在差异，导致这些知识难以被有效整合与自动化利用[5]，成为缺乏统一语义与结构化表示的知识体系，不足以支撑结构复杂的电站设备的智能故障诊断[6]。

失效模式及影响分析（failure modes and effects analysis，FMEA）广泛应用于复杂设备故障分析与知识组织[7-8]。王仲等[9]针对燃气轮机健康维护中知识分散、可移植性差、推理效率低的问题，结合 FMEA 与故障树分析提取设备健康知识，提出了一种“知识获取−表达−推理”框架。杨立飞等[10]通过对多个系统和设备的故障模式、后果及维修策略进行系统性分析，建立了关键设备 FMEA 数据库，实现了对潜在故障模式的识别与维护策略优化。但是传统 FMEA 存在排序不稳、信息丢失等缺陷，并且 FMEA 分析获得的故障知识以表单等非结构化形式存储，不利于计算机识别和利用[11]。

为了高效表达电站设备故障知识，提高知识的查询与分析能力，提出了很多知识表达方法，常见的方法包括本体理论[12-13]、知识图谱（knowledge graph，KG）[14-16]等。张彬桥等[17]针对水电站运维知识难以有效管理与复用问题，基于本体理论建立了水电站运维知识库，并设计了一套本体相似度计算算法。Li 等人[18]将 KG 与改进的 FMEA 相结合，引入模糊集理论与证据融合方法，处理专家评估中的不确定性问题。Bahr 等人[19]将 KG 和语义检索增强生成（retrieval-augmented generation，RAG）应用于 FMEA 数据，实现了基于集合论标准化的知识映射与嵌入，显著提升了系统的可解释性与检索精度。谢淙恺[20]针对传统知识驱动诊断在知识管理和交互方面的不足，将专家经验知识与机器可理解的图结构融合，构建了基于 KG 的汽轮机故障诊断问答系统[21]。

随着人工智能的快速发展，大语言模型（large language model，LLM）得到了广泛应用，其在自然

语言理解与知识分析领域取得了显著突破[22]。Karan 等人[23]通过多数据集融合与指令调优，使模型在医学问答基准上达到专家级水平，验证了 LLM 在领域知识理解与推理方面的潜力。张鹤译等[24]提出了一种将 LLM 与 KG 深度融合的中医药领域知识问答系统框架。尽管 LLM 引起了广泛的关注，但它仍然面临着诸多挑战，包括模型内部的可控性、解释性、语言数据质量的不确定性以及产生幻觉的潜在风险[25]。

综上，为了解决电站设备故障知识分散异构、术语不一致等问题，并提升知识检索精度与答案可信度，同时降低生成幻觉、支持多源知识融合，本文利用本体与 KG 技术构建统一语义体系与逻辑约束，建立“结构化事实+向量化索引”融合的数据底座，并引入检索增强生成机制以实现可控检索与可解释生成，结合 LLM 的语义理解与推理能力，最终形成可追溯、可交互的故障知识问答流程。

1 基于 KG-RAG 的故障知识问答

1.1 总体流程

基于 KG-RAG 的电站设备故障知识智能问答系统总体流程如图 1 所示。首先，通过构建 FMEA 表格、建立领域本体、完成 KG 构建与语义向量嵌入，为后续问答提供结构化和语义化的知识支撑。其次，依托 Cypher 查询与检索增强生成的双通道检索机制，结合大语言模型在语义理解、上下文记忆与推理方面的能力，实现对问题的多轮连续问答。在交互过程中，模型能够自动关联历史对话内容，进行上下文整合与总结反馈，从而完成对复杂故障知识的语义解析、动态检索以及智能推理。

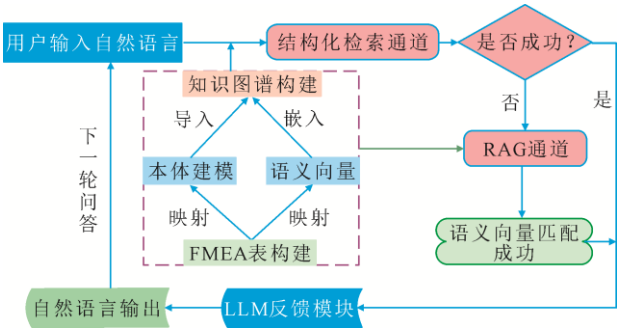


图 1 基于 KG-RAG 的智能问答总体流程

Fig.1 Overall workflow of the KG-RAG-based intelligent question-answering

1.2 FMEA 表构建

对电站设备中采集的故障信息进行清洗与标准化处理，进而生成了链式结构记录。每条记录均包含设备层次、故障模式及其对应的征兆、原因和影响等信息，构成了故障因果链，为后续的本体建模与故障知识三元组提取提供了基础数据，FMEA 表的建立如图 2 所示。

