文章编号：2097-5899(2025)09-0052-06　　中图分类号：U298　　文献标识码：A
DOI：10.16669/j.cnki.issn.2097-5899.20240427001
引用格式：李涛宇，张振海. 基于知识图谱的编组站控制系统故障应急处置研究[J]. 铁路物流，2025，43(9)：52-57，64.
LI Taoyu，ZHANG Zhenhai. Research on Fault Emergency Response of Marshalling Station Control System Based on Knowledge Graph[J]. Railway Logistics，2025，43(9)：52-57，64.

基于知识图谱的编组站控制系统故障应急处置研究

李涛宇[1,2]，张振海[1]

(1. 兰州交通大学 自动化与电气工程学院，甘肃 兰州 730070；2. 中国铁路武汉局集团有限公司 武汉北站，湖北 武汉 430012)

摘　要：以铁路编组站控制系统的应急处置为背景，研究构建编组站控制系统安全应急处置知识图谱。在编组站控制系统安全应急处置实体和关系的获取工作中，对于非结构化数据，采用一种基于正则表达式和自然语言处理的技术，从特定格式的.docx文件中抽取实体和关系。在编组站控制系统安全应急处置知识图谱的实现工作中，基于Neo4j构建编组站控制系统安全应急处置知识图谱，实现编组站控制系统运行过程中故障数据集的可视化表达，利用Cypher语言实现故障发生时紧急处理流程的查询，研究成果可以为编组站控制系统故障状态下的应急处置提供参考。

关键词：知识图谱；编组站控制系统；知识存储；Neo4j；辅助决策

0　引言

编组站是铁路网上货物列车汇集、分流和中转的枢纽节点，其运转效率和货车最大保有量直接关系到全线的运输畅通和货物列车周转等[1]。随着国家“公转铁”战略的提出和现代化铁路运输服务体系建设的推进，铁路干线办理车辆激增，路网点线运能之间不匹配问题更加凸显，枢纽的运能已成为路网点线能力提升的“瓶颈”[2]。编组站综合集成自动化系统(CIPS)是编组站控制系统的一种，也是一种重要的铁路工业控制系统，在现代铁路运输中扮演着至关重要的角色[3-4]。然而，随着互联网技术的不断发展，编组站控制系统也面临着越来越多的安全威胁，一旦遭受到恶意攻击或者发生意外故障，可能会导致严重的后果。

为此，通过构建基于知识图谱的应急处置方法，可以更有效地识别和应对系统中的安全漏洞，提高系统的安全性和稳定性。同时，对相关领域知识进行整合和利用，更好地为系统运行中出现的问题提供解决方案，保障系统正常运行。在国外，最著名的是谷歌的知识图谱，还有DBpedia、Wikidata等。谷歌首次提出知识图谱理念时，创建了一张覆盖5.5亿个实体、3.5万个知识链接和180亿个知识三元组的知识图谱；DBpedia的知识图谱，为维基百科的知识提供了一套包含460万个实体、1 400个知识链和53.8万个知识三元组的知识图谱；Wikidata的知识图谱，旨在处理Wikidata中的非结构性数据难题，该图谱覆盖了7 000万个实体、1 000个知识链接、4.8亿个知识三元组。在国内，比较著名的知识图谱有“知心”百度、“知立方”搜狗等，其特征在于，能将一些与用户搜索有关的信息呈现出来，从而降低用户的搜索数量，提高工作效率。我国部分科研院所也在此方面取得了一定成果，如清华大学的XLore55、上海交通大学的Zhishi.me等。

随着互联网的发展，对编组站控制系统的安全性和可靠性要求越来越高。因此，研究编组站控制系统

收稿日期：2024-04-27　**修订日期**：2025-08-29
基金项目：国家自然科学基金项目(61763025)

安全应急处置方法，不仅可以提升我国相关行业在这一领域的技术水平，也能够保障我国铁路系统的安全运行。知识图谱能够为编组站控制系统中多源异构的应急处理方法信息资源的集成提供有力手段，能够更便捷地利用与突发事件有关的信息资源，协助管理者在系统发生故障时迅速作出决策，更好地完成编组站控制系统故障应急处置工作。

