

基于领域知识图谱的建筑工程标准数字化研究

李刚, 苏义坤, 刘均昊, 郑志哲

(东北林业大学土木与交通学院, 哈尔滨 150040)

摘要: 为推动中国建筑工程领域标准的数字化, 总结建筑工程标准和文本结构规律, 细化知识图谱构建流程, 运用人与自然语言处理技术, 构建建筑工程标准领域知识图谱。通过对标准的实际案例应用, 验证建筑工程标准领域知识图谱在结构化表达、抽取和调用相关知识方面具有良好的可用性。构建的知识图谱可以有效地对建筑工程领域标准知识进行结构化表达。

关键词: 标准数字化; 知识图谱; 标准结构化; 图数据库

中图分类号: TU18 **文献标志码:** A **文章编号:** 1671-1807(2025)13-0246-06

标准是为了在一定范围内获取最佳秩序, 经协商一致制定并由公认机构批准, 共同使用和重复使用等一种规范性文件, 是对一定范围内的重复性事物和概念所做的统一规定^[1]。工程标准规范作为建筑工程全生命周期的指导准则, 在设计、施工、验收和运维等各个环节扮演着不可或缺的角色^[2]。然而, 目前建筑工程领域的标准制定及服务尚处于较为基础的信息检索或 PDF(portable document format, 可移植文档格式)浏览阶段^[3], 受制于信息系统的限制, 标准制定过程耗时长、工作量大, 标准易出现不符合使用需要, 无法采用先进技术等情况, 并且标准之间的协调能力较差, 易出现内容交叉或知识冲突等现象^[4-6]。

信息技术的不断发展为标准制定、实施、维护的全生命周期提供全新选择。建筑工程标准数字化的总体目标是通过数字技术提高标准的可读性、可理解性、可计算性和可操作性^[7]。利用数字技术将海量的标准文献结构化、知识化, 使标准的协同研制与信息实时共享成为可能, 有效缩短标准的制定周期, 提高标准整合力度, 也为标准知识提供了更为广阔的应用场景。考虑到在促进知识的系统化表达和传播中的重要作用, 知识图谱(knowledge graph)技术已在多个领域得到广泛应用^[8-11]。可为建筑工程标准的数字化提供准确、全面、及时的知识支持^[3]。本文旨在利用知识图谱技术实现建筑工

程标准、规范文本的结构化表达, 构建建筑工程标准知识图谱, 以促进建筑工程标准知识的传播, 提供建筑工程标准编制与实施水平。

1 相关研究

标准文件的自身特点可以较好适配知识图谱结构, 实现文件内容的局部检索、语义关联、可视化^[12]。并且, 利用语义建模、人工智能、知识图谱等技术实现标准知识的抽取、分类与表达已成为未来标准数字化的发展趋势^[7]。例如, 王佳等^[13]提出了面向 BIM(building information modeling, 建筑信息模型)消防智能审查的规范语义模型构建方法, 并基于 IFC(industry foundation classes, 工业基础类)标准进行知识本体构建, 从而达到面向 BIM 模型的消防智能审查的目的。穆磊等^[14]提出了一种基于建筑信息模型和知识图谱的建筑消防智能审图方法。这部分研究工作集中在将标准部分内容结构化表达, 可以形象表示标准知识之间的关系, 但缺乏对标准文献整体结构的研究。另一部分研究如杨跃翔等^[4]通过分析标准文献的结构特征, 并通过 XML(extensible markup language, 可扩展标记语言)标准标签集拓展, 构建适用于中国标准文献结构的标准标签集, 实现标准文献机器可读和知识抽取, 进而借助知识图谱构建技术实现标准文献知识图谱构建。这部分研究着重于从标准文献整体结构出发, 对标准文献整体内容反映全面, 但反映标

收稿日期: 2025-01-04

基金项目: 国家重点研发计划(2021YFF0602002)