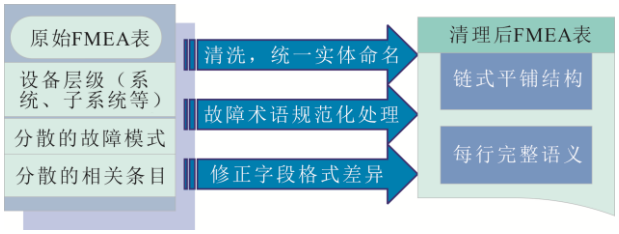


图 2 FMEA 表的建立

Fig.2 Construction of the FMEA table

1.3 本体建模

本体基于描述逻辑和集合论，为领域概念和它们之间的关系提供了严谨的数学基础及逻辑约束，可将异构、分散的 FMEA 信息统一到同一个语义框架下，消除术语混杂和结构差异。借助本体公理和推理机制，可实现故障知识自动一致性验证、隐含知识推导和复杂因果链的查询优化，保障后续 KG 构建与智能问答的准确性与可维护性，本体的语义和语法如表 1 所示。本体模型的建模过程包括概念建模、属性定义和关系设计及个体填充等，建模流程如图 3 所示。

表 1 本体的语义和语法

Tab.1 Semantics and syntax of the ontology

语法	中文释义	语义
Individual	个体	类的实例，个体之间利用属性来连接
Class	类	具有相同属性个体的集合
Property	属性	类与个体、个体与数据之间的关联关系
Domain	定义域	声明具有某种属性的类
Range	值域	声明属性作用对象的类
Annotation	注释	为本体、公理及个体添加补充说明

根据电站设备知识问答的需要，构建了包含设备结构、故障模式、故障征兆、故障原因、故障影响和运维措施在内的电站设备故障知识本体类，如图 4 所示。

设备结构反映了故障发生位于的实体，也是维护工作的对象单元，设备结构类有子类“系统”和“设备”。故障模式一般处于故障传递链的顶端（顶事件），是系统层面不希望发生的事件。复杂系统内的故障具有传递性的特征，部件上发生的故障原因可以进一步造成系统层面故障模式的发生，故障原

因可能涉及设计和制造、安装和维修、运行和调整等方面。故障征兆是故障发生后的表现特征，也是诊断的依据。由于故障的复杂性，征兆具有表现形式和检测手段多样的特点。如果要对故障原因进行识别，需要进一步寻找故障原因的征兆。故障影响是故障发生的后果，涉及运行经济性、设备和人员安全和环保等多个方面，而运维措施是为了防止故障发生或者消除故障所采取的各类运行和维护手段。根据各类故障知识间的关系，给出了属性关系的名称、类型、定义域和值域，具体见表 2。

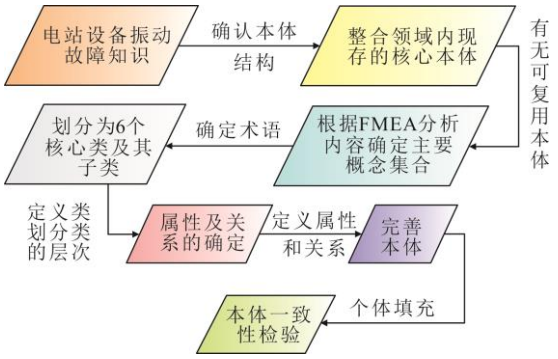


图 3 本体的建模流程

Fig.3 Ontology modeling process

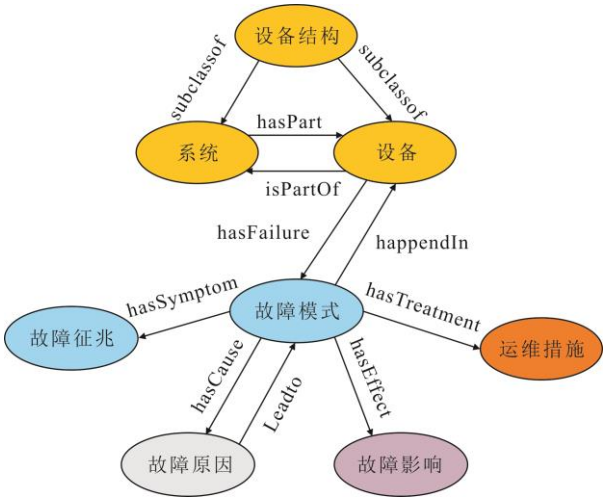


图 4 本体类和属性关系

Fig.4 Ontology classes and property relations

Protégé作为本体构建的开源工具，它支持多种类型的语义声明，包括类层级、公理约束、对象属性链等复杂逻辑结构。Protégé 提供的 Cellfie 插件支持通过 MappingMaster 语言，将 Excel 表格数据按指定映射规则自动转换为本体结构，实现批量实例的快速构建。利用该工具可便捷地在 Protégé 中导入大量故障知识个体，同时支持属性、注释及限制条件的导入，极大提升了本体实例化效率。

