Computer Science and Application 计算机科学与应用, 2022, 12(12), 2983-2993
Published Online December 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.1212302

Hans汉斯

基于知识图谱技术的风险作业智能化管理研究

赵　丹，杨　辉，罗舒元，郭丽雪

中国石油西南油气田分公司川中油气矿，四川　遂宁

收稿日期：2022年11月26日；录用日期：2022年12月23日；发布日期：2022年12月30日

摘　要

作业风险因素管理和风险预警管控作为遏制重特大事故、实现安全生产超前预防的主要措施途径，对提升企业安全生产保障水平、提高政府安全监管效能具有至关重要的作用。据国务院安全生产委员会历年安全生产大检查情况通报显示，2007年至2015年全国共排查事故隐患5298万项，年均排查隐患589万项；海量事故隐患数据被科学地记录与存储，成为辅助安全生产的重要因素。然而，由于大数据存续时间较短，大数据分析与处理能力相对薄弱，大多数企业与机构仅能利用事故隐患数据指导整改工作，无法分析数据背后的潜在价值，对安全生产管理工作的指导作用有限，我国安全生产形势仍不容乐观。石油企业现有高风险作业类型多、数量大、管理要求高，作业的风险识别、风险等级划分、控制措施制定等风险因素因人员能力和素质差距导致分析的结果差异较大，同类风险的管控措施不一致、标准不统一，容易出现风险管控的漏项缺项，目前无有效手段将公司每年2万余项高风险作业数据信息进行整合利用，高风险作业管控难度大。本文研究的基于知识图谱技术构建的石油天然气行业风险作业知识库，运用人工智能技术，建立知识自动加工处理流程，通过深入研究和应用，实现风险作业的“知识标签”及“关联关系”。通过整合共享全公司风险作业步骤、作业风险因素、风险控制措施等数据，提升作业风险数据的应用价值，减少同类不安全行为、不安全状态发生的频率，实现各基层单位风险防控措施标准化管理，提醒高风险作业全过程管理的风险源、风险控制措施，能够更高效、规范和安全的开展风险作业，建立人工智能及知识图谱技术平台，构建自主学习词库，完善和丰富高风险作业知识体系。综上，通过智能化信息手段，利用知识图谱技术建立风险作业行业知识图谱，采集并整合各类风险数据，实现作业风险提示、风险控制措施、事故事件等信息的精准推送，进一步提高风险管控水平，助推智能化油气田建设工作。

关键词

风险作业，专业词库，知识图谱，智能搜索，精准推送

Research on Risk Operation Intelligent Management Based on Knowledge Graph Technology

Dan Zhao, Hui Yang, Shuyuan Luo, Lixue Guo

文章引用: 赵丹, 杨辉, 罗舒元, 郭丽雪. 基于知识图谱技术的风险作业智能化管理研究[J]. 计算机科学与应用, 2022, 12(12): 2983-2993. DOI: 10.12677/csa.2022.1212302

Central Sichuan Oil and Gas Distract of Petrochina Southwest Oil and Gas Field Company, Suining Sichuan

Received: Nov. 26th, 2022; accepted: Dec. 23rd, 2022; published: Dec. 30th, 2022

Abstract

Operation risk factor management and risk early warning control, as the main measures to curb major accidents and realize advance prevention of production safety, play a crucial role in improving the level of production safety guarantee of enterprises and improving the efficiency of government safety supervision. From 2007 to 2015, 52.98 million hidden dangers were investigated nationwide, with an average of 5.89 million being checked annually. Massive accident data are recorded and stored scientifically, which becomes an important factor to assist production safety. However, due to the short existence time of big data, big data analysis and processing ability is relatively weak, most enterprises and organizations can only use the data of hidden accidents to guide the rectification work, unable to analyze the potential value behind the data, the guidance role of the safety production management is limited, our safety production situation is still not optimistic. The existing high-risk operations of petroleum enterprises have many types, large quantities, and high management requirements. The risk factors such as risk identification, risk grade classification, and control measures formulation of the operations, due to the gap in personnel ability and quality, lead to large differences in the analysis results. The control measures of similar risks are inconsistent and the standards are not uniform, which is prone to the occurrence of missing items in risk control. At present, there is no effective means to integrate and utilize the data information of more than 20,000 high-risk operations of the company every year, which makes it difficult to control high-risk operations. In this paper, the knowledge base of risk operations in oil and gas industry is built based on knowledge graph technology. Artificial intelligence technology is used to establish automatic knowledge processing process. Through in-depth research and application, the "knowledge label" and "association relationship" of risk operations are realized. Through the integration and sharing of the whole company's risk operation steps, operational risk factors, risk control measures and other data, improve the application value of operational risk data, reduce the occurrence frequency of similar unsafe behaviors and unsafe states, realize the standardized management of risk prevention and control measures of all grass-roots units, and remind the risk sources and risk control measures of the whole process management of high-risk operations. It can carry out risky operations in a more efficient, standardized and safe manner, establish artificial intelligence and knowledge graph technology platform, construct self-learning thesaurus, and improve and enrich the knowledge system of high-risk operations. In summary, through intelligent information means, knowledge graph technology is used to establish the knowledge graph of risk operation industry, collect and integrate all kinds of risk data, realize the accurate push of operation risk tips, risk control measures, accidents and other information, further improve the level of risk management and control, and boost the construction of intelligent oil and gas fields.