作者简介: 李刚(1991—), 男, 吉林梨树人, 硕士研究生, 研究方向为智能建造; 通信作者苏义坤(1972—), 男, 黑龙江哈尔滨人, 博士, 教授, 研究方向为智能建造、工程建设标准化等; 刘均昊(1999—), 男, 山东潍坊人, 硕士研究生, 研究方向为智能建造、工程建设标准化; 郑志哲(1992—), 男, 黑龙江哈尔滨人, 博士研究生, 研究方向为智慧建造、安全管理、项目管理。

准知识之间关系的能力较弱。

在知识图谱构建及知识抽取上,刘均昊等^[3]采用基于人工和自然语言处理技术的方式进行知识抽取,构建了建筑施工安全标准知识图谱。林洁如和胡最^[15]使用 BERT-BiLSTM-CRF(预训练语言模型-双向长短时记忆网络-条件随机场)深度学习模型进行命名实体识别,以从海量的文本信息中提取关键信息。林佳瑞等^[7]对基于深度学习的实体关系抽取进行研究,系统总结了近几年来基于深度学习的实体关系抽取研究进展,并对未来可能的研究方向进行了探讨和展望。

综上所述,语义模型及知识图谱技术是未来标准数字化建设的一个重要方向。尽管现有研究已将知识图谱应用到建筑工程领域,但是,基于建筑施工领域标准知识图谱的研究相对较少。同时目前基于语义模型与知识图谱的建筑工程领域标准数字化的研究没有形成在标准知识结构化表达的同时,充分反映标准文献的整体内容以及相互之间关系的成熟做法。因此,本文基于知识图谱技术实现建筑工程领域标准等规范文本的结构化表达,以揭示建筑工程领域标准中知识及标准间的内在联系,提升中国建筑工程领域标准的数字化水平。

2 建筑工程标准文本结构特征

建筑工程标准的文本结构规范、层次清晰、表达准确简洁,对建筑工程标准文本的内容及其结构进行分析是构建标准知识图谱的前提。根据建筑工程标准文本内容,标准文本可以分为纯文本数据与半结构化数据(表格、列表等)。半结构化数据语义表达清晰,结构相对固定,易于进行知识的表示。表 1 显示,建筑工程标准的纯文本数据可以根据是否易于结构化表达而分为可结构化标准条文和不可结构化标准条文两类^[16]。可结构化标准条文的特点是可以特定的表达式进行表达,从而抽取出实体、关系和属性,进而形成标准知识图谱。不可结构化标准条文多为文本描述,或是由相关领域专家给予的经验值,无法以特定的结构表达式进行表示。

表 1 建筑工程标准知识分类

分类	条文示例
可结构化标准条文	硅酸盐水泥初凝时间不小于 45 min,终凝时间不大于 300 min 本标准主要起草单位:中国建筑材料科学研究院
不可结构化标准条文	鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新条款

根据标准要素的存在状态,建筑工程标准文本可以分为必备要素和可选要素。根据《标准化工作导则 第 1 部分:标准化文件的结构和起草规则》(GB/T 1.1—2020)(下简称《标准化工作导则》)的明确规定,标准文本的构成要素有封面、目次、前言、引言、范围、规范性引用文件、术语和定义、符号和缩略语、分类和编码/系统构成、总体原则和/或总体要求、核心技术要素、其他技术要素、参考文献、索引,其中封面、前言、范围、规范性引用文件、术语和定义、核心技术要素为必选或必选/可选要素,其他要素为可选要素。必选要素和必选/可选要素可统称为共性结构要素^[4]。共性结构要素的文本内容一般有明确的规定,知识描述形式与文本结构相对固定,可以以特定的表达式进行表示,从而进行标准知识图谱的构建。

标准化工作导则中也对标准进行了分类,标准可以按照功能分为术语标准、符号标准、分类标准、实验标准、规范标准、规程标准和指南标准。不同类型的标准其核心技术要素的内容在知识表述与表达逻辑上不尽相同。不同类型的建筑工程标准其核心技术要素的具体内容见表 2。