表 2 故障知识本体中属性声明

Tab.2 Attribute declaration of fault knowledge ontology

名称	注释	定义域	值域
isPartOf	是部分	设备	系统
hasPart	包含	系统	设备
happendIn	发生在	故障模式	设备
hasCause	原因是	故障模式	故障原因
hasEffect	影响是	故障模式	故障影响
hasTreatment	措施是	故障原因	运维措施
hasSymptom	征兆是	故障模式	故障征兆

1.4 基于 KG-RAG 的知识问答

通过 KG 将故障知识要素组织成可查询的语义网络，并作为 RAG 框架中的外部知识底座。故障知识问答时，首先利用 KG 进行结构化检索，优先从图谱中精确定位相关设备和故障链路；当结构化结果不足以回答问题时，再通过嵌入向量进行语义检索，将与问题相关的故障节点及其上下文召回，由 LLM 在这些检索证据的基础上完成答案生成。

在上述过程中，以故障的 FMEA 表为基础，利用本体统一领域知识表达，通过 KG 实现因果关系的语义关联，并结合 LLM 的语言理解与推理能力，实现知识结构化表达和智能问答。

KG 具体构建流程如图 5 所示，包括以下几个步骤：首先，读取并预处理经标准化整理的 FMEA 表，对字段进行清洗、规范和类型转换，使其与本体中定义的类与对象属性一一对应。随后，将 FMEA 记录转换为 RDF 三元组，完整刻画“设备–故障模式–原因/征兆/影响/措施”之间的因果与隶属关系。生成的三元组被序列化为 OWL/RDF 文件，通过 Neo4j 的 n10s 插件导入图数据库，由插件自动识别类、实例与关系，并物化为节点-边的结构，以保证后续故障知识查询的一致性与高效性。

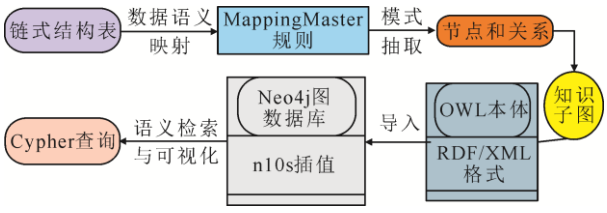


图 5 知识图谱构建

Fig.5 Construction of knowledge graph

在完成图谱结构构建后，为支持基于 RAG 的智能问答，需进一步为图谱节点生成向量表示并与 LLM 融合。具体步骤包括：首先，为每个故障模式节点构建一段概括性描述文本，将系统/设备名称、典型征兆、主要原因、影响与处置措施等字段拼接

为语义连贯的说明语句；随后，调用 OpenAI 文本嵌入模型，对这些描述文本进行批量编码，得到语义向量，并将其作为节点的 embedding 属性写回 Neo4j 数据库，从而实现图谱节点在“结构关系+向量表示”层面的知识统一。

在此基础上，将所构建的 KG 嵌入 RAG 问答流程，形成“结构优先、语义补充、双轨结合”的 LLM 融合应用框架：一方面，系统首先尝试将用户自然语言问题解析为 Cypher 查询，通过 KG 精准定位相关设备、故障模式及其因果链路，并将返回的局部子图（包括节点属性与关系路径）作为结构化证据输入 LLM，由其整合生成基于事实的答案；另一方面，若结构化查询未能检索到足够信息或解析失败，系统则转入基于向量语义的检索路径，即对用户问题进行嵌入编码，计算其与图谱中故障节点嵌入向量的语义相似度，召回相关节点及其邻域子图，再交由 LLM 在检索证据的约束下生成回答。通过上述机制，LLM 不再仅依赖自身参数中的隐式知识，但能够动态调用领域知识库，从而在电站故障问答场景中实现更准确、可解释的知识驱动型智能交互。

2 案例分析

2.1 数据集

本文基于给水泵的故障知识分析，整理并构建了 100 多张 FMEA 表，涉及主泵、前置泵、电机、液力耦合器等设备的多种故障模式，表 3 给出了 FMEA 表示例。以此为基础，开展了故障知识的智能问答测试，验证所提出方法的有效性。

表 3 FMEA 表示例
Tab.3 FMEA table example

系统	设备	故障模式	故障原因	故障征兆	故障影响	运维措施
给水泵	液力耦合器	滑动轴承烧瓦	油量不足 安装工艺不良	振动增大 一倍频 油温升高	造成停泵 威胁机组安全	对滑动轴承进行检修
给水泵	电动机	转子不平衡	平衡校正不当 转子变形 部件安装不良 部件松动或脱落	振动大 一倍频	导致轴承受力情况恶化 会造成停车，威胁机组安全	动平衡