Keywords

Risk Operation, Professional Thesaurus, Knowledge Graph, Intelligent Search, Accurate Delivery

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

石油天然气行业涉及风险作业种类多、数量大、要求高，作业的分级分类管理、危害因素识别、控制措施制定等因素受人员能力素质、工作经历影响较大，同类风险的管控措施不一致、标准不统一，容易出现风险管控的漏项缺项，目前，暂无有效手段将公司每年开展的风险作业数据信息进行整合、分享、运用。本研究通过整合共享风险作业步骤、作业危害因素、风险控制措施等数据资源，建立风险作业知识图谱技术应用平台，构建自主学习词库，完善和丰富风险作业知识体系，进一步提升作业风险数据的应用价值，减少同类不安全行为、不安全状态发生的频率，促使各单位标准化制定风险防控措施，最终实现风险作业更高效、规范和安全地开展。

综上所述，通过智能化信息手段，利用知识图谱技术建立风险作业知识图谱，采集并整合各类风险数据，实现作业风险提示、风险控制措施、事故事件等信息的精准推送，进一步提高风险管控水平，助力“油公司”模式下质量安全环保业务数字化转型。

2. 风险作业知识图谱技术

图谱技术有两大阵营，一是以谷歌为代表的通用知识图谱[1]，主要是用互联网思维，解决通用的搜索、推荐、问答等业务，由于它强调的是广度，因而更多的是强调实体，很难生成完整的全局性本体层的统一管理；另一种是行业知识图谱，也就是我们拟采用的技术手段，具有以下特点：首先这是面向特定专业领域的知识图谱；其次是面向各级人员、各种专业操作，业务深度及完备性上更有保障；再次是数据精度高、对决策辅助及分析提供强有力支撑；最后是有严格和丰富的数据模式，实体属性具备专业性和复杂相关性，能有效地解决业务问题。

构建作业风险知识图谱，用到的技术主要包括文本分析、文本处理、知识库管理及知识图谱工具等[2]。从架构设计来说，分为数据来源、知识抽取、知识融合和知识应用四个层级。整体架构如图 1。

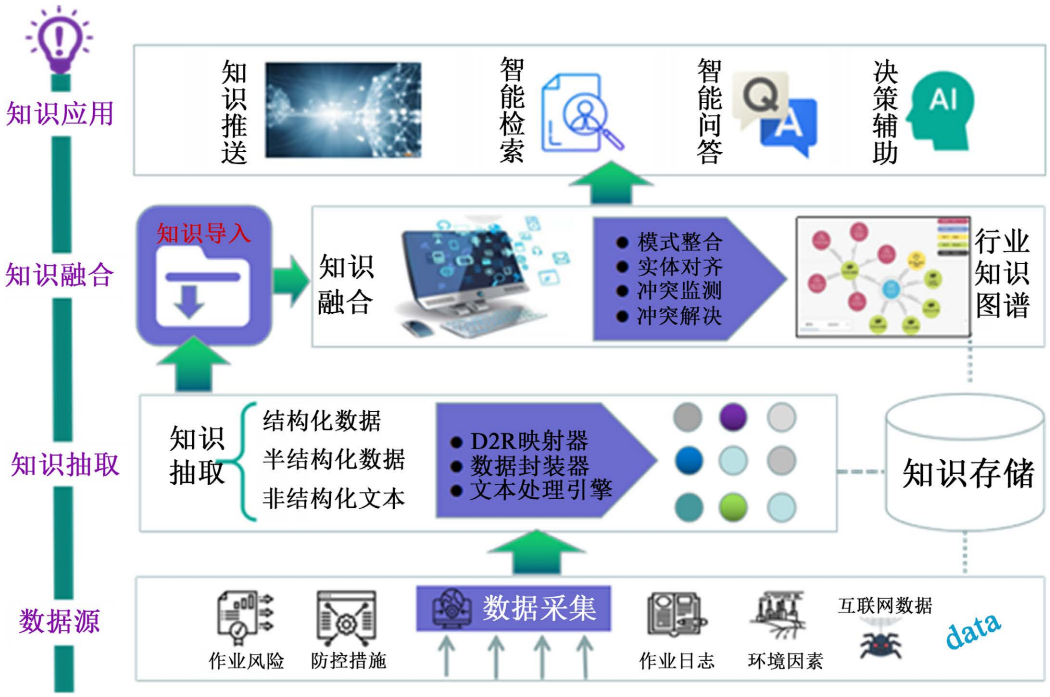


Figure 1. Overall architecture of knowledge graph
图 1. 知识图谱整体架构

1) 数据来源，主要包括日常风险作业管理工作中，具有指导作用的操作规程、风险防控办法、工作前安全分析、风险作业日志等内部数据，从互联网上也获取了部分行业内的作业风险管理及作业指导文献及安全事故案例等文档。

2) 知识抽取，主要是指从数据源中，利用文本分析，知识体系管理技术，将高风险作业知识数据提炼出来，通过自动化处理、人工审核，将知识数据结构化存储起来。

3) 知识融合，是对高风险作业中的风险因子的识别，抽取及融合处理，通过映射与归并处理，对风险类型，风险描述等内容进行整理，通过关系模型，标签模型等建立高风险数据之间的关联关系，规范化的管理知识节点、标签、链路等数据，以便开展数据分析。