表 2 建筑工程标准核心技术要素

标准类型	核心技术要素内容
术语标准	界定概念的指称及其含义
符号标准	界定符号的表现形式及其含义或名称
分类标准	确立分类体系
试验标准	描述实验活动以及得出结论的方式
规范标准	描述用于判定产品或服务是否满足的证实方法
规程标准	规定明确的程序并判定是否履行的证实方法
指南标准	给予指导或给出相关建议或信息的标准

3 领域知识图谱

领域知识图谱的本质是一种结构化的知识表示形式,旨在利用图结构建模、识别和推断事物之间的复杂关联关系和沉淀领域知识。相对于文本而言,结构化数据更容易被机器处理。知识图谱采用图的方式描述和表达知识,既可以表达复杂的事物间关系,也可以避免复杂的逻辑约束。知识图谱领域常用的知识表示方式有属性图、RDF(资源描述框架)图模型和 OWL 本体语言 3 种^[17],其中属性图的表达能力与推理能力适中,表达方式灵活,查询计算方便,因此本文选取属性图为知识表达方式,采用 Neo4j 进行数据保存。

基于属性图的知识图谱 K_G 的表示方式为

$$K_G = \{L, E, R, P\} \quad (1)$$

式中： L 为标签集； E 为标准知识实体集； R 为关系集； P 为属性集。其中关系集 P 可以分为实体属性和关系属性，标签集 L 可以抽象为概念集。图数据库表示形式如图 1 所示。

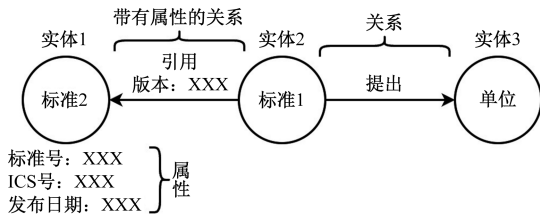


图 1 数据库示例

通过对知识图谱的描述，将以自然语言方式表达的标准文本分解为标签、实体、关系和属性，最终形成标准知识图谱。其构建流程如图 2 所示，标准知识图谱构建关键在于实体、关系和属性的识别与抽取。

4 建筑工程标准领域知识图谱构建

根据标准文本的结构分析，将标准文本分为共性关键要素和可选要素两种结构，半结构化数据与纯文本数据两种内容，针对标准文本的不同结构和内容，宜采用不同的方法进行标准知识的抽取。

经过第 2 节对标准文本的结构分析，核心技术要素为必备要素，是建筑工程标准的关键内容。然而核心技术要素的内容不尽相同。面对核心技术要素内容的特点，对标准条文结构规律与条文语义信息进行分析，是实现标准知识图谱各要素的识别与抽取的关键。

4.1 建筑工程标准条文结构规律

总结条文结构规律是进行知识抽取的基础。在王佳等^[13]对逻辑表达式研究的基础上，将标准文本短句进一步划分为实例条件型和实例关系型两

种，并扩充一种词语类型，即关系词，使其更加适合应用于标准知识图谱的知识抽取。一段标准文本可以由用逗号划分的不同短句组成，每条短句又可分割为不同数量的词语，词语是标准文本的最小组成单元，如式(2)所示。

$$[\text{词 } 1] + [\text{词 } 2] + [\text{词 } 3] + [\text{词 } 4] + \dots + [\text{词 } n] \quad (2)$$

根据对标准知识图谱 K_G 的定义以及对文本语义的分析，可以将词语分为 5 种，即实例、关系词、属性词、比较词和规定值，将句子分为两类，即实例条件型和实例关系型，如式(3)、式(4)所示。

实例条件型：

$$[\text{实例}] + [\text{属性词}] + [\text{比较词}] + [\text{规定值}] \quad (3)$$

实例关系型：

$$[\text{实例}] + [\text{比较词}] + [\text{实例}] \quad (4)$$

实例为标准条文中各种知识对象，属性词是各种知识的属性，比较词是各种属性的约束词语，如不得大于、不小于等词汇，规定值是标准条文所规定的具体参数，关系是描述实例之间的联系词汇。