1　编组站控制系统故障应急处置知识图谱构建

1.1　知识图谱概述

知识图谱是显示一个知识发展的进程和特征的各种图，并通过可视化方法，将这一串由图所表达的知识和知识的本体，以及它们与知识属性的关系展现出来。知识图谱实质上是一个由3个知识元构成的网格知识库，目前知识三元组的形式有2种，分别是＜实体1，关系，实体2＞和＜实体1，属性1，属性值＞[5-7]。

1.2　知识图谱构建流程

构建知识库常用的方式有自动搭建、半自动搭建和手工搭建。针对编组站控制系统知识图谱对知识库的需求，采用基于半自动搭建和手工搭建相结合的知识图谱构建方法，将经典的正规匹配方法与高级自然语言处理方法相结合，使已有的知识图谱更加精确和完整。知识图谱构建流程如图1所示。针对编组站控制系统故障应急处置的知识图谱，以“自顶向下”与“自底向上”相结合的方法，对编组站控制系统故障应急处置过程中的重要问题，提出了一种新的研究思路。

1.3　数据获取及处理

1.3.1　故障类型框架获取

明确编组站控制系统故障应急处置知识图谱的设计需求后，结合编组站控制系统领域知识的研究，归纳和整理编组站控制系统故障案例等非结构化和半结构化数据，总结编组站控制系统的故障类型。在确定故障类型的基础上，精确地描述编组站控制系统的领域知识体系和知识点间的关联关系，从而形成精确、完备的编组站控制系统的故障类型集合。以CIPS常见的故障类型为例，CIPS故障框架如图2所示。

1.3.2　实体及属性抽取

实体及属性抽取是指从编组站控制系统故障分析数据中提取代表实体或属性的词汇，并对其进行词性标记，通过对属性及属性值的提取，完成实体类型的构建[8]。在知识图谱的构建中，实体和关系的自动化抽取是关键的一步，因为它们构成了知识图谱的核心结构。

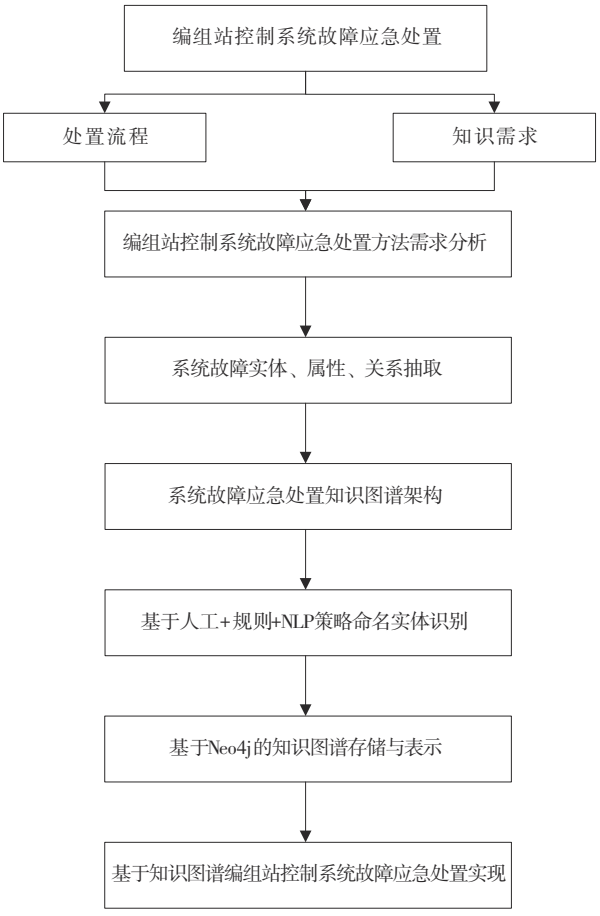


图1　知识图谱构建流程

Fig.1　Knowledge graph construction process

本研究采用一种基于正则表达式和自然语言处理的技术，从特定格式的.docx文件中抽取实体和关系。通过识别“处理”关键字的出现，能够区分“故障”和“步骤”这2种不同的实体类型。这种方法利用了中文语境中的语言结构，比传统的关键字匹配方法更具灵活性。

