

文章编号:1671-251X(2021)03-0041-05

DOI:10.13272/j.issn.1671-251x.2020090013

煤矿装备维护知识图谱构建及应用

曹现刚^{1,2}, 张梦园^{1,2}, 雷卓^{1,2}, 段欣宇^{1,2}, 陈瑞昊^{1,2}

(1. 西安科技大学 机械工程学院, 陕西 西安 710054;

2. 陕西省矿山机电装备智能监测重点实验室, 陕西 西安 710054)



扫码移动阅读

摘要:利用大数据管理系统对煤矿装备维护信息进行管理缺乏对煤矿装备维护知识的表示能力,没有形成相对完整的煤矿装备维护知识管理体系,无法实现知识挖掘及知识间关系链接,导致大量具有深度挖掘价值的信息不能得到有效利用。针对上述问题,构建了煤矿装备维护知识图谱。首先通过定义煤矿装备维护的关键概念、关系及属性,进行基于本体的知识建模;然后从结构化、半结构化和非结构化数据源中获取知识,通过命名实体识别、关系抽取及事件抽取完成煤矿装备维护知识抽取;最后利用图数据库 Neo4j 存储煤矿装备维护知识,形成煤矿装备维护知识图谱。煤矿装备维护知识图谱在智能语义搜索、智能问答及可视化决策支持等方面的应用可提高煤矿装备维护知识管理效率,为煤矿装备智能化动态管理的实现提供有力支持。

关键词:煤矿装备维护;知识图谱;知识建模;知识抽取;知识存储

中图分类号:TD421

文献标志码:A

Construction and application of knowledge graph for coal mine equipment maintenance

CAO Xiangang^{1,2}, ZHANG Mengyuan^{1,2}, LEI Zhuo^{1,2}, DUAN Xinyu^{1,2}, CHEN Ruihao^{1,2}

(1. College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China;

2. Shaanxi Key Laboratory of Mine Electromechanical Equipment Intelligent Monitoring, Xi'an 710054, China)

Abstract: Using big data management system to manage coal mine equipment maintenance information lacks the ability to express coal mine equipment maintenance knowledge, has not formed a relatively complete coal mine equipment maintenance knowledge management system, and cannot realize knowledge mining and inter-knowledge relationship links, resulting in a large number of information with in-depth mining value cannot be used effectively. In order to solve the above problems, a coal mine equipment maintenance knowledge graph is constructed. Firstly, ontology-based knowledge modeling is carried out by defining key concepts, relationships and attributes of coal mine equipment maintenance. Secondly, obtain knowledge from structured, semi-structured and unstructured data sources, and coal mine equipment maintenance knowledge extraction is completed through named entity identification, relationship extraction and event extraction. Finally, the graph database Neo4j is used to store coal mine equipment maintenance knowledge and form a coal mine equipment maintenance knowledge graph. The application of coal mine equipment maintenance knowledge graph in intelligent semantic search, intelligent question and answer and visualization decision support can improve the efficiency of coal mine equipment maintenance knowledge management and provide strong support for the realization of intelligent dynamic

收稿日期:2020-09-07;修回日期:2021-03-09;责任编辑:盛男。

基金项目:国家自然科学基金重点资助项目(51834006);国家自然科学基金资助项目(51875451)。

作者简介:曹现刚(1970—),男,山东莒南人,教授,博士,主要研究方向为设备状态监测与故障诊断技术、现代设备维护理论与技术,E-mail:582060393@qq.com。

引用格式:曹现刚,张梦园,雷卓,等.煤矿装备维护知识图谱构建及应用[J].工矿自动化,2021,47(3):41-45.

CAO Xiangang, ZHANG Mengyuan, LEI Zhuo, et al. Construction and application of knowledge graph for coal mine equipment maintenance[J]. Industry and Mine Automation, 2021, 47(3): 41-45.

management of coal mine equipment.

