

港口企业知识库智能问答系统多维评价体系 构建与实证研究

张昭

华北理工大学

摘要: 港口企业在数字化转型中积累了大量非结构化知识资源, 基于大语言模型与检索增强生成 (RAG) 技术构建的智能问答系统为港口知识管理提供了新的技术路径。然而, 现有评价研究多聚焦于算法性能指标, 缺乏面向高安全、高合规工业场景的综合效能评价框架。本文面向港口企业知识库智能问答系统的应用评价需求, 基于 D&M 信息系统成功模型, 引入认知负担理论与 SECI 知识转化模型, 构建了包含知识与内容质量、系统与效能、用户体验与管理效益 3 个一级指标和 9 个二级指标的多维评价体系。运用德尔菲法进行指标筛选, 采用层次分析法 (AHP) 确定指标权重, 并通过一致性检验保障权重分配的逻辑合理性。在此基础上, 以某港口企业实际部署的融合知识图谱与向量检索的 KG-RAG 智能问答系统为评价对象, 基于自建 PortBiz-QA-200 测试集开展消融实验与真实业务盲测。结果表明, KG-RAG 架构较传统单路 RAG 在事实准确与无幻觉率、多跳逻辑关联能力两项核心指标上分别提升了 25.2% 和 45.5%, 系统综合加权得分达 91.35 分。研究表明, 所构建的评价体系能够较为全面地反映智能问答系统在高风险业务场景中的技术性能与管理应用价值, 可为港口及其他高安全行业的知识管理系统评价提供参考。

关键词: 智能问答系统; 多维评价体系; 层次分析法; 检索增强生成; 知识图谱

DOI:10.65976/3105-8450.2026.01.007

Construction and Empirical Study of a Multi-Dimensional Evaluation System for Intelligent Q&A Systems in Port Enterprise Knowledge Bases

Zhang Zhao

(North China University of Science and Technology)

Abstract: Port enterprises have accumulated a vast amount of unstructured knowledge resources during digital transformation. Intelligent question-and-answer (Q&A) systems built upon large language models (LLMs) and Retrieval-Augmented Generation (RAG) technology offer a promising technical pathway for port knowledge management. However, existing evaluation studies predominantly focus on algorithmic performance metrics, lacking a comprehensive effectiveness evaluation framework tailored to high-safety, high-compliance industrial scenarios. Addressing the evaluation needs of intelligent Q&A systems for port enterprise knowledge bases, this study draws upon the DeLone and McLean (D&M) Information Systems Success Model, and further incorporates Cognitive Load Theory and the SECI knowledge conversion model to construct a multi-dimensional evaluation framework comprising three first-level dimensions—Knowledge and Content Quality, System and Technical Effectiveness, and User Experience and Management Benefits—encompassing nine second-level indicators. The Delphi method is employed for indicator screening, the Analytic Hierarchy Process (AHP) is applied to determine indicator weights, and consistency tests are conducted to ensure the logical rationality of weight allocation. On this basis, a KG-RAG intelligent Q&A system integrating knowledge graph reasoning with vector retrieval, deployed in an actual port enterprise, serves as the evaluation object. Ablation experiments and real-world blind business tests are carried out using a self-constructed domain-specific test set, PortBiz-QA-200. The results demonstrate that the KG-RAG architecture achieves improvements of 25.2% and 45.5% over a conventional single-path RAG system in two core indicators—Factual Accuracy and Hallucination-Free Rate, and Multi-hop Logical Association Capability, respectively—with a comprehensive weighted score of 91.35. The findings indicate that the proposed evaluation framework can comprehensively reflect both the technical performance and management application value of intelligent Q&A systems in high-risk business scenarios, and may serve as a reference for knowledge management system evaluation in port and other high-safety industries.