编组站控制系统故障知识图谱是用图模型来表达的，其实体抽取所产生的就是节点，而属性抽取指的就是构建节点与关系的属性[9]。在对实体或属性进行定义时，应考虑在冗余程度最低的条件下，满足应用及可视化呈现。编组站控制系统故障知识图谱实体的属性可归纳为名称、类型、概况、分析、方法抽取等，实体及属性抽取如表1所示。

1.3.3　关系抽取

关系抽取的目的是从编组站控制系统故障分析语料中，抽取出实体之间的语义关系，对各实体间的关联进行说明，对于用户来说，这是一个必不可少的步骤，可以进行查询和视觉显示。运用本体和属性的词类，定义编组站控制系统中的故障知识库，关系抽取

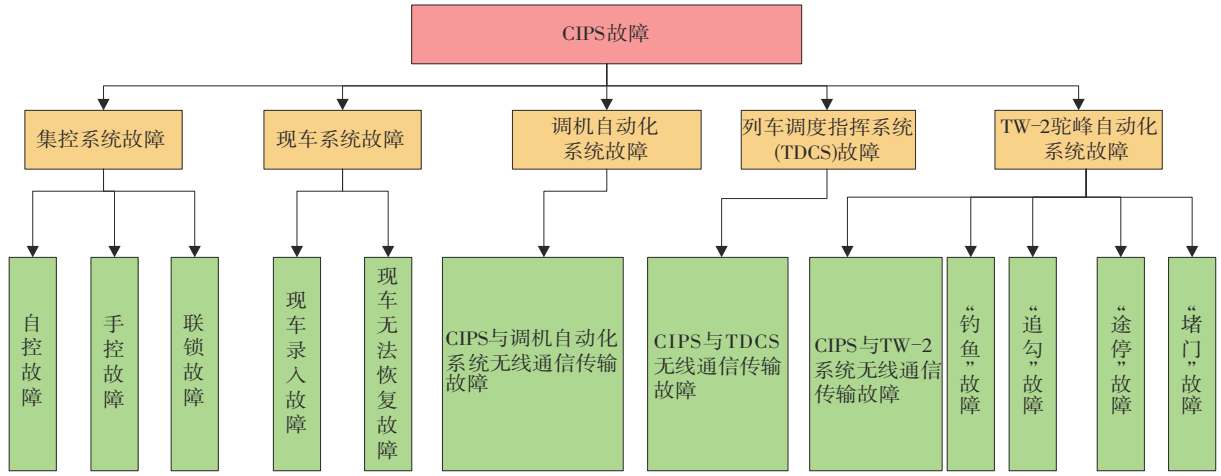


图2　CIPS故障框架

Fig.2　CIPS fault framework

表1　实体及属性抽取

Tab.1　Entity and attribute extraction

实体序号	实体名	属性	实体类型
1	自控故障	故障名称	故障
2	CIPS集控系统无法实现接发列车、调车进路自动排列	故障概况	故障概况
3	登记汇报：总值班员应及时登记运统46，通知电务部门处理，并按安全信息管理办法进行汇报，通知车间管理人员上岗	应急处置流程	应急处置流程1
4	对应场的车站值班员可在集控界面点击“转集控/自控”按钮，将系统切换为手控模式，在集控界面点击始终端按钮，人工排列列车、调车信号	应急处置流程	应急处置流程2

如表2所示。

1.3.4　知识存储

本研究以Neo4j Graph Database(Neo4j)为基础，为编组站控制系统故障应急处理提供知识库。Neo4j属性图模型由节点和关系2部分构成。本研究通过定义包含.docx文件的目录路径和初始化最终的DataFrame，用于存储所有的实体和关系。

对于目录中的每一个.docx文件，提取里面的表格数据，并根据其内容生成实体和关系。打开.docx文件并提取表格数据，根据文件内容，特别是表格的内容，生成实体和关系，使用pandas库的DataFrame数据结构存储这些实体和关系。

处理完所有的文件之后，将所有的实体和关系列表合并成2个大的DataFrame，然后将其保存到.xlsx文件中。

2　基于知识图谱的编组站控制系统故障应急处置展示及应用

2.1　知识图谱数据展示

以所收集到的编组站控制系统故障应急处置分析

表2　关系抽取

Tab.2　Relationship extraction

实体1	实体2	关系
集控系统故障	自控故障	实体1包含实体2
集控系统故障	手控故障	实体1包含实体2
自控故障	CIPS集控系统无法实现接发列车、调车进路自动排列	实体1的概况是实体2
自控故障	在集控界面点击“转集控/自控”按钮，将系统切换为手控模式	实体1的应急处置流程是实体2