结合以上分析，将两种标准短句分别定义为：实例条件型 $R_{EC} = (i, c, p, n)$ ，实例关系型 $R_{ER} = (i, r, i)$ ，其中 i 为实例， c 为条件， p 为比较词， n 为规定值， r 为关系词。以表 1 内容为例，将标准短句以逗号分隔后可将标准短句分解为式(3)、式(4)的形式，短句 1：[硅酸盐水泥] + [初凝时间] + [不小于] + [45 min]，短句 2：[终凝时间] + [不大于] + [300 min]，短句 3：[本标准] + [主要起草单位：] + [中国建筑材料科学研究院]，并按照定义进行划分，如图 3 所示，其中，紫色为实例，绿色为属性词，黄色为比较词，蓝色为规定值，棕色为关系词，短句 1 与短句 2 为实例条件型，短句 3 为实例关系型。

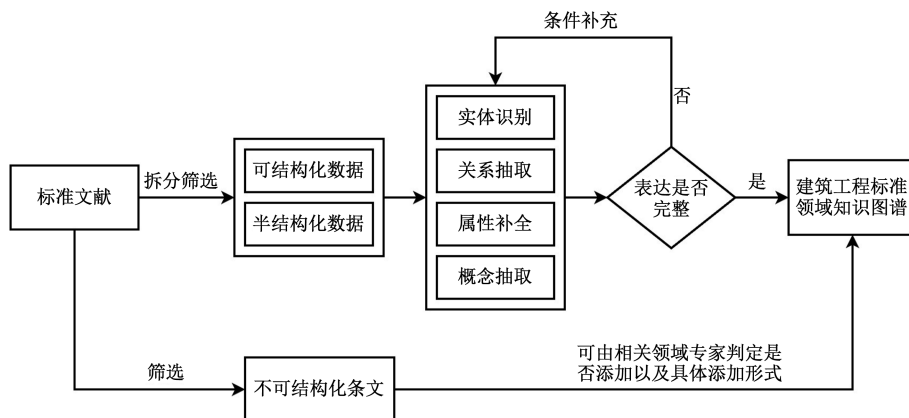


图 2 领域知识图谱构建流程

短句1: 硅酸盐水泥 初凝时间 不小于 45 min
 短句2: 终凝时间 不大于 30 min
 短句3: 本标准 主要起草单位: 中国建筑材料科学研究院

图3 标准条文分解

4.2 建筑工程标准领域知识图谱要素抽取

根据以上对标准条文结构规律的分析不难发现,对标准知识图谱 K_G 的定义与标准短句 R_{EC} 、 R_{ER} 的定义可以总结出相应的对应关系。即建筑工程标准领域知识图谱的实体、关系可分别对应标准短语的实例、关系词,而建筑工程标准领域知识图谱的属性则由标准短语的属性词、比较词和规定值共同组成,可对式(1)中的各元素进行对齐,即标准知识实体集 E 可表示为 $E = \{i_1, i_2, i_3, \dots, i_n\}$,关系集 R 可表示为 $R = \{r_1, r_2, r_3, \dots, r_n\}$,属性集 P 可表示为 $P = \{(c_1, p_1, n_1), (c_2, p_2, n_2), (c_3, p_3, n_3), \dots, (c_n, p_n, n_n)\}$,如图4所示。最终,经对齐后可形成能够存储入知识图谱的数据结构。

概念是人类在认识过程中,把所感知事物的共同本质特点抽象出来,加以概括的表达^[3]。一个概念可以对应多个实体,同时一个实体也可以有多个概念,如图5中的实体硅酸盐水泥可以抽象为概念“建筑材料”,中国建筑材料科学研究院可以抽象为概念“单位”。所有这些概念又可以总结为概念“建设工程标准相关”。