4) 知识应用，主要是对知识体系的所有内容进行智能检索，智能推送，为业务系统，及安全生产工作进行推送服务，信息化支撑服务，对作业事件或安全信息进行捕获与分析，例如，主要针对每日新增的作业信息，运用高风险知识图谱，进行数据处理，风险提醒等操作。

3. 基于文本分析技术的风险作业词库设计

3.1. 文本挖掘流程

中文文本挖掘是分析中文文本数据，挖掘中文文本信息，进而获取中文文本知识的过程[3]。完整的文本挖掘流程应包括海量多元中文文本集合预处理(文本选择、清洗、分词、特征表述、特征提取等)、文本挖掘分析(文本分类、文本聚类、关联分析等)和文本知识可视化等步骤，见图 2。

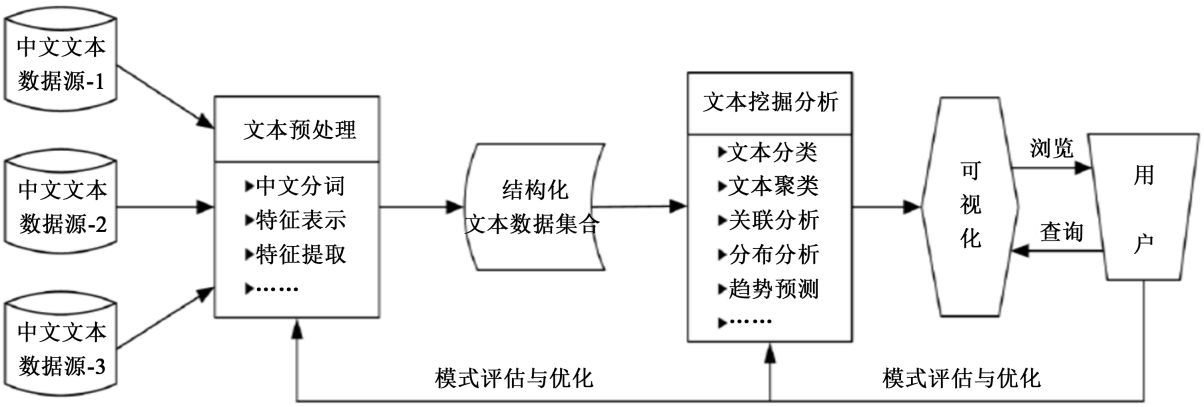


Figure 2. Text mining process
图 2. 文本挖掘流程

1) 中文文本预处理通常包括中文分词、特征表示和特征提取三个环节。a) 中文词汇与词汇之间没有固定间隔符，需要进行分词处理，主要有基于词库的分词方法(逐词遍历匹配法、正向最大匹配法、逆向匹配法等)和无词典分词技术(词频统计算法、依存句法分析、命名实体识别等)；b) 特征表示，利用描述性特征(名称、日期、大小、形状等)和语义性特征(标题、内容等)等特征项来代表文档，是非结构化向结构化转换的阶段，处理模型有布尔逻辑模型、向量空间模型、概率模型等；c) 特征提取，文本特征项数量往往较高，增加机器学习时间，影响分类学习准确率，需提取关键特征，降低特征向量维数，常用评估函数有信息增益等。

2) 文本挖掘分析，中文文本经中文分词、特征表示和特征抽取后，可利用机器学习、数据挖掘等技术方法提取分析文本中的潜在知识和模式，常用分析技术包括：文本分类技术、文本聚类技术、文本关联分析等。

3) 数据可视化，运用计算机图形学和图像处理技术，将文本信息、数据信息转换为图像信息。

4) 模式评估与优化，利用已经定义好的评估指标对提取分析的潜在知识和模式进行评价，如果结果符合评价标准，就存储知识模式供用户使用；否则返回到数据预处理和挖掘分析环节调整和优化，进行新一轮文本知识发现。

3.2. 风险作业专业词库

词库是文本分析的基础，词库有通用词库和专业词库，通用词库获取渠道为互联网[4]，我们参考了Jieba、IK、WORD 等中文 NLP 词库。专业词库，可以通过自行整理、学习训练来进行管理。词库数据源经过调研与论证，我们先明确了石油天然气行业相关的安全方面的国标、行标，分公司、油气矿内部相关的规范和制度文档。通过词库模型进行学习和训练，逐步充实和完善高风险作业专业词库。由油气矿安全管理业务部门提供基础数据；词库管理人员利用词库管理工具进行数据处理及词库管理；数据审核人员负责对词库数据质量整体把控。数据源有：国家标准、行业标准及单位安全制度、操作规范数据；互联网数据、行业数据；风险作业日志、计划数据等。词库业务流程见图 3。