语料为基础，通过对实体、关系等知识的抽取和知识存储，构建编组站控制系统故障应急处置知识图谱。图数据库中已创建的实体节点总数为151个，包括集控系统故障类节点13个、调机自动化故障类节点5个、TDCS故障类节点7个、现车系统故障类节点40个、TW-2驼峰自动化系统故障类节点86个，此外，还建立了151个关系属性。

利用Neo4j图形库，通过映射节点、关联边等方式，对编组站控制系统中的事故与紧急处理知识进行描述，编组站控制系统故障应急处置部分节点及关系如图3所示。图中的节点包括了各类系统故障的名称、故障概况、故障应急处置流程节点，每一种节点都携带着自己的属性信息，使用不同的颜色来区别不同的属性。其中，系统名称母节点是绿色的，故障名称类节点是粉红色的，故障概况类节点是蓝色的，故障应急处置流程类节点是黄色的，每一条有向边都含有开始节点、关联属性和终结点。

利用精确的知识抽取和合理的节点关系属性定义，建立编组站控制系统故障应急处置知识图谱。Neo4j图数据库的存储结构将编组站控制系统故障应急处置知识图谱中的所有实体和之间的关系都清楚地呈现出来，将不同的实体按不同的属性进行归类和存储，便于对不同的实体进行维护。同时，本研究还将为知识图谱的内容拓展提供便利，当出现新的数据集时，仅需要在图库中添加与之对应的实体和关系属性，保持数据库的整体存储框架不变，方便知识的更新。

2.2　Cypher语句查询

2.2.1　声明性语言

本研究所用的查询语言Cypher，是用于Neo4j的图数据库查询语言，具有查询效率高、表达能力强等优点。Cypher是一种声明性语言，这是其最主要的特点。与命令式语言中需要用户告诉数据库如何对数据进行处理和检索不同，声明性查询可以对用户想要查看的模式进行声明，并让数据库来查询和处理数据。

2.2.2　数据处理实现

Cypher语言使用户可以更简单地操作图数据库，不需要编写复杂的查询语句。在Neo4j图形库中，通过与节点及关系的模式匹配，完成数据的查询、更新等操作。

（1）查询编组站控制系统故障类实体节点。Cypher声明是一种特殊的语言，用来对图数据库进行查询和操纵，与SQL声明很相似，但只在图数据库中使用。在实践中，使用Cypher声明进行实体、关系和属性的查询。例如，可以使用MATCH关键字匹配符合条件的节点和关系，使用WHERE关键字添加约束条件，使用RETURN关键字返回查询结果。此外，Cypher语句还支持对查询结果进行排序、分页显示等操作，使用户能够更加灵活地处理知识图谱数据。

查询编组站控制系统故障类实体节点的步骤如下。①获取所有故障：通过定义一个函数，从Neo4j中查询所有类型为Fault的节点，并返回名称列表。②词频-逆文本频率指数(TF-IDF)向量化：通过使用TF-IDF对文本进行向量化，能够有效地捕捉文本中每个单词的重要性。③余弦相似度匹配：输入文字和全部Fault文字的余弦相似性，找出最符合输入的Fault。④主查询函数：通过定义一个函数整合上述功能，根据输入的文字，首先找到最相似的故障，然后查询该故障的解决步骤。⑤可视化故障：通过使用NetworkX库创建一个有向图来表示故障，并利用Matplotlib对其进行可视化。该方案的创新之处是采用TF-IDF和余弦相似