Keywords: intelligent question-and-answer system; multi-dimensional evaluation framework; Analytic Hierarchy Process; Retrieval-Augmented Generation; knowledge graph

作者简介: 张昭 (1993—), 男, 硕士, 研究方向为工程管理。

1 研究概述

1.1 研究背景与问题提出

随着我国“交通强国”战略的深入实施，港口企业正处于数字化与智能化转型的关键阶段。港口在复杂的工程建设与运营管理过程中积累了大量规章制度、操作手册和设备维保台账等非结构化知识资源。然而，受制于异构系统分散、传统关键词检索语义理解能力不足以及专家经验难以显性化沉淀等因素，现有知识管理方式难以有效支撑港口高风险业务场景下的快速、准确和可追溯问答需求^[1-2]。

近年来，大语言模型（LLMs）与检索增强生成（RAG）技术为企业知识管理提供了新的解决方案^[3-4]。在此基础上，部分研究进一步引入知识图谱（KG）技术，构建 KG-RAG 融合架构以增强复杂业务逻辑的关联推理能力^[5-7]。然而，现有研究在系统评价方面仍存在明显不足，多数文献主要关注 BLEU、ROUGE 等文本生成质量指标或检索准确率等技术性能指标^[8]，较少从工程管理视角考察系统在高风险业务场景中的实际应用成效。对于港口企业这类高安全、高合规的工业组织而言，系统不仅要“能回答”，更需要在答案准确性、可追溯性、认知减负和知识转化等方面具备可验证的管理价值。如何构建兼顾技术性能与管理效益的综合评价框架，成为智能问答系统从技术原型走向工程应用的关键问题。

1.2 文献综述

1.2.1 信息系统评价理论研究

DeLone 和 McLean 提出的信息系统成功模型（D&M Model）是信息系统评价领域的经典理论框架^[9]，其核心维度包括系统质量、信息质量、服务质量、用户使用、用户满意和组织净收益。该模型已被广泛应用于企业信息系统、电子政务和电子商务等领域的效能评估。近年来，部分学者开始将 D&M 模型拓展应用于人工智能系统的评价。张金营等^[10]针对电力领域专业问答系统构建了包含多维度的评价模型，采用量化打分验证了系统的综合效能。然而，面向大语言模型驱动的智能问答系统，现有评价研究仍以技术指标为主，较少将信息质量、用户认知负担和组织知识转化等管理维度纳入统一框架。

1.2.2 认知负担理论与用户体验评价

认知负担理论（Cognitive Load Theory）强调，当系统输出信息质量不足或交互流程复杂时，用户需要投入更多时间与精力进行甄别、理解和二次确认，从而形成较高的认知负荷^[11]。在港口生产场景中，一线作业人员频繁面临跨部门、跨系统的信息检索任务，传统检索方式返回大量无关文档的问题显著增加了用户的信息筛选

成本。将认知减负指标纳入智能问答系统评价体系，有助于从用户视角衡量系统的实际应用价值。

1.2.3 知识管理理论在工业场景中的应用

Nonaka 提出的 SECI 知识转化模型^[12]将组织知识转化过程划分为社会化、外化、组合和内化四个阶段。港口企业中大量核心作业经验长期掌握在资深员工手中，属于难以结构化表达的隐性知识。智能问答系统通过知识图谱建模与自然语言交互，有望在一定程度上支撑隐性经验的显性化沉淀与跨岗位共享。然而，如何评估系统在知识外化与组织传承方面的实际贡献，目前尚缺乏系统性的指标设计。

1.2.4 文献述评

综上，现有研究在大语言模型问答系统的技术性能评价方面已取得一定成果，但面向港口企业等高安全、高合规工业场景，仍存在以下不足：（1）评价维度单一，偏重算法指标，缺乏从工程管理视角对系统应用效能的综合考量；（2）未将用户认知负担和知识转化等组织管理要素纳入评价框架；（3）缺少基于真实工业业务场景的实证评价案例^[13-14]。针对上述不足，本文尝试构建面向港口企业知识库智能问答系统的多维评价体系，并通过实证测试验证其有效性。