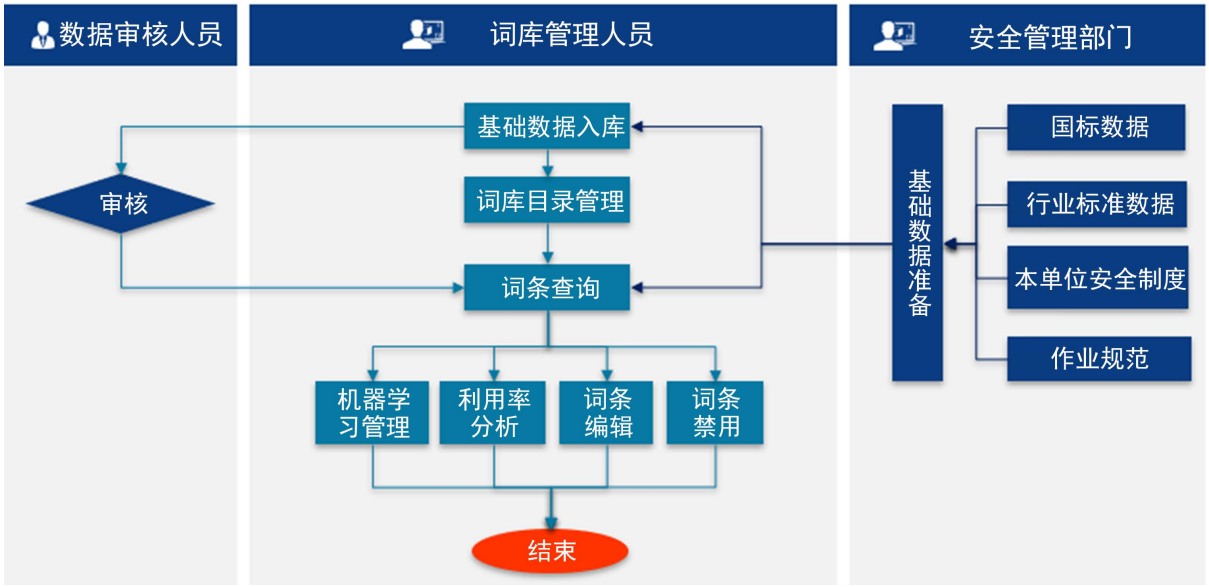


Figure 3. Thesaurus business process
图 3. 词库业务流程

业务场景设计，场景一：词库查询，见图 4。用户在任何需要调取词库的时候，可以通过数据接口和页面集成两种方式进行调用。数据接口方式，通过接口授权 TOKEN，以及传入词库类型，或者关键字，查询出相关类型的词库数据，或者符合指定查询条件的词库。页面集成方式，支持将词库类型，词库搜索框，以及词库的查询结构，封装成可以集成访问的页面，在分公司 SOA 单点登录验证通过后，发起访问。

场景二：近义词提醒。

每个人的词汇量，知识范围是有一定局限性的，我们在申请作业许可，或者编写制定安全防护措施的时候，如有词条提醒功能，便可更直观的发现编写内容是否符合标准规范，也能从相近的词条中得到思维的延展，将防护措施、作业指导内容编写得更为完善。经过训练、使用后的词库，近义词是根据词汇聚合关系的学习，不断累积后形成的，更具备专业性，见图 5。

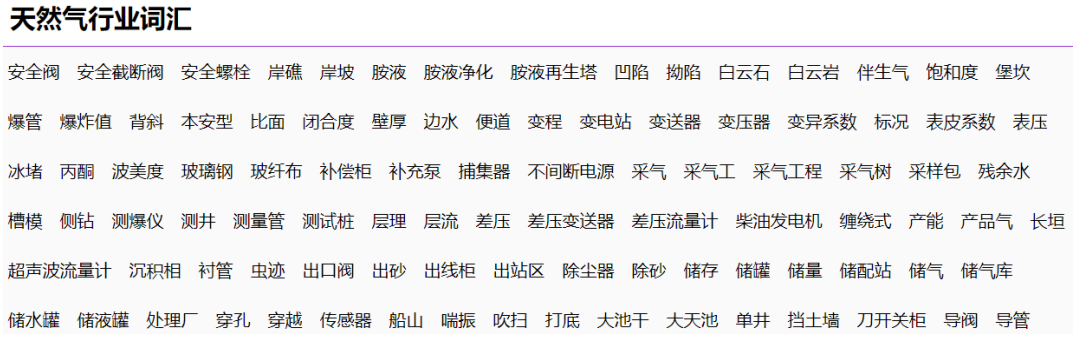


Figure 4. Professional word database query
图 4. 专业词库查询

Figure 5. Knowledge graph professional thesaurus synonyms
图 5. 知识图谱专业词库近义词

4. 基于知识图谱技术的风险作业管理平台设计

4.1. 风险作业知识图谱管理流程

基础数据由系统建立自动接口及定期导入等方式进行知识库数据的管理；知识图谱通过数据加载引擎进行入库管理；支持智能问答及可视化查询流程。一是知识库管理：知识体系目录、知识节点及知识标签管理。二是知识库数据审核。三是知识图谱数据自动加载、自动关联及数据可视化展示，见图 6。

基于知识图谱融合的海量知识和数据，用户通过自然语言进行搜索[5]。以下三种业务场景。一是语义搜索：基于知识图谱中的知识，解决传统搜索中遇到的关键字语义多样性及语义消歧的难题[6]；通过实体链接实现知识与文档的混合检索。二是智能问答：针对用户输入的自然语言进行理解，从知识图谱中或目标数据中给出用户问题的答案。答案经过关联分析，还具备向用户推送相关联的答案，相似度较高的答案。例如风险作业步骤，步骤的先后顺序，相关的风险控制措施，JSA 数据，事故案例，最近的作业申请等。三是可视化决策支持：通过提供统一的图形接口，结合可视化、推理、检索等，为用户提供信息获取的入口。