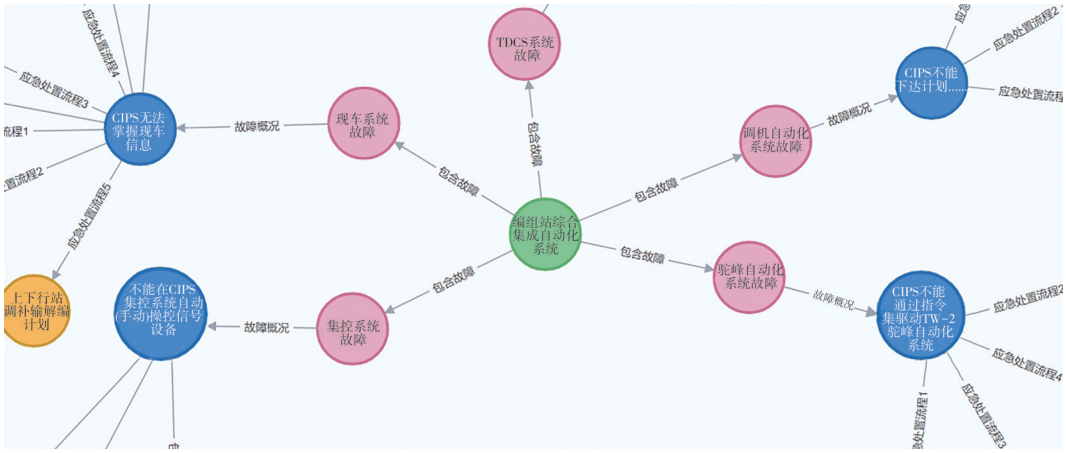


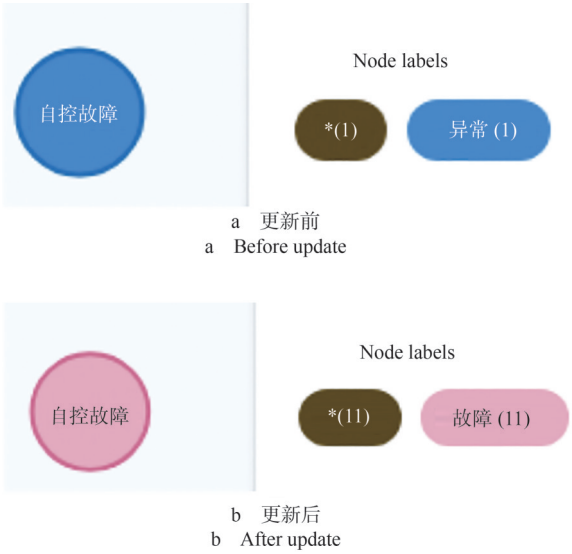
图3　编组站控制系统故障应急处置部分节点及关系

Fig.3　Some nodes and relationships for emergency response to faults in marshalling station control system

性，实现文本的向量化，并利用余弦相似度进行匹配，该方法能够非常高效地找到与输入文本最相似的故障。与传统的完全匹配或模糊匹配相比，这种方法在处理文本时更为灵活，能够更好地捕捉到文本的语义信息。该方案提供了一种基于文本相似度匹配的方法来查询知识图谱中的故障和解决步骤。

（2）更新编组站控制系统故障类实体节点。除了查询之外，Cypher 语句还可以用来更新图中的数据。在进行知识更新时，首先利用 MATCH 匹配图样，再利用 SET 对其属性进行设置，再根据节点的 ID 属性对其进行查找、修改和删除。

例如，“自控故障”的节点标记定义是错误的，更新完成后，根据节点 ID 查询节点，检验更新是否生效。节点属性更新结果如图 4 所示，更新前节点标签 Labels 为“异常”，更新后为“故障”，节点颜色也相应地发生变化。



a　更新前
a　Before update

b　更新后
b　After update

图 4　节点属性更新结果

Fig.4　Node attribute update result

2.3　应用分析

编组站控制系统故障应急处置知识图谱能够依据现场出现的故障状态，实现可视化的语义检索，为缺乏经验的人员提供辅助决策，也有助于编组站现场管理者和操作人员对编组站控制系统故障应急处理知识的有效获取和系统学习。

以武汉北编组站的 CIPS 现车系统出现故障为例，编组站现车系统操作人员(调度车间的车号员、车站调度员、总调度员、值班站长)，在“知识图谱”中查找到与这个故障有关的实体节点和关系，检索到该情况下的应急处置流程，先登记汇报，然后由值班站长通知涉及现车系统使用的管理人员(技术科和调度车间的分管调度、车号、分析管理人员)上岗盯控，根据故障时间来恢复 10，20，30，60 min 前的现车数据，现车恢复后，总值班员通知设备单位恢复设备故障前最近的现车信息，值班站长通知车号员补输到达和出发编组信息，通知车站调度员补输解体及编组调车作业计划并组织现场作业人员(编组场和交换场的为连结员，到达场和出发场的为货检人员)核对现车。现车系统故障应急处置节点及关系如图 5 所示。应急处置流程：登记运统 46，通知汇报相关人员上岗如图 6 所示。