2.2 案例展示

在完成基于 Protégé 的 OWL 本体设计与验证后，将其导出为 Turtle 格式文件，该文件包含了经逻辑验证的完整本体三元组集合。随后，在保持与 Protégé 中 Cellfie 插件（MappingMaster 规则）所定义映射规则一致的前提下，开发了 Python 转换脚本，实现了从原始 FMEA 数据到 KG 与语义向量索引的全流程自动化构建。整体流程可分为以下 4 个阶段。

1）数据预处理，对原始 FMEA 表格数据进行系统化清洗，包括去重、缺失值填充及格式统一等操作，形成完整、一致的标准化数据集，为后续实例化提供可靠输入。

2）基于映射规则集合 $\Phi=\{\phi_k\}$，将每条清洗记录 $r_i \in X_{\text{prep}}$ 转化为本体三元组，为确保映射过程语义一致，在构建三元组时引入了本体的概念层、关系属性层与实例层三部分，如图 6 所示。

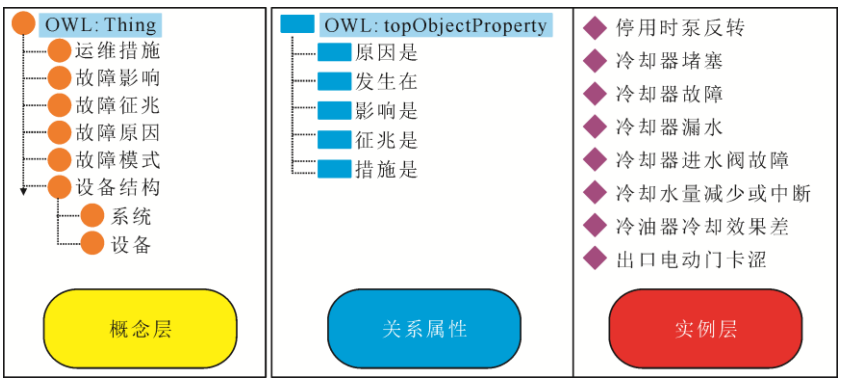


图 6 本体结构示意
Fig.6 Schematic diagram of the ontology structure

3）为支持后续的语义检索与相似匹配，在生成三元组后，为每个故障模式节点预计算语义向量表示。具体流程如下：

①生成描述文本，自动组织设备层次、典型征兆、可能原因、影响及运维建议等信息，形成连贯的语义上下文；

②调用 OpenAI 文本嵌入模型，将描述文本编码为高维浮点向量，以捕获节点语义；

③通过 Neo4j 驱动接口，将向量作为属性写入对应节点。遍历所有故障模式节点后，即构建起完整的语义索引底座，实现从“结构化知识”到“语义化表示”的拓展，为语义检索、相似匹配与问答生成提供支撑。

4）将映射所得的三元组、原始本体三元组及新增的语义向量属性合并，通过 Neo4j 批量导入，构建最终的 KG。该图谱不仅完整覆盖“设备−故障−原因−征兆−影响”的因果链，还附有高维语义嵌入，实现了结构关系与语义特征的统一表达。

在 Neo4j 中，设备（如“主给水泵”）可通过结构部件关联到一系列典型故障，这些故障进一步表现为多种可观测征兆。借助语义关系的可视化分析，可识别“主给水泵”引发的征兆分布。该故障 KG（fault KG，FKG）如图 7 所示，该图谱不仅清晰呈现故障传播路径，还显示多个征兆节点可能对应多个潜在故障，提示诊断中需关注“多因多果”的交叉关联现象。