Key words: coal mine equipment maintenance; knowledge graph; knowledge modeling; knowledge extraction; knowledge storage

0 引言

煤矿装备维护是基于装备类型、维护方案、工作环境等多种因素综合作用的结果,其覆盖的各类相关信息具有复杂、分散等特点,存在共享度低及缺乏有效管理等问题,制约了煤矿智能化发展^[1]。近年来,许多研究通过建立大数据管理系统以实现煤矿装备维护信息的管理。曹现刚等^[2]提出了一种基于 Hadoop 的煤矿机电设备运行状态大数据管理平台,以实现煤矿机电设备信息的高效管理。谭章禄等^[3]构建了煤炭大数据平台管理协同构架,旨在实现煤炭大数据全生命周期信息管理。程曜安等^[4]提出了一种面向大型装备的 MRO (Maintenance, Repair and Operations, 维护、维修和大修) 支持系统,实现了以中性物料清单为核心的维修知识管理。大数据管理系统的应用从一定程度上提高了煤矿装备维护信息的管理效率,但缺乏对煤矿装备维护知识的表示能力,没有形成相对完整的煤矿装备维护知识管理体系,无法实现知识挖掘及知识间关系链接,导致大量具有深度挖掘价值的信息不能得到有效利用。

知识图谱是一种自带语义、蕴含逻辑含义与规则的特殊图数据,具有人类可识别、对机器友好的特点,运用知识图谱能够提升网络环境下知识互联和知识共享的效率^[5-6]。因此,本文构建了煤矿装备维护知识图谱,可将煤矿装备维护信息聚类为具有利用价值的知识。

1 煤矿装备维护知识图谱构建流程

2012 年,Google 提出了知识图谱的构建原则,即通过概念、实体和关系形成的结构化语义关系网络,以节点关系图的形式来进行存储^[7-8]。知识图谱构建主要有自顶向下和自底向上 2 种方式^[9],可衍生为专家法、参照法、归纳法及混合法。专家法是指基于相关领域专家知识,利用自顶向下的构建方式进行总体规划;参照法是指根据已有的行业标准进行标杆对照与业务适配;归纳法是指以自底向上构建方式为主,进行单点切入及模式抽取;混合法是将自顶向下和自底向上 2 种构建方式结合。由于煤矿装备维护没有统一的行业标准来进行标杆对照,且相关领域的复合型专家不足,不适合使用参照法与专家法,而归纳法的构建方式不够灵活,所以选择混

合法作为煤矿装备维护知识图谱构建的主要方法。煤矿装备维护知识图谱构建流程如图 1 所示。首先通过定义知识图谱概念、属性及关系模式进行基于本体的煤矿装备维护知识建模;然后从数据源中获取知识,通过命名实体识别、关系抽取及事件抽取,实现煤矿装备维护知识抽取;最后选择置信度高的知识存储至 Neo4j 图数据库,完成煤矿装备维护知识存储。

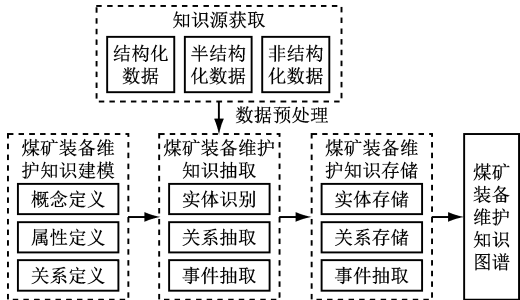


图 1 煤矿装备维护知识图谱构建流程
Fig. 1 Construction process of knowledge graph of coal mine equipment maintenance

2 煤矿装备维护知识图谱构建关键技术

2.1 煤矿装备维护知识建模

为更好地描述知识本身与知识之间的关联,在知识建模过程中,需要选择合适的知识表示方法。传统的知识表示方法在复杂知识推理过程中存在组合爆炸的可能性,因此将本体的概念^[10]引入知识模型中,以解决传统知识表示方法的局限性^[11]。

根据本体的语义清晰度、概念一致性、可扩展性、可重复利用性的构造原则来构建煤矿装备维护本体^[12],主要步骤:① 确定知识本体的范围与目标。基于煤矿装备维护的研究需求,以煤矿装备为研究对象,旨在解决其信息共享、知识重用及本体扩展的问题,实现煤矿装备维护知识的智能化管理。② 本体分析。主要任务包括总结重要术语、定义概念类及提取关系属性。根据煤矿装备维护知识特点,利用本体构建工具 Protégé^[13]构建煤矿装备维护本体。根据煤矿装备维护的特征,将煤矿装备维护知识的核心概念分为设备名称、设备状态、维护方案、故障维修、工作人员、维护案例 6 个类别,如图 2 所示。煤矿装备维护本体的类属性包括对象属性与数据属性。对象属性主要表示对象之间的关联,一般用来描述类的不可量化特征;数据属性主要表示对象与数值之间的关联,一般用来描述类的某些可