3　结束语

目前关于编组站控制系统故障紧急处理过程的大多数知识都是分散在铁路法规、系统说明书和专家经验等零星的知识库中，不利于编组站控制系统故障紧急处置，也不便于现场作业人员进行控制系统学习培训。知识图谱具有丰富的知识管理和语义整合能

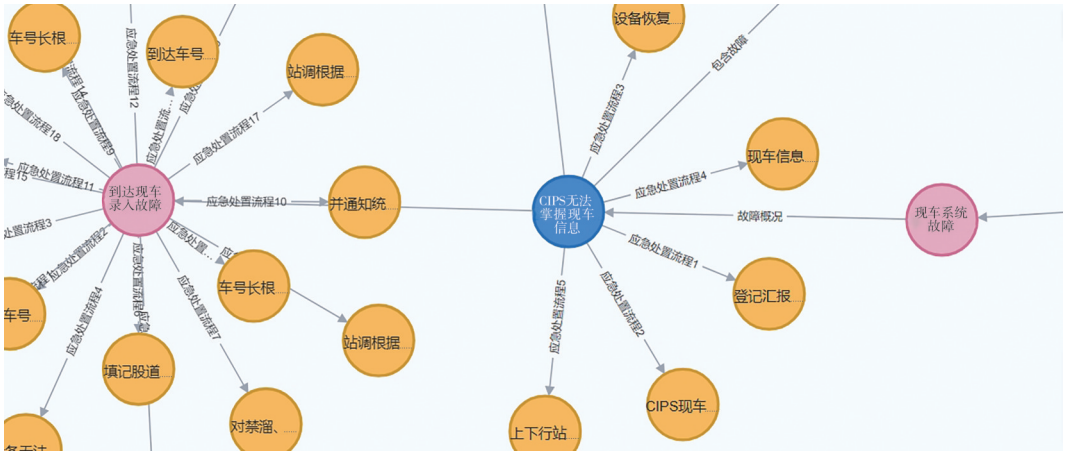


图 5　现车系统故障应急处置节点及关系

Fig.5　Emergency response nodes and relationships for current vehicle system faults

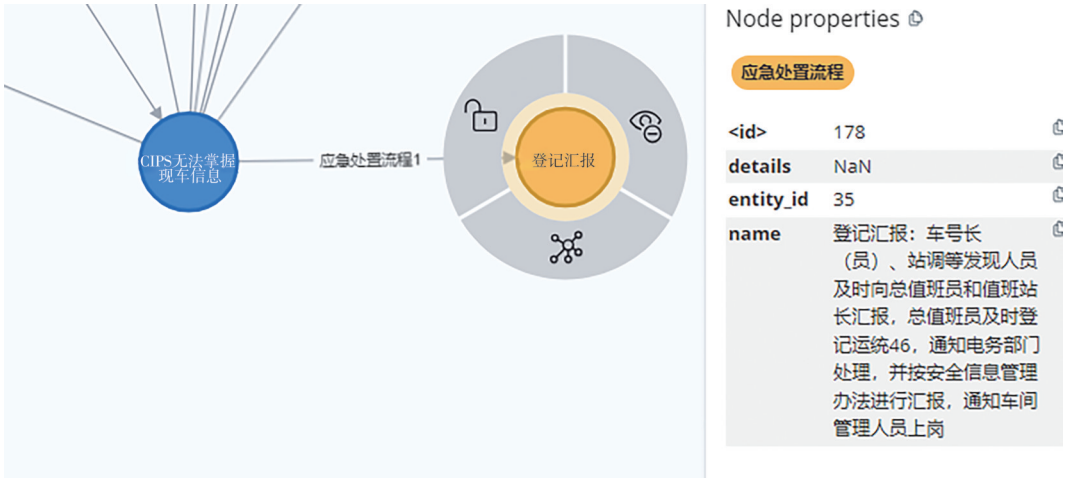


图6　应急处置流程：登记运统46，通知汇报相关人员上岗

Fig.6　Emergency response process: register with Yuntong 46 and notify relevant personnel to take up their posts