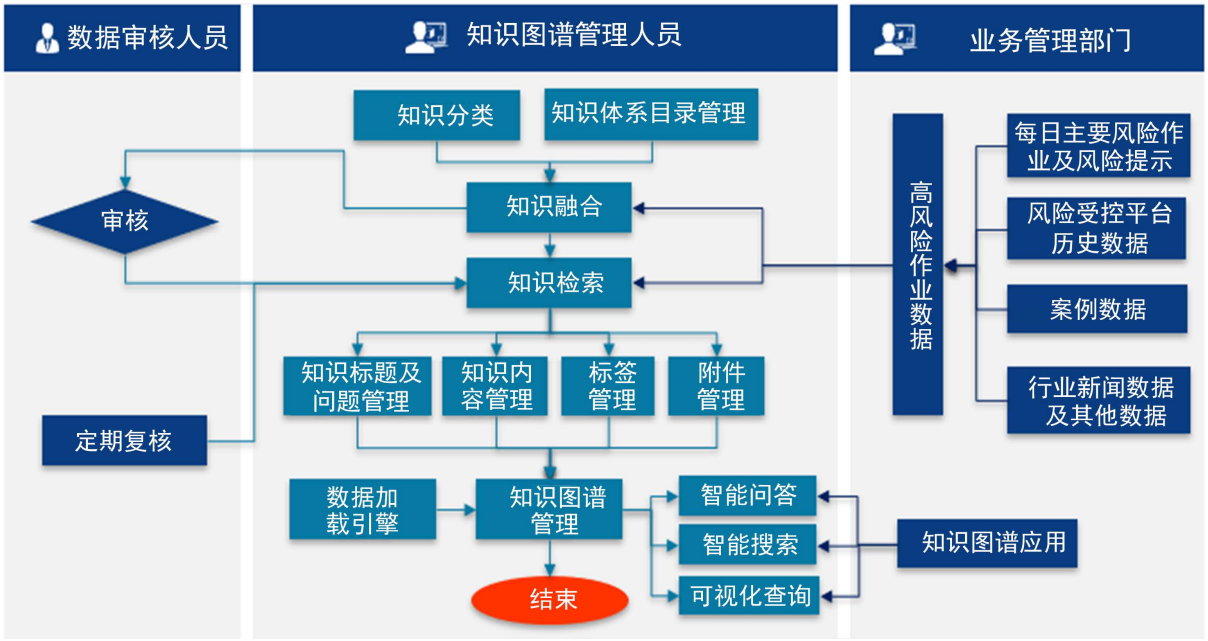


Figure 6. Knowledge graph business process
图 6. 知识图谱业务流程

4.2. 风险作业管理平台建设

4.2.1. 知识库管理

功能包括风险作业知识分类管理、风险作业知识目录结构管理、图文知识自动接口、图文知识导入、导出、图文知识编辑修改、知识标签智能推荐及专家经验管理、知识审核管理等[7]，见图 7。

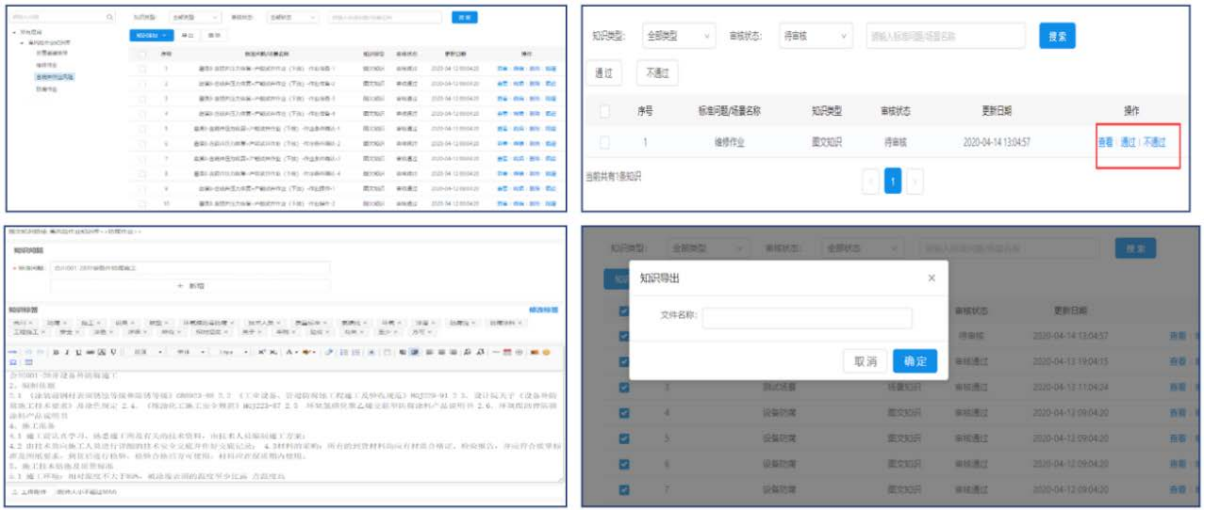


Figure 7. Knowledge base management function
图 7. 知识库管理功能

4.2.2. 知识图谱工具软件集成

通过技术手段，将 Neo4J 工具软件与本项目进行集成，完成风险作业知识库数据加载到知识图谱的脚本管理[8]，以及相关的 ETL 工具、知识数据和标签数据的关联关系、知识图谱可视化数据预览工具、