量化特征。③ 本体构建。OWL (Web Ontology Language,网络本体语言)作为国际通用本体语义

描述语言之一,具有强大的本体推理能力,因此本文选择 OWL 作为煤矿装备维护本体的描述语言。

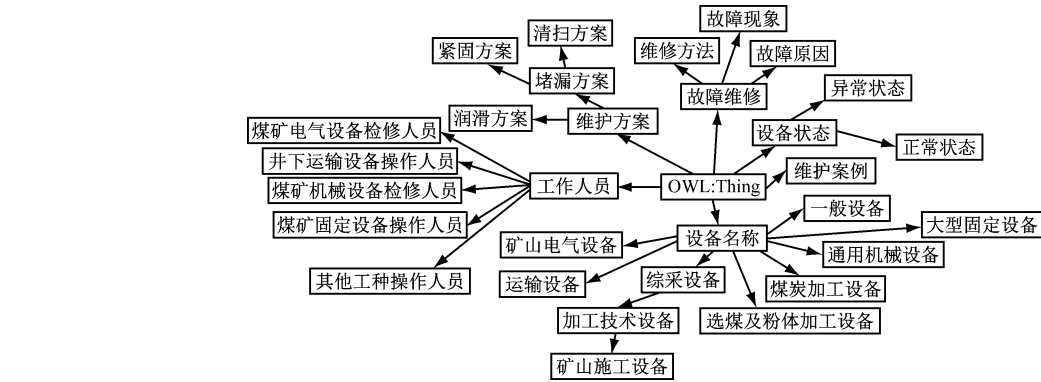


图 2 煤矿装备维护部分概念本体

Fig. 2 Some concept ontology of coal mine equipment maintenance

2.2 煤矿装备维护知识抽取

知识抽取是指从海量的多源异构数据中抽取特定的知识^[14]。知识抽取的数据源类型多样,包括结构化数据、半结构化数据、非结构化数据,针对不同类型的数

据源,知识抽取的方法不同。结构化数据具有良好布局结构,一般存储于关系型数据库中。利用 Ultrawrap 直接实现关系型数据库到 RDF (Resource Description Framework,资源描述框架)或 OWL 的映射,以完成对结构化数据的知识抽取。

半结构化数据是指在一定程度上具有某种特征的数据,例如网页数据。通过网络爬虫实现对煤矿装备维护半结构化数据的知识抽取,具体实现方法有手工方法、归纳方法及自动抽取方法。手工方法通过构建适合网页数据的抽取规则,在人工分析的基础上,编写适用于当前网页的抽取表达式,抽取目标明确、速度快,但成本高且置换性不强;归纳方法通过事先从已标注的数据训练集中学习抽取规则,再对其他具有相同模板的网页数据进行知识抽取,但标注数据的质量会影响知识抽取的结果;自动抽取方法通过自主挖掘网页中的相似规律进行无监督学习,可免于人工标注,但需要对知识抽取的内容进行降噪处理。

非结构化数据是指符合自然语言规范的文本数据,例如煤矿装备维修工单、专家知识经验、维修手册等,煤矿装备维护知识主要来源于非结构化数据。面向非结构化数据的知识抽取主要步骤:通过网络爬虫及实验室资料收集等方式获取煤矿装备维护相关文本语料并进行预处理;利用 HanLP 等工具对预处理后的语料进行分词、标注与词向量转换;利用 TensorFlow 工具训练 Lattice-LSTM (Lattice-Long Short-Term Memory,网格结构-长短期记忆网络)模型^[15-16],采用Lattice-LSTM 模型从文本中抽取

实体的信息元素,实现命名实体识别;采用基于弱监督学习的 Bootstrapping 方法从文本中抽取 2 个或多个实体之间的语义关系,完成关系抽取;利用基于深度学习的事件联合抽取方法从文本中抽取需要的事件信息,并以结构化方式呈现,实现事件抽取。

2.3 煤矿装备维护知识存储

知识图谱通常采用关系型数据库、RDF 三元组及图数据库来实现知识存储^[17]。关系型数据库在每一次处理过程中都需添加新的表与字段,增加了开发难度的同时降低了系统稳定性;利用 RDF 进行语义描述时不仅需要提前制订满足特定领域所需的 RDF 规则词汇表,还需要使用 RDF 的特定工具才能完成相应操作,不够简单、灵活。而图数据库只需插入节点与边即可实现数据的高效存储与查询^[18],因此本文使用图数据库 Neo4j 来实现煤矿装备维护知识存储。Neo4j 的标签表示煤矿装备维护知识的概念,节点和节点属性分别表示实体与实体属性,边和边属性分别表示实体间的关系和关系属性。基于 Neo4j 的知识存储方案见表 1。