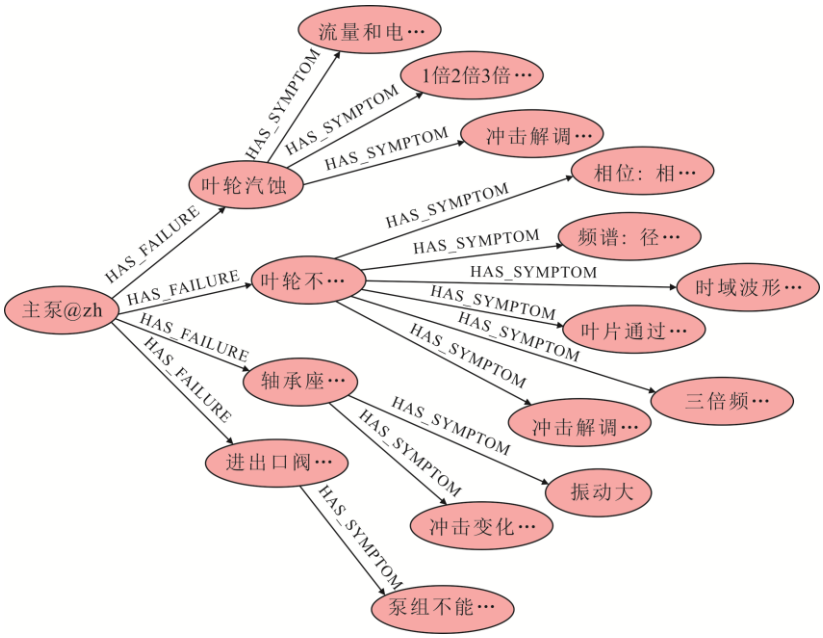


图 7 故障知识图谱示意

Fig.7 Schematic diagram of the fault knowledge graph

随后，系统基于用户输入检索语义相似的故障模式，生成包含多条可能原因与处理建议的候选问答结果，并通过多轮对话形式进行展示。具体流程如图 8 与图 9 所示。



图 8 首轮问答结果

Fig.8 First-round question-answering result



图 9 第二轮问答结果

Fig.9 Second-round question-answering result

图 8 展示了首轮交互中，系统对用户问题进行的语义检索与信息整合过程，LLM 基于检索结果生成了初步的问答总结并反馈给用户；图 9 则展示了在后续对话轮次中，系统进一步响应用户追问的界面：左侧区域呈现基于上一轮结果更新后的语义检索与判断反馈，右侧区域则输出结构化的 AI 问答结果，包括候选故障模式、可能原因及对应处理建议。该系统以多轮、递进式的对话输出多条推理结论，从而辅助工程人员进行综合判断与决策。

3 结果分析和讨论

3.1 对比方案

为了验证不同技术路线在复杂工程场景下的适用性与推理深度，共选取 20 个典型问题场景，从响应时延、信息覆盖率与语义适配度三方面进行量化评价，设计并对比了 4 种故障知识问答方案：

1）纯大模型（LLM） 直接将用户提问提交给大模型，无任何领域知识注入。

2）整表上下文注入（FMEA-LLM） 完整 FMEA 表格拼接为一段长文本，作为模型上下文一次性输入。

3）向量检索增强（FMEA-RAG） 先将用户问题与 FMEA 文档中的各条记录做向量匹配，筛选 Top-K 最相似段落，再将这些段落作为上下文注入大模型生成答案。

4）KG-RAG 融合 KG 的结构化检索与向量检索双通道。

3.2 测试集与评价指标

围绕故障模式，模仿真实用户口吻的模糊/并列/嵌套/错别字问法提问，检验在非标准输入下仍能覆盖关键因果要素。

主要评价指标包括：

1）响应时延（s） 从提问到模型返回答案的平均耗时，引入戒上型隶属度函数来呈现响应时延和得分的关系，如式(1)所示。

$$\mu(x)=\begin{cases}100, & x\leqslant \mathrm{c}\\ \dfrac{100}{1+(a(x-c))^{b}}, & x>\mathrm{c}\end{cases} \tag{1}$$

式中：x 为响应时延；c 为时延阈值；a 为衰减速度的尺度系数；b 为控制曲线形状的指数系数。

2）信息覆盖度 用于衡量模型生成的回答中是否完整覆盖用户问题所涉及的关键内容要素。基于 KG 构建了标准的链式因果结构，包括以下 5 类核心要素：①故障，问题的核心现象或对象；②原因，导致故障发生的诱因；③征兆，故障可能表现出的异常特征；④影响，故障可能引发的后果；⑤措施，可采取的应对策略或维修建议。

3）语义适配度 用于衡量回答在表达结构与内容组织上的清晰度与专业度，主要包括以下要点：①回答是否围绕问题展开，避免与问题无关的陈述；②回答内容是否具备清晰的因果关系链条，体现因果递进（如“由于……导致……从而……”）；③信息组织是否条理清晰，段落划分合理，避免堆砌或语义重复；④回答信息量是否适中：通常 3~4 条具体内容为佳，若回答中存在超过 4 条信息却无逻辑结构或缺乏依据，视为结构性混乱。

在评分过程中，采用 Claude 模型对每个标准要素进行判断：识别回答中是否包含与该要素语义等价或高度相近的表述。模型运作的原理可形式化表示为式(2)与式(3)。

覆盖度 “被覆盖要素数/标准要素总数”，这对应信息检索中的召回率。

$$C_{\text{coverage}}=\frac{1}{|S|}\sum_{i=1}^{N} f_{\theta}(e_i,A) \tag{2}$$

式中：$|S|$表示标准要素的总数；A 表示回答中抽取出的信息集合；e_i 表示第 i 个标准要素，判定函数 $f_\theta(e_i, A)$表示“e_i 是否在 A 中语义覆盖”。