拓扑关系管理及关系校验、节点管理等，见图 8。

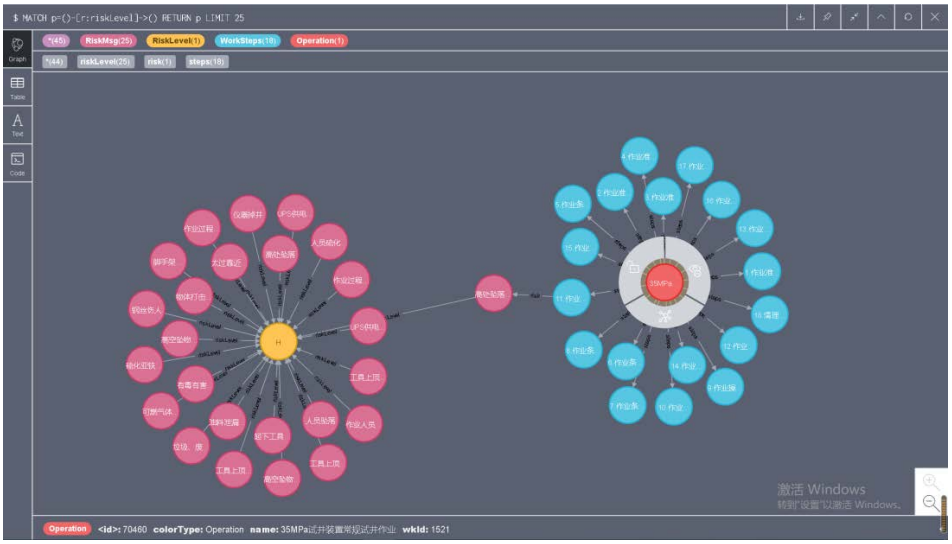


Figure 8. NEO4J features integration
图 8. NEO4J 功能集成

5. 风险作业管理平台建设效果

本文利用知识图谱技术把复杂的业务数据和知识体系通过数据挖掘、信息处理和图形绘制显示出来。通过整合结构化及非结构化数据，实现从目前零散分布的业务数据中发现知识，并智能融入知识体系，便捷的展示数据分类规则及内在联系，为风险作业信息的研究提供全方位、整体性、关系链的参考，从而帮助用户实现业务智能化。利用知识图谱技术采集并整合各类风险数据，从而建立风险作业知识图谱应用平台，对作业风险提示、风险控制措施、事故事件等信息的精准推送，进一步提高风险管控水平，助推智能化油气田建设工作。

1) 建成石油天然气行业首个“风险作业专业词库”。通过汇集各类风险作业数据以及文档资料，利用“文本分析”技术，实现数据的“内容分析、信息拆分、键值提取、关系校验”，制定数据管理标准，实现风险作业信息的规范入库。

2) 建成西南油气田公司首个基于知识图谱技术的“风险作业信息知识库”。利用自然语言处理(NLP)技术、Neo4J 架构、智能推送、智能问答等技术手段，实现知识体系的加工及迭代训练体系，以及应用支撑体系。

3) 实现风险作业信息管理工作的“IT 赋能”，助力业务部门及管理人员科学高效的获取作业风险数据，实现数据的共享及精准推送服务。

4) 通过作业风险因素智能化信息平台建立，实现了对知识进行有效管理、快速挖掘数据价值、有效信息快速推送、专业知识智能关联、业务视角综合分析、石油知识经验有效传承的价值目标。

本文研究成果已推广应用于公司各个单位，系统上线以来，系统提供稳定的应用及数据服务。1) 风险作业申请填报智能提醒：一线人员填写风险作业申请时，可以得到标准的作业前准备、工作步骤及描述、对应危害、对应控制措施等信息，提高申请填报的效率。2) 风险作业数据智能提醒及信息推送场景：系统每天从气矿网站获取当日的风险作业，并将所有数据解析解析、结构化、关联并存储到知识库。系统后台进行数据分析运算，将风险作业统计信息及分类、重点关注等内容以短信的形式将作业情况反馈给负责人。3) 风险作业数据智能提醒及信息推送场景：用户在作业区数字化管理平台进行作业申请后，

可以在后续工作步骤中，实时推送相关风险注意事项，以及对应的措施。

通过对风险知识体系进行集中管理，实现多触点的应用延伸，推进智能化油气田的建设工作，对油气安全生产工作提供更精准、可靠和全面的数据支撑。将风险作业信息及风险因素，精准及时推送到正在进行作业中去，让操作员第一时间获知相关的注意事项以及风险提示信息，为决策管理提供风险数据的支撑，提升 QSHE 整体管理水平。

5.1. 智能搜索、智能问答

风险作业人员进行风险作业时，只需要输入简单的关键词，机器人就可以自动查询相关风险作业的所有信息，实现高效精准查询。针对用户输入的自然语言进行理解，从知识图谱中或目标数据中给出用户问题的答案。答案经过关联分析，还具备向用户推送相关联的答案，相似度较高的答案，见图 9。可实现问题分析、语义理解、知识检索，智能推荐、机器人智能问答体系、智能纠错体系、使用感知反馈、知识利用率管理。