表 1 基于 Neo4j 的知识存储方案

Table 1 Knowledge storage scheme based on Neo4j		
类别	作用	对象范围
标签	描述知识概念类	设备名称、设备状态、维护方案、故障维修、工作人员、维护案例等
节点	描述知识实体	一般设备、综采设备、运输设备等
边	描述实体关系	等价关系、跟随关系、等级关系等

利用Neo4j专属的声明式查询语言——Cypher 语言,实现对图数据库中知识节点与关系的增删改查等操作,具体操作方法:使用 CREATE 语句创建煤矿装备维护实体节点;使用 MATCH 语句更新或查询实体节点或关系;使用 WHERE 语句设置查询过程中的匹配条件等。

3 煤矿装备维护知识图谱应用

煤矿装备维护知识图谱可实现智能语义搜索、智能问答、可视化决策支持等应用,有利于用户高效查询和学习煤矿装备维护概念、维修方案等知识,并可推理潜在的知识,进而增进知识共享,提高煤矿智能化背景下的煤矿装备维护水平。

(1) 智能语义搜索。当前基于关键词的传统搜索在知识图谱的知识支持下可上升到基于实体和关系的搜索,称之为语义搜索^[19]。语义搜索利用知识图谱的实体链接,解决了传统搜索中关键词语义消歧的难题。煤矿装备维护人员借助语义检索可得到准确性高、关联度强的查询结果,有利于提高工作效率,保障煤矿装备正常运行。

(2) 智能问答。问答系统是一种能够让计算机对用户提出的问题进行自动回复的高级形式信息服务模式^[20]。区别于现有的搜索引擎,问答系统以精准的自然语言搜索答案取代与关键词相关度高的文档并返回给用户。智能问答可针对煤矿装备维护人员输入的自然语言进行理解,从知识图谱或目标数据中给出问题的答案,并对返回答案进行评分评定以确定优先级顺序,实现了煤矿装备维护人员依据经验开展维护工作到依据数据开展维护工作的转变,有利于高效搜索的实现。

(3) 可视化决策支持。通过统一的图形接口,结合可视化、推理、检索等为用户提供信息获取的入口被称之为可视化决策支持。例如,决策支持可以通过图谱可视化技术对煤矿装备维护知识图谱中的采煤机故障部件、故障部件运行参数等信息进行解读,有利于辅助煤矿装备维护人员实行最佳决策,提高维护工作效率与能力。

4 结语

从知识建模、知识抽取及知识存储 3 个方面描述了煤矿装备维护知识图谱构建流程。首先通过定义概念、属性、关系进行基于本体的煤矿装备维护知识建模;然后从结构化、半结构化和非结构化数据源中获取知识,通过命名实体识别、关系抽取和事件抽取完成煤矿装备维护知识抽取;最后基于图数据库 Neo4j 实现煤矿装备维护知识存储,形成煤矿装备维护知识图谱。煤矿装备维护知识图谱可实现智能语义搜索、智能问答及可视化决策支持等应用,能有效提高煤矿装备维护知识利用率,为煤矿装备智能化动态管理的实现提供有利支持。

参考文献 (References):