1.3 研究目的与技术路线

本文旨在解决“港口企业知识库智能问答系统如何进行综合效能评价”这一核心问题。研究按照“理论映射—指标构建—权重确定—实证验证”的思路展开。首先，基于 D&M 信息系统成功模型、认知负担理论和 SECI 知识转化模型，建立评价理论框架；其次，运用德尔菲法进行指标筛选，采用 AHP 层次分析法确定权重；最后，以某港口企业实际部署的 KG-RAG 智能问答系统为对象，开展消融实验与真实业务盲测，验证评价体系的适用性^[15]。

2 评价体系构建

2.1 理论映射与设计原则

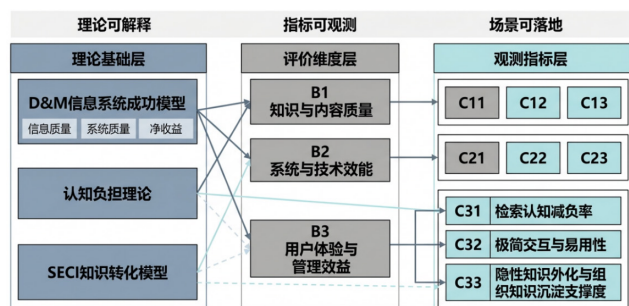
港口企业知识库智能问答系统兼具信息系统、人工智能工具和知识管理载体三重属性，单一理论难以完整覆盖其评价需求。本文在 D&M 模型基本逻辑基础上，结合港口高风险业务特征、用户减负需求以及知识转化目标，对评价维度进行适应性调整，形成“理论—维度—指标”的映射关系。

（1）D&M 模型中的信息质量、系统质量与净收益维度，分别映射至“知识与内容质量”“系统与技术效能”“用户体验与管理效益”三个一级评价维度；（2）认知负担理论重点体现在“检索认知减负率”和“极简交互与易用性”两个指标中；（3）SECI 知识转化模型主要对应“隐性知识外化效能”指标，用以评估系统对隐

性经验显性化与组织知识沉淀的支撑作用。

评价体系构建遵循以下设计原则：（1）理论可解释，每项指标均有明确的理论依据；（2）指标可观测，具备量化测量或专家评分的可操作性；（3）场景可落地，紧密结合港口业务痛点与管理实际。

理论与指标的映射关系如图 1 所示。



通过复合理论映射实现“理论基础—评价维度—观测指标”的结构化衔接
 图 1 评价理论与指标体系映射关系图

2.2 基于德尔菲法的指标筛选

2.2.1 初始指标池构建

本文以 D&M 信息系统成功模型和 RAG 系统常见评价维度为基础，参照《港口危险货物安全管理规定》等行业标准，初步梳理出包含 32 项潜在评估要素的指标海选池。

2.2.2 专家咨询与指标收敛

邀请 8 位具有 10 年以上港口从业与信息化经验的资深专家（包括 2 名安监部主任、2 名调度中心高级指挥、2 名港口信息化架构师及 2 名工程管理领域学者）组成专家咨询小组。通过内网邮件与背靠背问卷的形式，对初选指标进行了“剔除、合并与重构”操作。

专家组一致认为：（1）BLEU、ROUGE 等传统 NLP 指标无法衡量大模型在复杂工业场景中的事实性表现，应予以剔除；（2）显存占用率、并发吞吐量等纯 IT 运维指标偏离管理效能评估范畴，予以删减；（3）同质化

较高的指标进行了科学归并。

经过两轮专家咨询，最终确立了 3 个一级指标和 9 个二级指标的评价体系。指标筛选过程与结果如表 1 所示。

3 系统架构设计

3.1 评价指标体系

3.1.1 一级指标设计（见表 2）

3.1.2 二级指标设计

（1）知识与内容质量（B1）维度：C11 事实准确与无幻觉率——评估模型输出是否严格遵从现行规章制度，是否存在伪造安全标准或虚构设备参数等事实偏差；C12 多跳逻辑关联能力——评价引入知识图谱后，系统面对多重约束叠加问题时能否完整提取前置条件并给出综合判断；C13 知识溯源与引用可靠性——评估答案是否准确附带规章文件的具体来源。

