

建筑工程标准数字化与智能化：现状与未来

林佳瑞^{1,2}, 陈柯吟¹, 潘鹏^{1,2}

(¹清华大学土木工程系, 北京 100084; ²住房和城乡建设部数字建造与孪生重点实验室, 北京 100084)

摘要: 为了推动工程标准数字化与智能化, 实现标准规范内容的机器可读、可算, 系统调研了标准数字化研究现状, 提出了建筑工程标准数字化发展阶段的分级模型及其典型应用场景, 为标准数字化与智能化发展提供了清晰路径。同时, 建立了覆盖存量标准智能解译、增量标准智能编制与标准多场景应用 3 个方面的建筑工程标准数字化关键技术框架, 并系统综述了框架各部分关键技术研究应用现状、未来需求及关键挑战。研究表明, 相关研究已在知识抽取、条文解译、协同设计、智能审查等方面取得显著进展, 为标准数字化解译提供了有力支持; 但完全实现标准数字化转型尚需时日, 仍需结合多模态大模型 (LMM) 等技术, 在标准数字模型、标准智能解译与编制、多场景应用等方面开展更多探索。

关键词: 标准体系; 数字化转型; 机器可读标准; 规则解译; 数字标准; 综述

中国图书分类号: F203

Digital and intelligent standards for building and construction engineering:

Current status and future

Lin Jiarui^{1,2}, Chen Keyin¹, Pan Peng^{1,2}

(¹Department of Civil Engineering, Tsinghua University, Beijing 100084, China; ²Key Laboratory of Digital Construction and Twinning, Ministry of Housing and Urban-Rural Development, Beijing 100084, China)

Abstract: To promote the development of digitalization and intelligence of construction engineering standards, achieving the machine readability and computability of standard specifications, this paper has conducted research on the current state of standard digitalization, proposed a grading model for the development stages of construction engineering standard digitalization and its typical application scenarios, providing a clear development path for the process of standard digitalization. Meanwhile, a key technical framework for the digitalization of construction engineering standards covering the intelligent interpretation of existing standards, intelligent compilation of incremental standards, and multi-scenario application of standards has been established. A systematic review of the research and application status, future needs, and key challenges of the key technologies in each part of the framework has also been conducted. Currently, significant progress has been made in knowledge extraction, rule interpretation, collaborative design, intelligent review, and other areas, providing strong support for the digitalization of standards. However, the complete realization of the digital transformation of standards still requires time and further exploration in aspects such as standard digital models, intelligent interpretation and compilation of standards, and multi-scenario applications, combining technologies such as large multimodal models (LMMs).

Key words: standard system; digital transformation; machine-readable standards; rule interpretation; digital standards; review

建筑业是我国国民经济的重要支柱产业。随着行业规模不断扩大, 传统的粗放型发展模式导致资源能

收稿日期: 2024-08-15.

作者简介: 林佳瑞 (1987—), 男, 博士, 副研究员, lin611@tsinghua.edu.cn; 潘鹏 (通信作者), 男, 博士, 教授, 博士生导师, panpeng@tsinghua.edu.cn.

基金项目: 国家重点研发计划资助项目(2023YFC3804600); 国家自然科学基金资助项目(52378306).

源消耗剧增、质量安全风险显著提升，这些问题已成为制约建筑业可持续发展的关键瓶颈^[1]。在此背景下，通过新兴数字化技术创新驱动建筑业的转型升级，是实现建筑业高质量发展的必由之路^[2]。建筑工程标准规范是行业长期研究与实践经验的重要体现，为建筑活动主体行为提供了明确的指导纲领，对保证建筑工程全生命期的安全性、可持续性和舒适性至关重要，是保障建筑业高质量发展的重要依据。例如，规划设计阶段基于相关标准的设计图纸审查是确保建筑满足规范要求、保障质量安全的关键环节^[3]，而施工运维阶段的安全隐患识别排查、质量验收以及合同管理等各环节同样依赖于相关标准规范的约束^[4]。

随着建筑信息模型（building information modeling, BIM）、人工智能等先进信息技术与建筑机器人^[5]的不断推广应用，探索智能建造新模式已成为建筑业转型升级的重要战略举措。相应地，如何让计算机或机器人实现工程图纸、文档、模型、图像等异构多模态数据的高效获取、智能分析与充分利用，是建筑业数字化转型与智能化升级的关键难题。然而，当前标准规范的表现形式仍以文本为主，主要面向工程师人工解读与利用，面临查询检索困难、人工校核低效等问题，已难以满足智能设计、建筑机器人等未来人机协同场景的需求。因此，探索建筑工程标准规范内容数字化表达、拆解与推理技术，实现相关标准规范的机器可读、高效推理与融合计算，构建兼顾人工阅读与机器推理的标准规范新形态，实现建筑工程标准的数字化发展，对推动建筑业数字化转型与智能建造发展具有重要意义。具体来说，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，实现标准内容的智能化检索、自动解译与协同化编制，促进标准信息与建筑工程各环节的无缝集成与应用。