经以上步骤,即完成了标准知识的实体、关系、属性和概念的识别与抽取,进而形成建筑工程标准领域知识图谱。同时在构建标准知识图谱的过程中也应注意实体间所存在的隐性关系,即无法通过条文结构分析所得出关系词,进而形成关系。以图5的信息为例,实体“硅酸盐水泥”与实体“本标准(通用硅酸盐水泥)”之间不存在通过文本分析即可得出的关系词,但因其皆出自同一标准,因此两者之间存在隐性关系“属于”。

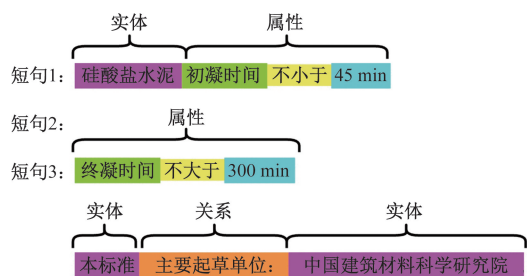


图4 标准知识图谱相关元素抽取

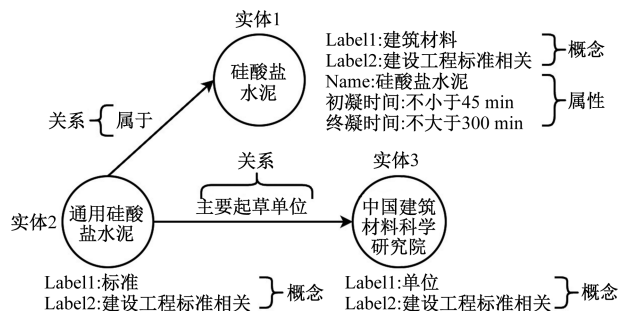


图5 知识图谱表现形式

5 应用案例演示

以建筑工程领域标准为研究对象,相关标准来源于住房和城乡建设部和国务院发布及主管的相关标准、规范和条例。通过全国标准信息公共服务平台、中华人民共和国住房和城乡建设部网站和工程建设标准化信息网进行收集,收集结果见表3。虽然相关标准、规范和条例是公开披露和可获取的,但文件通常是以PDF格式文件提供,因此使用OCR(optical character recognition,光学字符识别)技术进行内容提取。

首先依照本文4.1节的步骤对建筑工程领域标准文本进行相关实例、关系词、属性词、比较词和规定值的提取。以《建筑施工作业劳动防护用品配备及使用标准》(JGJ 184—2009)为例,其具体内容为:“抹灰工应配备高腰布面胶底防滑鞋和手套,宜配备防护眼镜。”对以上内容进行识别,可识别出“抹灰工”“高腰布面胶底防滑鞋”“手套”等实例,以及如“应配备”与“宜配备”等关系词。同时,“抹灰工应配备高腰布面胶底防滑鞋”条文结构可识别为实例关系型。然后,按照4.2节的步骤将以上内容与建筑工程标准领域知识图谱中的节点、关系和属性进行对齐,使用人工的方法为它们赋予概念,如“抹灰工”可对应概念“人员”。

表3 文件收集结果统计

类别	数量	代表文件
国家标准	45	《施工升降机安全规程》(GB/T 34023—2017)
		《振动沉拔桩机安全操作规程》(GB/T 13750—2023)
行业标准	30	《建筑深基坑工程施工安全技术规范》(JGJ 311—2013)
		《建筑施工作业劳动防护用品配备及使用标准》(JGJ 184—2009)
法律法规、规章	17	《危险性较大的分部分项工程安全管理规定》

经提取后的节点、关系和属性使用 JSON (JavaScript object notation, JavaScript 对象表示法)文件作为中间文件进行储存。使用 Python 的第三方库 py2neo 将 JSON 文件中的数据写入 neo4j 图数据库。图数据库保存结果如图 6 所示。