适配度　“有效要点数/回答总要点数”来刻画适配度，这对应检索中的精确率。

$$F_{\text{fitness}} = \frac{1}{|A|}\sum_{j=1}^{M} g_\theta(a_j, S) \tag{3}$$

式中：a_j 表示第 j 个回答要素，判定函数 $g_\theta(a_j, S)$表示“a_j 是否与 S 相关”。

具体评分流程如下：1）针对每道评测题目，依据 KG 提取标准原始数据，形成包含故障、原因、征兆、影响与措施等要素的集合，并基于此提炼出信息覆盖度和语义匹配度 2 项评分准则。2）Claude 模型根据预设提示词解析回答中的内容结构、信息链条与表述逻辑，对比标准数据生成结构化的 JSON 评分结果。3）解析评分 JSON，记录模型在该问题下的各项表现指标。4）所有题目评估完成后，汇总数据并输出平均覆盖度与匹配度结果，进而形成模型对比结论。

上述机制不仅涵盖图谱因果逻辑的完整性评估，还兼顾语言表达质量的判断，可为结构化问答模型性能评价提供了可复现、可解释的评估框架。

3.3 测试结果

评价结果见表 4。由表 4 可见，4 种方案在覆盖率与适配度 2 项指标上的对比情况如下：FMEA-LLM 与 FMEA-RAG 的得分相对偏低，LLM 单模型的表现略低于知识增强类方案，KG-RAG 表现最优，其覆盖率和适配度最高，表明基于 KG 的检索能够实现更全面、更精准的知识匹配。

表 4 评价结果
Tab.4 Evaluation result

项目	覆盖度/%	适配度/%	时延/s
LLM	84.1	88.8	2.0
FMEA-RAG	82.3	88.4	7.0
FMEA-LLM	70.5	83.8	15.2
KG-RAG	87.4	94.8	4.8

根据领域知识确定式(1)中的参数 c=2，根据 2 个中心点值反推出 a=0.17，b=3.29，最后得到隶属度图形如图 10 所示。

引入层次分析法（AHP）[26]用以计算 3 个指标的权重，根据领域知识，确定判断因子矩阵为：

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 1/3 & 1/3 & 1 \end{bmatrix} \tag{4}$$

采用几何平均法（方根法）对表 4 的判断矩阵进行计算，并按式(5)对所得权向量进行归一化后，得出 3 个评价指标的权重分别为 0.43、0.43 和 0.14，最终得到每种问答方案的综合加权得分（表 5）。

$$W_i = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}}{\sum_{i=1}^{n}\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}} \quad i = 1,\ 2,\ \ldots,\ n \tag{5}$$

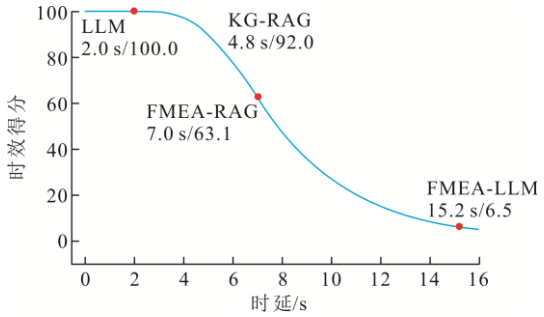


图 10 时延和得分隶属度关系
Fig.10 Relationship between time delay and score membership degree

表 5 最终评价结果
Tab.5 Final evaluation results

项目	时延/s	时效得分	覆盖度得分	适配度得分	综合得分
LLM	2.0	100.0	84.1	88.8	88.3
FMEA-RAG	7.0	63.1	82.3	88.4	82.2
FMEA-LLM	15.2	6.5	70.5	83.8	67.3
KG-RAG	4.8	92.0	87.4	94.8	91.2

从表 5 可见，所提出 KG-RAG 方法综合得分最高（91.2），在覆盖性与适配性方面均显著优于其他方案，能够有效提升多场景下的知识推理准确性与覆盖能力。

4 结　论

本文提出的 KG-RAG 知识问答方法，有效实现了对专业领域故障知识的智能问答支持。通过将 FMEA 知识以本体图谱形式深度融入 RAG 框架，在面对复杂的知识问答与推理问题时，展现出更高的准确性、专业性与实用性。其核心优势主要体现在以下 3 个方面。

1）知识的结构化语义融合　突破传统文本检索模式，通过构建领域本体对 FMEA 知识进行语义化建模，形成结构化的 FKG。这不仅实现了对设备、故障模式及其因果关系的深度理解，更支持沿图谱关系路径进行逻辑推理，使问答过程由简单的信息