（2）系统与效能（B2）维度：C21 系统响应与推理时延——评估双路召回与模型推理的总耗时是否满足应急调度需求；C22 非结构化数据解析完整度——评估预处理管道对复杂 PDF、表格等异构数据的清洗能力；C23 数据隔离与安全防护能力——评估本地化部署方案的数据安全性。

（3）用户体验与管理效益（B3）维度：C31 检索认知减负率——评估相比传统检索方式，系统直接输出结论所节省的员工信息筛选成本；C32 系统易用性与交互体验——评估一线人员能否通过自然语言无障碍获取所需信息；C33 隐性知识外化效能——评估系统能否将非结构化经验转化为可复用知识。

AHP 层次结构模型如图 2 所示。

3.2 基于 AHP 的权重确定

3.2.1 一级指标权重

利用 1-9 标度法，对 8 位专家的打分进行几何平均，构建一级指标判断矩阵如表 3 所示。

表 1 初始指标海选池与德尔菲法专家筛选结果

初选维度归属	初始海选指标数量	专家组筛选决议	最终确立指标
知识与内容质量	10 项	剔除 3 项、合并为 3 项	C11 事实准确与无幻觉率、C12 多跳逻辑关联能力、C13 引用溯源可靠性
系统与效能	10 项	剔除 4 项、合并为 3 项	C21 检索与推理延迟、C22 异构数据解析完整度、C23 数据私有化隔离度
用户体验与管理效益	12 项	剔除 4 项、合并为 3 项	C31 检索认知减负率、C32 极简交互与易用性、C33 隐性知识外化效能

表 2 一级指标设计

一级指标	维度	评价作用
知识与内容质量（B1）	核心业务维	衡量系统输出答案的准确性、合规性及逻辑严密性
系统与效能（B2）	工程支撑维	衡量系统响应时延、异构数据解析能力与数据安全隔离能力
用户体验与管理效益（B3）	组织赋能维	衡量系统在认知减负、知识流转方面的实际管理价值

表3 一级指标判断矩阵 (A-B)

总目标 A	B1	B2	B3
B1	1	5	3
B2	1/5	1	1/2
B3	1/3	2	1

运用方根法计算得出权重向量 $W=(0.637,0.105,0.258)$ ，最大特征根 $\lambda_{max}=3.038$ ，一致性指标 $CI=0.019$ ，一致性比率 $CR=0.032 < 0.1$ ，通过一致性检验。

结果表明，在港口高风险业务场景中，专家组普遍认为知识与内容质量 (B1) 具有最高重要性，其次为用户体验与管理效益 (B3)，系统与技术效能 (B2) 居后。这一权重分布与港口企业对安全合规和管理减负的核心需求高度吻合。

3.2.2 二级指标权重

采用同样方法计算各一级指标下属二级指标的局域权重，判断矩阵及权重计算结果如表 4-6 所示。

表4 B1-C 层判断矩阵及权重

指标	C11	C12	C13	局域权重
C11	1	2	3	0.540
C12	1/2	1	2	0.297
C13	1/3	1/2	1	0.163

$\lambda_{max}=3.009$ ， $CI=0.0045$ ， $CR=0.0078 < 0.1$ ，通过检验。

表5 B2-C 层判断矩阵及权重

指标	C21	C22	C23	局域权重
C21	1	2	1	0.400
C22	1/2	1	1/2	0.200
C23	1	2	1	0.400

$\lambda_{max}=3.000$ ， $CI=0$ ， $CR=0 < 0.1$ ，通过检验。

表6 B3-C 层判断矩阵及权重

指标	C31	C32	C33	局域权重
C31	1	3	1	0.450
C32	1/3	1	1/2	0.200
C33	1	2	1	0.350

$\lambda_{max}=3.018$ ， $CI=0.009$ ， $CR=0.0155 < 0.1$ ，通过检验。

3.2.3 全局权重分布

综合各层级计算结果，得出 9 个二级指标相对于总目标的全局绝对权重，如表 7 所示。

从全局权重可见，C11 (0.344) 与 C12 (0.189) 合计占比超过 53%，反映了港口业务对操作合规性与安全性的极高要求。C31 (0.116) 位列第三，凸显了系统为员工减负的核心管理目标。