使用 Cypher 语言可实现知识图谱内容多维度、多层次的查询与统计。通过对建筑工程标准领域知识图谱进行查询与统计,可迅速获取清晰直观的建筑工程领域标准知识。使用 MATCH 命令查询图谱中的节点、关系及属性的相关信息。例如查看所有关系属性“相关文件”中的内容,输入命令“MATCH ()-[r]-() WHERE (r.`相关文件`)

IS NOT NULL RETURN DISTINCT “relationship”AS entity, r.`相关文件` AS `相关文件`”运行结果如图 7 所示。

6 结论

(1)构建的知识图谱可以有效地将建筑工程领域中的标准知识进行结构化表达,向建筑从业人员提供准确、全面的建筑工程领域相关知识。对提升建筑工程领域标准的数字化水平具有重要意义。

(2)结构化后的建筑工程领域标准知识更加系统、全面、易获取。可为建筑工程的设计、施工和维护提供准确、全面、及时的标准知识支持。

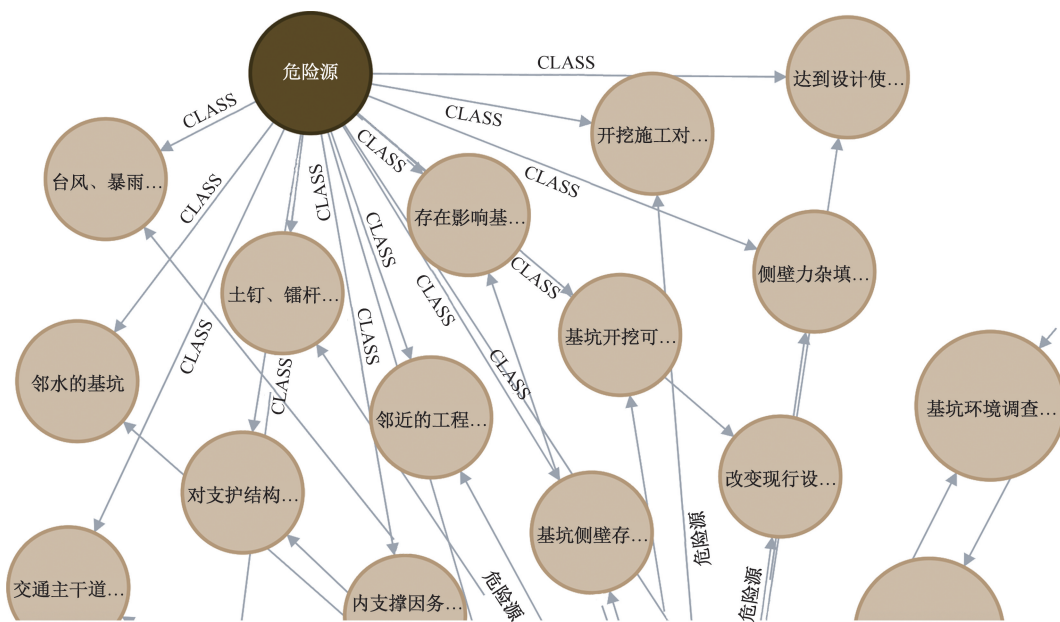


图 6 建筑工程标准领域知识图谱(部分)

```

1 MATCH (n) WHERE (n.`相关文件`) IS NOT NULL
2 RETURN DISTINCT "node" as entity, n.`相关文件` AS `相关文件`
   LIMIT 25
3 UNION ALL
4 MATCH ()-[r]-() WHERE (r.`相关文件`) IS NOT NULL
5 RETURN DISTINCT "relationship" AS entity, r.`相关文件` AS `相关
   文件` LIMIT 25

```

entity	相关文件
1 "node"	"《JGJ94-2008建筑桩机技术规范》"
2 "node"	"《JGJ94-2008建筑桩机技术规范》《JGJ46-2022施工现场临时用电安全技术规范》"
3 "node"	"《JG/T571-2019玻纤增强聚氨酯节能门窗》"
4	