力[10-11]，利用其对铁路编组站控制系统中的突发事件进行综合管理，研究知识图谱与编组站控制系统安全应急处置方法，有助于提高编组站控制系统故障情况下的应急处理效率。在未来的研究中，可以进一步完善知识图谱的构建方法，提高知识获取的效率，以期实现编组站控制系统故障处置更加智能和高效。

参考文献：

[1] 段蕴桔. 编组站货运调车协同监控系统设计与应用[J]. 铁道货运，2024，42(1)：37-43.
DUAN Yunjie. Design and Application of Collaborative Monitoring System for Freight Shunting in Marshalling Stations[J]. Railway Freight Transport，2024，42(1)：37-43.

[2] 郭凯生，何宇浩，刘　学，等. 畅通宁西线枢纽节点南阳西站、信阳北站的策略[J]. 铁道货运，2024，42(9)：6-12.
GUO Kaisheng，HE Yuhao，LIU Xue，et al. Strategy for Smooth Transportation of Ningxi Line Hub Nodes Nanyangxi Station and Xinyangbei Station[J]. Railway Freight Transport，2024，42(9)：6-12.

[3] 王志维. 铁路编组站综合集成自动化系统方案研究[J]. 交通建设与管理，2024(S1)：200-202.

[4] 董云逸. 编组站综合自动化系统可靠性建模与评估[D]. 北京：中国铁道科学研究院，2020.

[5] CHEN X L，XIE H R，LI Z X，et al. Topic Analysis and Development in Knowledge Graph Research：A Bibliometric Review on Three Decades[J]. Neurocomputing，2021，461：497-515.

[6] 庄传志，靳小龙，朱伟建，等. 基于深度学习的关系抽取研究综述[J]. 中文信息学报，2019，33(12)：1-18.
ZHUANG Chuanzhi，JIN Xiaolong，ZHU Weijian，et al. Deep Learning Based Relation Extraction：A Survey[J]. Journal of Chinese Information Processing，2019，33(12)：1-18.

[7] 张云中，郭　冬，王亚鸽，等. 基于知识图谱的红色历史人物知识问答服务框架研究[J]. 图书情报工作，2021，65(16)：108-117.
ZHANG Yunzhong，GUO Dong，WANG Yage，et al. Framework of Knowledge Q & A Service for Red Historical Figures Based on Knowledge Graph[J]. Library and Information Service，2021，65(16)：108-117.

[8] 张吉祥，张祥森，武长旭，等. 知识图谱构建技术综述[J]. 计算机工程，2022，48(3)：23-37.
ZHANG Jixiang，ZHANG Xiangsen，WU Changxu，et al. Survey of Knowledge Graph Construction Techniques[J]. Computer Engineering，2022，48(3)：23-37.

[9] 康　雁，李　涛，李　浩，等. 融合知识图谱与协同过滤的推荐模型[J]. 计算机工程，2020，46(12)：73-79，87.
KANG Yan，LI Tao，LI Hao，et al. Recommendation Model Fusing with Knowledge Graph and Collaborative Filtering[J]. Computer Engineering，2020，46(12)：73-79，87.

[10] 孔庆玮，苗　茁，冯小芳，等. 基于知识图谱的高速铁路客运调度命令闭环管理方案研究[J]. 铁道运输与经济，2023，45(10)：89-98.
KONG Qingwei，MIAO Zhuo，FENG Xiaofang，et al. Research on Closed-Loop Management Scheme of High Speed Railway Dispatching Command Based on Knowledge Graph[J]. Railway Transport and Economy，2023，45(10)：89-98.

[11] 姚　剑，白　伟，杨国元，等. 铁路客站设备健康管理知识图谱构建与应用研究[J]. 铁道运输与经济，2023，45(3)：94-102.
YAO Jian，BAI Wei，YANG Guoyuan，et al. Research on the Construction and Application of Knowledge Graph of Equipment Health Management of Railway Passenger Stations[J]. Railway Transport and Economy，2023，45(3)：94-102.

（责任编辑：李丹丹）

（下转第64页）