3.3 综合评价模型

基于 AHP 全局权重，系统综合效能得分计算公式为：

$$S_{total} = \sum_{i=1}^n (W_i \times E_i) \quad (1)$$

式中， S_{total} 为系统综合管理效能得分， W_i 为各二级指标的全局绝对权重， E_i 为系统在各指标上的实际测试评分 (归一化至 0-100 分制)。

4 实证验证

4.1 评价对象

以某港口企业实际部署的智能问答系统为评价对象。该系统采用融合知识图谱与向量检索的 KG-RAG 双路召回架构，底层基座模型为本地化部署的 DeepSeek-R1 (14B 参数，INT4 量化)，知识底座包含约 450 份核心业务文档 (约 200 万字)，构建了包含约 15,000 个业务实体和 32,000 条逻辑关系边的港口领域知识图谱。

4.2 测试集与实验设计

4.2.1 PortBiz-QA-200 测试集

联合港口企业安监部与调度中心，基于历史真实工单与安全规程，人工构建了包含 200 个问答对的 PortBiz-QA-200 测试集。测试集按业务复杂度划分为三类：事实检索类 (80 题，40%)、多约束逻辑推理类 (80 题，40%) 和安全红线抗诱导类 (40 题，20%)。

4.2.2 消融实验设计

设置三种模型配置进行对比，如表 8 所示。

模型 B 与模型 C 采用相同的 BGE-m3 嵌入模型与 BGE-Reranker 重排策略，唯一变量在于模型 C 引入了 Neo4j 知识图谱的双路召回机制。

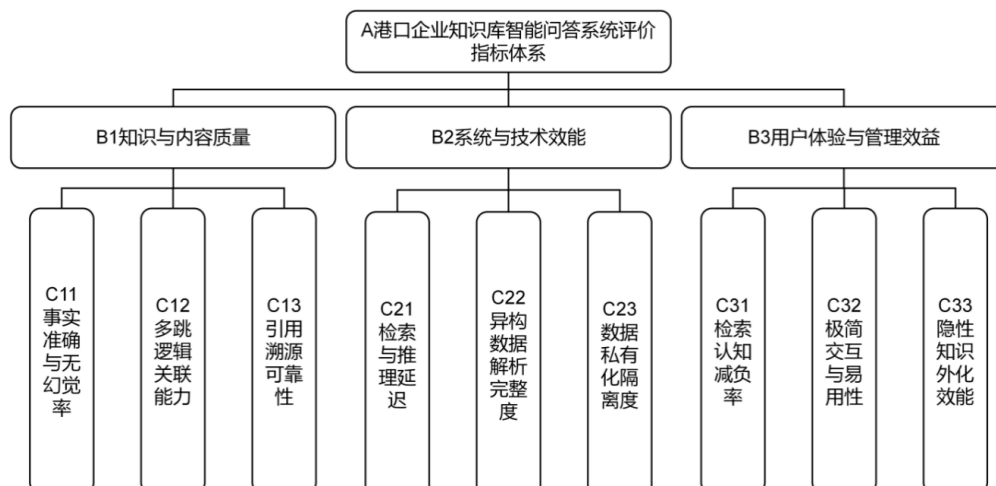


图2 港口企业智能问答系统 AHP 层次结构模型图

表 7 多维评价体系全局权重分布

一级指标 (权重)	二级指标	局域权重	全局权重	排序
B1 知识与内容质量 (0.637)	C11 事实准确与无幻觉率	0.540	0.344	1
	C12 多跳逻辑关联能力	0.297	0.189	2
	C13 引用溯源可靠性	0.163	0.104	4
B2 系统与技术效能 (0.105)	C21 检索与推理延迟	0.400	0.042	7
	C22 异构数据解析完整度	0.200	0.021	9
	C23 数据私有化隔离度	0.400	0.042	7
B3 用户体验与管理效益 (0.258)	C31 检索认知减负率	0.450	0.116	3
	C32 极简交互与易用性	0.200	0.052	6
	C33 隐性知识外化效能	0.350	0.090	5