图 7 领域知识图谱关系属性“相关文件”中的内容

参考文献

- [1] 田世宏, 标准化理论与实践[M]. 北京: 中国标准出版社, 2023: 12-20.
- [2] 林佳瑞, 陈柯吟, 郑哲, 等. 建筑工程标准规范智能解译关键技术及应用[J]. 工程力学, 2024(2): 1-14.
- [3] 刘均昊, 苏义坤, 张智博, 等. 面向建筑施工安全的标准知识支持研究[J]. 中国安全生产科学技术, 2024, 20(7): 156-162.
- [4] 杨跃翔, 涂新雨, 刘文玲. 标准文献知识图谱构建与应用研究[J]. 数字图书馆论坛, 2022(6): 22-30.
- [5] 秦丽, 郝志刚, 李国亮. 国家食品安全标准图谱的构建及关联性分析[J]. 计算机应用, 2021, 41(4): 1005-1011.
- [6] 康延领, 尤日淳, 唐文哲, 等. 中外工程建设标准体系差异及发展策略研究[J]. 建筑经济, 2020, 41(8): 10-13.
- [7] 林佳瑞, 陈柯吟, 潘鹏. 建筑工程标准数字化与智能化: 现状与未来[J]. 东南大学学报(自然科学版), 2025, 55(1): 16-29.
- [8] 高欣, 王引兰. 基于知识图谱与智能问答系统的淮扬美食可视化研究[J]. 美食研究, 2024, 41(4): 38-45.
- [9] CHEN Y, LU G, WANG K, et al. Knowledge graph for safety management standards of water conservancy construction engineering[J]. Automation in Construction, 2024, 168: 105873.
- [10] 姚卫华, 葛悦, 邓兵兵, 等. 城镇燃气管网应急处置知识图谱构建与应用[J]. 科学技术与工程, 2024, 24(10): 4343-4351.
- [11] 许雅玺, 孟天宇, 王欣, 等. 航空不安全事件知识图谱构建[J]. 科学技术与工程, 2024, 24(32): 13925-13934.
- [12] 刘慧琳, 牛力. 标准文件的知识图谱组织模式探究[J]. 档案学通讯, 2021(5): 58-65.
- [13] 王佳, 刘玥, 李继宝. 面向 BIM 消防智能审查的规范语义模型研究[J]. 消防科学与技术, 2022, 41(2): 242-246.
- [14] 穆磊, 王佳, 李继宝, 等. 基于 BIM 和知识图谱的消防智能审图研究[J]. 消防科学与技术, 2019, 38(12): 1765-1768.
- [15] 林洁如, 胡最. 基于深度学习模型的文化景观基因命名实体识别[J]. 地球信息科学学报, 2025, 27(1): 207-225.
- [16] 刘玥. 面向 BIM 消防审查的规范语义模型构建方法研究[D]. 北京: 北京建筑大学, 2022.
- [17] 王茹, 赵俊浩, 黄炜, 等. 基于知识图谱的建筑施工安全风险量化与分析[J]. 安全与环境学报, 2024, 24(6): 2138-2147.

Research on the Digitization of Construction Engineering Standards Based on Domain Knowledge Graph

LI Gang, SU Yikun, LIU Junhao, ZHENG Zhizhe

(College of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China)

Abstract: To promote the digitization of standards in the field of construction engineering in China, the patterns of construction engineering standards and text structures were summarized, the process of knowledge graph construction was elaborated, and both manual and natural language processing techniques were applied to build a knowledge graph in the domain of construction engineering standards. Through practical case applications of the standards, it is verified that the knowledge graph in the field of construction engineering standards exhibits good usability in structured representation, extraction, and retrieval of relevant knowledge. The constructed knowledge graph can effectively structure and represent standard knowledge in the field of construction engineering.

Keywords: digitization of standards; knowledge graph; standard structurization; graph database