[1] 王国法,刘峰,孟祥军,等. 煤矿智能化(初级阶段)研

究与实践[J]. 煤炭科学技术,2019,47(8):1-36.
WANG Guofa, LIU Feng, MENG Xiangjun, et al. Research and practice on intelligent coal mine construction(primary stage)[J]. Coal Science and Technology,2019,47(8):1-36.
[2] 曹现刚,罗璇,张鑫媛,等. 煤矿机电设备运行状态大数据管理平台设计[J]. 煤炭工程,2020,52(2):22-26.
CAO Xiangang, LUO Xuan, ZHANG Xinyuan, et al. Design of big data management platform for operation status of coal mine electromechanical equipment[J]. Coal Engineering,2020,52(2):22-26.
[3] 谭章禄,马营营,袁慧. 煤炭大数据平台建设的关键技术及管理协同架构[J]. 工矿自动化,2018,44(6):16-20.
TAN Zhanglu, MA Yingying, YUAN Hui. Key technologies and management collaborative architecture of construction of coal big data platform[J]. Industry and Mine Automation, 2018, 44(6):16-20.
[4] 程曜安,张力,刘英博,等. 大型复杂装备 MRO 系统解决方案[J]. 计算机集成制造系统,2010,16(10):2026-2037.
CHENG Yaoan, ZHANG Li, LIU Yingbo, et al. Solution of MRO support system for large complex equipment[J]. Computer Integrated Manufacturing Systems,2010,16(10):2026-2037.
[5] NGUYEN H L, VU D T, JUNG J J. Knowledge graph fusion for smart systems: a survey [J]. Information Fusion,2020,61:56-70.
[6] 王鑫,邹磊,王朝坤,等. 知识图谱数据管理研究综述[J]. 软件学报,2019,30(7):2139-2174.
WANG Xin, ZOU Lei, WANG Chaokun, et al. Research on knowledge graph data management; a survey [J]. Journal of Software, 2019, 30(7):2139-2174.
[7] DUAN Yucong, SHAO Lixu, HU Gongzhu, et al. Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph [C]//IEEE 15th International Conference on Software Engineering Research, Management and Applications, London, 2017:327-332.
[8] 刘峤,李杨,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.
LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques [J]. Journal of Computer Research and Development, 2016, 53(3):582-600.
[9] COLLARANA D, GALKIN M, TRAVERSO-RIBON I, et al. Semantic data integration for knowledge graph

construction at query time [C]//IEEE 11th International Conference on Semantic Computing, San Diego, 2017:109-116.

[10] 施昭,曾鹏,于海斌. 基于本体的制造知识建模方法及其应用[J]. 计算机集成制造系统, 2018, 24(11): 2653-2664.

SHI Zhao, ZENG Peng, YU Haibin. Ontology-based modeling method for manufacturing knowledge and its application[J]. Computer Integrated Manufacturing System, 2018, 24(11): 2653-2664.

[11] 张楠,谢国军,叶青,等. 矿山语义物联网自动语义标注方法[J]. 工矿自动化, 2020, 46(3): 27-33.

ZHANG Nan, XIE Guojun, YE Qing, et al. Automatic semantic annotation method of mine semantic web of things[J]. Industry and Mine Automation, 2020, 46(3): 27-33.

[12] DARAIO C, LENZERINI M, LEPORELLI C, et al. Data integration for research and innovation policy: an ontology-based data management approach [J]. Scientometrics, 2016, 106: 857-871.

[13] 余磊,谷宏强,孟晨,等. 基于本体的装备 PHM 知识化建模研究[J]. 火力与指挥控制, 2019, 44(10): 13-17.

YU Lei, GU Hongqiang, MENG Chen, et al. Research on ontology-based knowledge modeling of equipment PHM[J]. Fire Control & Command Control, 2019, 44(10): 13-17.

[14] 冯建周,马祥聪. 基于迁移学习的细粒度实体分类方法的研究[J]. 自动化学报, 2020, 46(8): 1759-1766.

FENG Jianzhou, MA Xiangcong. Fine-grained entity type classification based on transfer learning[J]. Acta Automatica Sinica, 2020, 46(8): 1759-1766.

[15] YE Na, QIN Xin, DONG Lili, et al. Chinese named entity recognition based on character-word vector fusion [J]. Wireless Communications and Mobile Computing, 2020(3): 1-7.

[16] ZHANG Yue, WANG Yile, YANG Jie. Lattice LSTM for Chinese sentence representation [J]. IEEE Transactions on Audio Speech and Language Processing, 2020, 28: 1506-1519.

[17] 宫法明,李翀然. 基于 Neo4j 的海量石油领域本体数据存储研究[J]. 计算机科学, 2018, 45(增刊 1): 549-554.

GONG Faming, LI Xiaoran. Research on ontology data storage of massive oil field based on Neo4j[J]. Computer Science, 2018, 45(S1): 549-554.

[18] 王红,张青青,蔡伟伟,等. 基于 Neo4j 的领域本体存储方法研究[J]. 计算机应用研究, 2017, 34(8): 2404-2407.

WANG Hong, ZHANG Qingqing, CAI Weiwei, et al. Research on storage method for domain ontology based on Neo4j [J]. Application Research of Computers, 2017, 34(8): 2404-2407.

[19] WU Qinyue, FU Duankang, SHEN Beijun, et al. Semantic service search in IT crowdsourcing platform: a knowledge graph-based approach [J]. International Journal of Software Engineering and Knowledge Engineering, 2020, 30(6): 765-783.

[20] RODRIGO A, HERRERA J, PENAS A. The effect of answer validation on the performance of question-answering systems [J]. Expert Systems with Applications, 2019, 116: 351-363.