4.2.3 评价方法

消融实验重点评估 C11 (事实准确与无幻觉率) 和 C12 (多跳逻辑关联能力) 两项权重最高的指标。由 5 名资深业务专家按照四级评分量表 (优秀 90-100、良好 75-89、及格 60-74、不及格 0-59) 进行独立盲评。

综合管理效能评价邀请 8 位 AHP 专家与 42 位一线员工 (共 50 人) 参与为期两周的真实业务盲测, 采用双盲机制打分, 基于 5 点 Likert 量表线性映射至百分制后代入公式 (1) 计算。

4.3 实验结果

4.3.1 核心技术指标对比

三组模型在 C11 与 C12 指标上的盲测得分如表 9 所示。

模型 C 对比模型 B, C11 提升 $(94.28-75.32)/75.32 \approx 25.2\%$, C12 提升 $(89.10-61.25)/61.25 \approx 45.5\%$ 。同时, 模型 C 的标准差远低于其他两组, 说明 KG-RAG 架构不仅提升了平均得分, 还增强了输出稳定性。

(1) 事实准确与无幻觉率 (C11) 分析。模型 A 由于缺乏私域知识支撑, 在面对特定设备参数、规章条款类问题时表现出较明显的事实性幻觉倾向。模型 B 通过引入向量检索获得了外部知识锚点, 但在规章版本更迭时容易召回废止文档。模型 C 依托图数据库中的时效约束关系边与安全熔断机制, 有效过滤了过期规章的误导。

(2) 多跳逻辑关联能力 (C12) 分析。该指标是验

证知识图谱核心价值的关键所在。模型 B 在处理跨文档多条件约束问题时, 因文本切片导致上下文物理割裂, 逻辑链条断裂。模型 C 利用图谱中的实体关联进行 N 跳广度优先遍历, 成功提取分散在异构文档中的前置条件, 在推理完整性方面表现显著优于单路 RAG。

4.3.2 综合管理效能评价

鉴于模型 A 不具备业务应用条件, 综合评价阶段仅比较模型 B 与模型 C。基于 AHP 全局权重和真实业务盲测得分, 计算综合管理效能, 结果如表 10 所示。

结果表明, KG-RAG 系统综合加权得分 (91.35 分) 显著优于传统 RAG 系统 (69.73 分)。从各维度来看: (1) 在 B1 维度中, 模型 C 在事实准确性、多跳逻辑关联能力和引用溯源可靠性方面均表现较优, 说明知识图谱与向量检索协同的双路召回机制能够增强系统在高风险场景中的答案可信度。(2) 在 B2 维度中, C21 指标模型 C 略低于模型 B (82 vs.88), 这是因为 KG-RAG 的双路召回增加了图谱遍历计算开销, 但差异在可接受范围内。C23 指标两者相同 (均为 95), 因为均采用本地化部署。(3) 在 B3 维度中, C31 指标模型 C (92) 显著高于模型 B (60), 表明 KG-RAG 系统能够更有效地直接输出结构完整的答案, 减少用户二次筛选负担。C33 指标模型 C (86) 显著高于模型 B (55), 说明知识图谱建模有助于将经验规则转化为可复用的数字知识资源。

表 8 实验模型配置

模型	接入向量库	接入知识图谱	重排机制	用途
Base-LLM (A)	否	否	否	零样本基线
Naive-RAG (B)	是	否	是	传统 RAG 对照
KG-RAG (C)	是	是	是	本文方法

表 9 三组模型核心技术指标盲测得分

模型	C11 均分	C11 标准差	C12 均分	C12 标准差
Base-LLM (A)	42.15	28.4	35.40	31.2
Naive-RAG (B)	75.32	14.6	61.25	18.5
KG-RAG (C)	94.28	5.3	89.10	7.1