匹配升级为基于结构化领域知识的诊断分析。

2）连贯的多轮对话与状态维持 具备对话状态跟踪与上下文记忆能力，能够在多轮交互中持续理解用户意图与历史语境。可以支持追问、澄清和递进式问答，使交互过程更贴合领域专业人员实际排查故障的连贯思维，提升了问答的连续性与场景适应性。

3）双通道检索增强的鲁棒性保障 创新性地融合了基于 FKG 实体的精确检索与基于向量空间的语义检索。双通道机制确保在用户提问表述精确时能直接定位图谱节点获取权威信息，而在提问模糊或表述多样化时，亦能通过语义相似性召回相关背景知识，从而在精确性与泛化能力之间取得平衡，保障了答案的全面性与系统鲁棒性。

与此同时，KG-RAG 框架在实际应用中也存在一些局限。其表现严重依赖于底层 FMEA 数据的完备性及本体建模的质量，若知识覆盖不全或建模存在偏差，可能导致对长尾故障模式的召回不足。此外，其结构化查询方法在跨设备、跨工况迁移时往往缺乏灵活性，常需针对新场景进行人工调整与适配。这些不足凸显了未来在本体动态扩展、增量知识学习以及跨场景自适应迁移等方向进行优化的重要性与改进空间。

[参 考 文 献]

[1] 李丰均, 徐云柯, 彭元苏, 等. 发电设备故障知识图谱可视化构建与智能化应用研究[J]. 电气时代, 2024(8): 57.
Li F J, Xu Y K, Peng Y S, et al. Research on visualization construction and intelligent application of power generation equipment fault knowledge graph[J]. Electrical Age, 2024(8): 57.

[2] 刻昌锋, 张永明, 艾科勇, 等. 基于本体的汽轮机多源异构知识建模与融合[J]. 兰州理工大学学报, 2021, 47(5): 30.
Yan C F, Zhang Y M, Ai K Y, et al. Ontology-based multi-source heterogeneous knowledge modeling and fusion for steam turbines[J]. Journal of Lanzhou University of Technology, 2021, 47(5): 30.

[3] 邢曜宇. 智能电厂建设中设备预测性维护技术的应用与探索[J]. 张江科技评论, 2025(8): 104.
Xing Y Y. Application and exploration of equipment predictive maintenance technology in intelligent power plant construction[J]. Zhangjiang Science Review, 2025(8): 104.

[4] 黄元平, 王仲, 韩旭东, 等. 燃气蒸汽联合循环发电机组健康管理大数据系统[J]. 热力发电, 2020, 49(12): 59.
Huang Y P, Wang Z, Han X D, et al. Big data system for health management of gas-steam combined cycle generating units[J]. Thermal Power Generation, 2020, 49(12): 59.

[5] 蓝梦琦, 黄路遥, 魏宸, 等. 面向发电领域文本数据结构化方法研究与应用[J]. 工业控制计算机, 2024, 37(9): 1.
Lan M Q, Huang L Y, Wei C, et al. Research and application of text data structuring methods for the power generation field[J]. Industrial Control Computer, 2024, 37(9): 1.

[6] 阳曦鹏, 李碧薇, 李德华, 等. 基于FMEA的电力设备故障管理系统设计[J]. 电子产品可靠性与环境试验, 2020, 38(增刊 1): 84.
Yang X P, Li B W, Li D H, et al. Design of power equipment fault management system based on FMEA[J]. Electronic Product Reliability and Environmental Testing, 2020, 38(S1): 84.

[7] Prajapati D R. Implementation of failure mode and effect analysis: a literature review[J]. International Journal of Managment, IT and Engineering, 2012, 2(7): 264.

[8] 房丽萍, 梁万顺, 杨凯. 基于 FMEA 和 FTA 分析的直接空冷机组冷端系统故障诊断研究[J]. 给水排水, 2017, 53(增刊 2): 82.
Fang L P, Liang W S, Yang K. Fault diagnosis study on the cold-end system of direct air-cooled units based on FMEA and FTA analysis[J]. Water & Wastewater Engineering, 2017, 53(S2): 82.

[9] 王仲, 顾煜炯, 韩旭东, 等. 燃气轮机健康维护知识获取与表达推理的框架[J]. 中国机械工程, 2021, 32(2): 235.
Wang Z, Gu Y J, Han X D, et al. Framework for knowledge acquisition, representation, and reasoning in gas turbine health maintenance[J]. China Mechanical Engineering, 2021, 32(2): 235.

[10] 杨立飞, 武涛, 青晨, 等. FMEA 技术在CPR1000机组关键敏感设备维修优化中的应用[J]. 核动力工程, 2019, 40(5): 175.
Yang L F, Wu T, Qing C, et al. Application of FMEA technology in maintenance optimization of key sensitive equipment in CPR1000 units[J]. Nuclear Power Engineering, 2019, 40(5): 175.