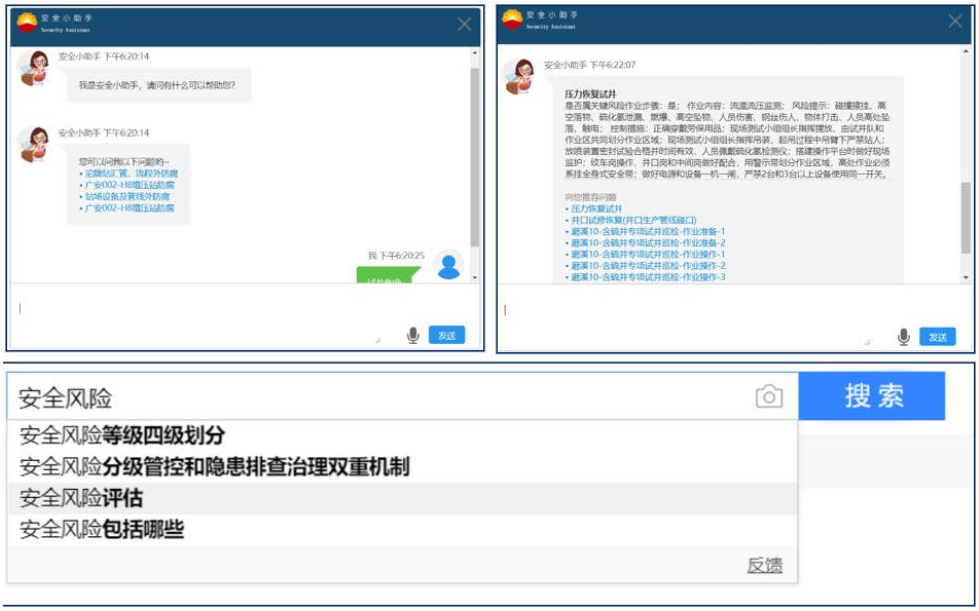


Figure 9. Intelligent question and answer service
图 9. 智能问答业务

5.2. 可视化查询

利用实体链接技术，进行基于知识图谱的语义搜索，同时通过应用平台提供数据共享发布服务接口，实现知识库数据的共享查询服务，见图 10。基于知识图谱可视化查询技术、完成数据的检索、锁定、关系及节点路径展示、结果保存、导出、拖拽式操作、节点延展、视图模式和数据表格模式切换。