表 10 综合管理效能评分对比

一级指标	二级指标	全局权重	模型 B 得分	模型 C 得分
B1	C11 事实准确无幻觉率	0.344	75	94
	C12 多跳逻辑关联	0.189	61	89
	C13 引用溯源可靠性	0.104	70	95
B2	C21 检索与推理延迟	0.042	88	82
	C22 解析完整度	0.021	65	90
	C23 数据私有化隔离度	0.042	95	95
B3	C31 检索认知减负率	0.116	60	92
	C32 极简交互易用性	0.052	80	88
	C33 隐性知识外化效能	0.090	55	86
综合加权总分		1.000	69.73	91.35

5 结语

本文面向港口企业知识库智能问答系统的应用评价需求，基于 D&M 信息系统成功模型、认知负担理论和 SECI 知识转化模型，构建了包含 3 个一级指标和 9 个二级指标的多维评价体系，运用 AHP 方法确定了各级权重，并通过消融实验与真实业务盲测进行了实证验证。主要结论如下：（1）港口企业智能问答系统的评价不能仅关注算法性能，还应统筹考虑安全合规性、认知减负效果和知识转化价值。本文构建的多维评价体系实现了技术性能与管理效益的综合衡量。（2）权重分析表明，事实准确与无幻觉率（0.344）和多跳逻辑关联能力（0.189）在港口业务场景中具有最高权重，检索认知减负率（0.116）位列第三，为系统优化方向提供了量化依据。（3）实证结果显示，KG-RAG 架构较传统 RAG 在核心技术指标和综合管理效能方面均有明显提升，综合加权得分从 69.73 分提升至 91.35 分，验证了评价体系的区分度与适用性。

本研究所构建的评价框架可为港口企业及其他高安全、高合规行业的知识管理系统效能评估提供参考。

参考文献：

[1] 张斌,魏扣,郝琦.国内外知识库研究现状述评与比较[J].图书情报知识,2016(3):15-25.
 [2] 张敏,邹盛,沈高峰,等.电力设计企业知识管理关键技术及应用功能探讨[J].企业改革与管理,2021(20):40-41.
 [3] Lewis P,Perez E,Piktus A,et al.Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//Advances in Neural Information Processing Systems 33.2020:9459-9474.
 [4] Pan S,Luo L,Wang Y,et al.Unifying large language models and knowledge graphs:A roadmap[J].IEEE Transactions on Knowledge and Data Engineering,2024,36(7):3007-3026.
 [5] 胡佳慧,李姣,姚宽达,等.大语言模型融合知识图谱的

医学问答系统构建研究[J].中国数字医学,2024,19(06):91-95.
 [6] 石致远,张佳蕾,孔志伟,等.结合知识图谱与大语言模型的风电装备智能问答系统[J].东方电气评论,2024,38(3):77-84.
 [7] Kang M, Kim J. A comprehensive survey on enterprise knowledge graph construction and application[J]. IEEE Access, 2022, 10: 106981-107005.
 [8] Chang Y, Wang X, Wang J, et al. A survey on evaluation of large language models[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(3): 1-45.
 [9] DeLone W H, McLean E R. The DeLone and McLean model of information systems success: A ten-year update[J]. Journal of Management Information Systems, 2003, 19(4): 9-30.
 [10] 张金营,王天堃,么长英,等.基于大语言模型的电力知识库智能问答系统构建与评价[J].计算机科学,2024,51(12):286-292.
 [11] Sweller J. Cognitive load theory, learning difficulty, and instructional design[J]. Learning and Instruction, 1994, 4(4): 295-312.
 [12] Nonaka I, Takeuchi H. The knowledge-creating company: How Japanese companies create the dynamics of innovation[M]. Oxford University Press, 1995.
 [13] 赵雪琴. 基于 RAG 的科技奖励知识库构建与应用研究[J]. 情报探索, 2024(11): 75-81.
 [14] 王东清, 芦飞, 张炳会, 等. 大语言模型中提示词工程综述[J]. 计算机系统应用, 2025, 34(1): 1-10.
 [15] Munim Z H, Dushenko M, Jimenez V J, et al. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions[J]. Maritime Policy & Management, 2020, 47(5): 577-597.