[11] Chakhrit A, Benharkat N E H, Guetarni I H M, et al. A literature review of risk assessment approaches in failure mode and effects analysis[J]. International Journal of Quality & Reliability Management, 2025, 42(9): 2415.

[12] 李恒杰, 李军权, 李明. 领域本体建模方法研究[J]. 计算机工程与设计, 2008, 29(2): 381.
Li H J, Li J Q, Li M. Research on domain ontology modeling methods[J]. Computer Engineering and Design, 2008, 29(2): 381.

[13] Ding Y, Zhu F, Li H, et al. A failure knowledge graph learning framework for offshore wind turbines with incomplete knowledge[J]. Renewable and Sustainable Energy Reviews, 2025, 215: 115561.

[14] Klein P, Malburg L, Bergmann R. Combining informed data-driven anomaly detection with knowledge graphs for root cause analysis in predictive maintenance[J]. Engineering Applications of Artificial Intelligence, 2025, 145: 110152.

[15] Du J, Yu A, Sugumaran V, et al. Ontological reasoning for automated tunnel defect diagnosis and root cause identification[J]. Automation in Construction, 2025, 178: 106362.

[16] 唐荻音, 丁奕州, 王轩, 等. 知识图谱技术在预测与健康管理中的应用现状与研究展望[J]. 电光与控制, 2024, 31(2): 1.
Tang D Y, Ding Y Z, Wang X, et al. Application status and research prospect of knowledge graph technology in

prognostics and health manage-ment[J]. Electronics Optics & Control, 2024, 31(2): 1.
[17] 张彬桥, 杨文娟, 葛苏叶, 等. 水电站运维本体知识库构建及应用[J]. 水力发电学报, 2022, 41(10): 86.
Zhang B Q, Yang W J, Ge S Y, et al. Construction and application of an ontological knowledge base for hydropower station operation and mainte-nance[J]. Journal of Hydroelectric Engineering, 2022, 41(10): 86.
[18] Li S, Wang J, Rong J. Combining improved DFMEA with knowledge graph for component risk analysis of complex products[J]. Reliability Engineering and System Safety, 2024, 251: 110340.
[19] Bahr L, Wehner C, Wewerka J, et al. Knowledge graph enhanced retrieval-augmented generation for failure mode and effects analysis[J]. Journal of Industrial Information Integration, 2025, 45: 100807.
[20] 谢淙恺. 基于知识图谱的汽轮机组故障诊断问答系统研发[D]. 杭州: 浙江大学, 2023: 1.
Xie C K. Development of a fault diagnosis question-answering system for turbine units based on knowledge graph[D]. Hangzhou: Zhejiang University, 2023: 1.
[21] Ciyuan P, Feng X, Mehdi N, et al. Knowledge graphs: opportunities and challenges[J]. Artificial Intelligence Review, 2023, 56(11): 31.
[22] 文森, 钱力, 胡懋地, 等. 基于大语言模型的问答技术研究进展综述[J]. 数据分析与知识发现, 2024, 8(6): 16.
Wen S, Qian L, Hu M D, et al. Review of research progress on question-answering technologies based on large language models[J]. Data Analysis and Knowledge Discovery, 2024, 8(6): 16.
[23] Karan S, Shekoofeh A, Tao T, et al. Publisher correction: large language models encode clinical knowledge[J]. Nature, 2023, 620(7973): E19.
[24] 张鹤译, 王鑫, 韩立帆, 等. 大语言模型融合知识图谱的问答系统研究[J]. 计算机科学与探索, 2023, 17(10): 2377.
Zhang H Y, Wang X, Han L F, et al. Research on question-answering systems integrating large language models and knowledge graphs[J]. Computer Science and Exploration, 2023, 17(10): 2377.
[25] 赵俊华, 文福拴, 黄建伟, 等. 基于大语言模型的电力系统通用人工智能展望: 理论与应用[J]. 电力系统自动化, 2024, 48(6): 13.
Zhao J H, Wen F S, Huang J W, et al. Prospects of general artificial intelligence for power systems based on large language models: theory and applications[J]. Automation of Electric Power Systems, 2024, 48(6): 13.
[26] 刘广进, 彭亚雄, 苏莹, 等. 基于 AHP-云模型的隧道断层破碎带施工风险评价[J]. 安全与环境工程, 2023, 30(3): 118.
Liu G J, Peng Y X, Su Y, et al. Construction risk assessment of tunnel fault-fracture zones based on AHP-cloud model[J]. Safety and Environmental Engineering, 2023, 30(3): 118.

（责任编辑　杜亚勤）