5.3. 风险作业知识信息精准推送

将知识图谱信息及风险因素，精准及时推送到正在进行作业中去，让操作员第一时间获知相关的注意事项以及风险提示信息。

业务场景一：支撑作业管理数据，在风险作业管理数据中，为当前作业步骤推送作业风险及措施数据，同时，将相关步骤的数据也进行推送和提醒，见图 11。

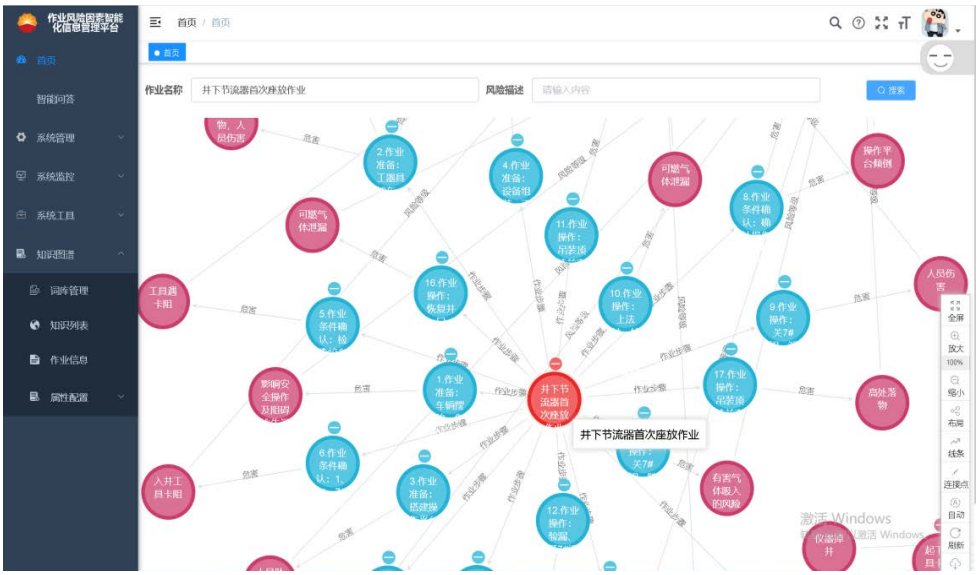


Figure 10. Visual query of risk operations graph
图 10. 风险作业图谱的可视化查询

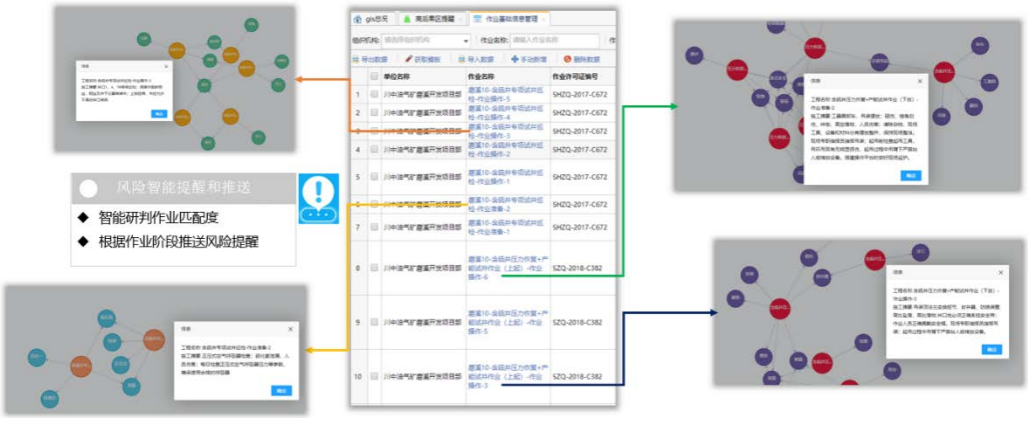
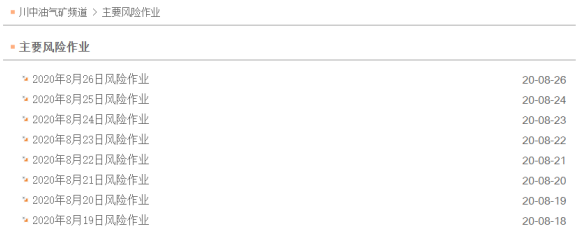


Figure 11. Risk job push service scenario 1
图 11. 风险作业推送业务场景一

业务场景二：为每日风险作业内容，补充智能推荐信息。减轻人工录入数据负担，见图 12。



今日作业内容	影响情况	今日风险作业主要提示	今日作业风险控制措施
围栏、平台拆除，生产流程拆除、设备吊装	无	1. 工器具伤人，人员绊倒，破坏地下隐蔽设施； 2. 人员灼烫伤、触电； 3. 吊车倾覆，吊绳损坏，人员伤亡、物体打击、挤压； 4. 硫化氢中毒、火灾爆炸、天然气泄漏。	1. 执行HSE目视化安全管理制度，作业区域设立警示隔离区域、警示标；保持场地整洁； 2. 穿戴好劳保用品，使用完好合格电器设备设施（发电机）；对电源导线定期进行检查，绝缘损坏的和有严重压痕的禁止使用； 3. 严禁超载超限（大于额定起重能力的75%），吊物及吊臂下严禁站人； 4. 穿戴好劳保用品；现场配备灭火器
站内土建施工，方井挡水条拆除	无	1. 工器具伤人，人员绊倒，破坏地下隐蔽设施； 2. 人员灼烫伤、触电； 3. 车辆损毁，人员伤亡、物体打击、挤压； 4. 硫化氢中毒、火灾爆炸、天然气泄漏。	1. 执行HSE目视化安全管理制度，作业区域设立警示隔离区域、警示标；保持场地整洁； 2. 穿戴好劳保用品，使用完好合格电器设备设施（发电机）；对电源导线定期进行检查，绝缘损坏的和有严重压痕的禁止使用 3. 穿戴好劳保用品；现场配备灭火器材，佩戴三合一气体监测仪；

Figure 12. Risk job push service scenario 2
图 12. 风险作业推送业务场景二

6. 结论

基于行业知识图谱技术构建石油天然气风险作业知识图谱，建立了风险作业因素智能化信息管理等应用模块，实现对作业风险提示、风险控制措施、事故事件等信息的精准推送及共享查询服务等功能。数据的精确推送能通过接口方式，对作业区管理平台“作业许可”申请等环节进行风险数据的智能提醒，支撑业务管理工作。

系统完成了专业词库建设，具备数据接口及自学习能力，能不断扩充风险因素等专业词库。风险作业知识库、知识标签及词库的管理及拓扑关系界面直观友好。系统通过收集整理风险作业数据，共完成了知识数据 16,876 条，词库数据 3000 余条，机器学习词条 1000 余条，生成 32,212 个风险作业信息知识节点、135,351 条关系数据。

本文研究的石油天然气安全生产领域作业风险管控，研究目的是利用知识图谱技术，构建专业词库，实现风险因素高效精准识别、实现作业风险提醒及数据共享应用。涉及的中文分词、关联规则等技术方法的性能指标还需要持续完善，以及大数据环境下，海量安全生产数据对各类技术方法的性能要求愈发苛刻，运算时间、加速性、可伸缩性、负载均衡性等性能指标将影响风险因素的获取和提醒及应用效果。因此，在未来研究中，应与计算机专业人员配合，设计算法并优化运行模式，降低中文分词运算时间，减少数据分析空间占用量，从多角度提高风险预警方法对安全生产大数据的运算效率。

参考文献

[1] 肖仰华, 等. 知识图谱[M]. 北京: 电子工业出版社, 2019.

[2] 欧艳鹏. 知识图谱技术研究综述[J]. 电子世界, 2018(13): 54+56.

[3] 刘峤, 李扬, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.

[4] 王元卓, 贾岩涛, 刘大伟, 等. 基于开放网络知识的信息检索与数据挖掘[J]. 计算机研究与发展, 2014, 52(2): 456-474.

[5] 陈学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.

[6] 卢道设, 杨世瀚, 吴尽昭, 等. 基于描述逻辑的组合知识库推理[J]. 计算机应用研究, 2013, 29(12): 4503-4506.

[7] 冯冬梅, 王乙慧. 基于知识图谱的国内工程风险管理可视化研究[J]. 广州大学学报(社会科学版), 2018, 17(10): 29-35.

[8] 帅健, 王伟, 梅苑, 金国清. 基于知识图谱的管道完整性管理研究特征与热点演化[J]. 油气储运, 2021(12): 1349-1357.