近年来，有关学者结合自然语言处理（natural language processing, NLP）、大语言模型（large language model, LLM）等人工智能技术，在标准规范知识提取、条文解译及设计审查等方面开展了诸多探索，极大推动了标准规范数字化的发展。本文旨在综述国内外工程标准数字化的相关研究进展，结合现有研究提出建筑业标准数字化发展的分级模型，并在此基础上系统总结标准数字化关键技术的研究现状与未来趋势，明确标准数字化面临的关键挑战，为未来建筑业标准规范的数字化与智能化发展提供建议。

1 研究思路

本文总体的技术路线如图 1 所示。首先，围绕建筑业标准数字化的关键技术，在 Web of Science 和 CNKI 数据库中对相关论文进行调研分析和筛选。其次，结合国内外现有研究中对标准数字化发展阶段的分级标准，提出了面向建筑业标准数字化分级模型。随后，对筛选后的 60 篇论文和相关报道进行分类总结，将其分为标准数字模型、标准内容的数字解译技术、新型编制技术、机器语言标准生成技术、标准推理校核技术及应用 5 个部分，详细阐述了各部分关键技术的研究成果和应用现状。最后，总结了建筑业标准规范数字化发展的未来趋势及面临的挑战。

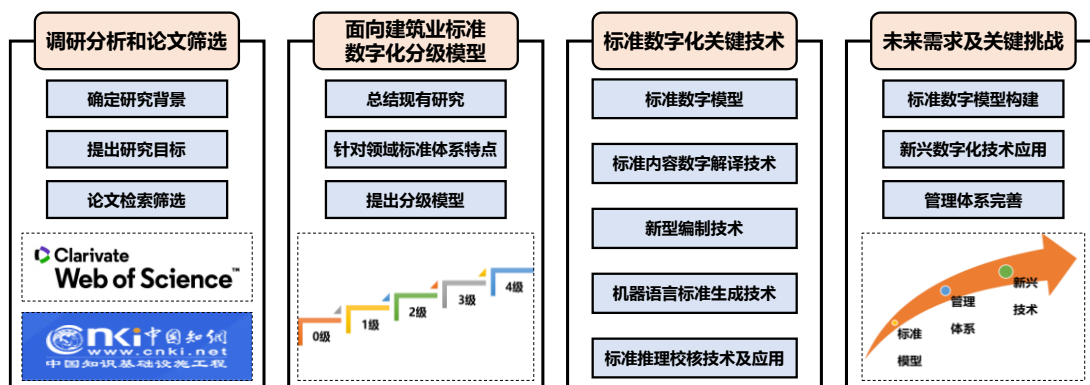


图 1 技术路线图

Fig.1 Technology roadmap

2 标准数字化发展的分级模型

标准数字化的发展不是一蹴而就的，而需要分级分阶段不断推进。当前，工程建设领域已经基本形成了一个结构复杂、内容丰富、体量庞大的标准体系^[6]，建立分级模型可以清晰地划分发展阶段，为标准数字化发展提供阶段目标。

国际标准化组织（ISO）和国际电工委员会（IEC）一直致力于推动标准数字化工作，建立了机器可读标准分级模型，该模型是目前使用最广泛的分级模型，已在众多标准组织中达成共识^[7-8]。在该分级模型中，0级阶段的标准以纸质文档的形式存在；1级阶段将纸质文档的标准转换为PDF等开放数据格式；2级“机器可读”阶段，重点在于对电子文档中的标准内容进行结构化处理；3级“机器可执行”阶段，进一步深化了对电子文档内容的处理，通过语义化增强了机器对标准内容的深入理解和应用能力；4级“机器可决策”阶段，旨在实现标准内容的自主学习和预测功能，从而赋予机器更高的智能水平和决策支持能力。ISO将自身的标准数字化水平定义在“机器可读”阶段，即借助可扩展标记语言（XML）实现标准内容的结构化，进而快速识别与检索标准内容^[9]。

上述分级模型更侧重于电子文档的格式和结构化程度，3级和4级之间的界限不够明确，缺乏对于建筑工程领域的实际应用和相关数据模型集成的相关说明。因此，本研究结合了现有研究中对机器可读标准分级模型的表述，考虑到领域知识和BIM等多源异构数据的集成、专业性、关联性强等工程领域的特点，提出了更具体的分级模型。同时，本研究进一步明确了各级别的关键支撑技术和典型应用场景，如图2所示，可以更好地适应建筑业标准数字化的发展需求，为标准数字化与智能化发展提供了清晰路径。

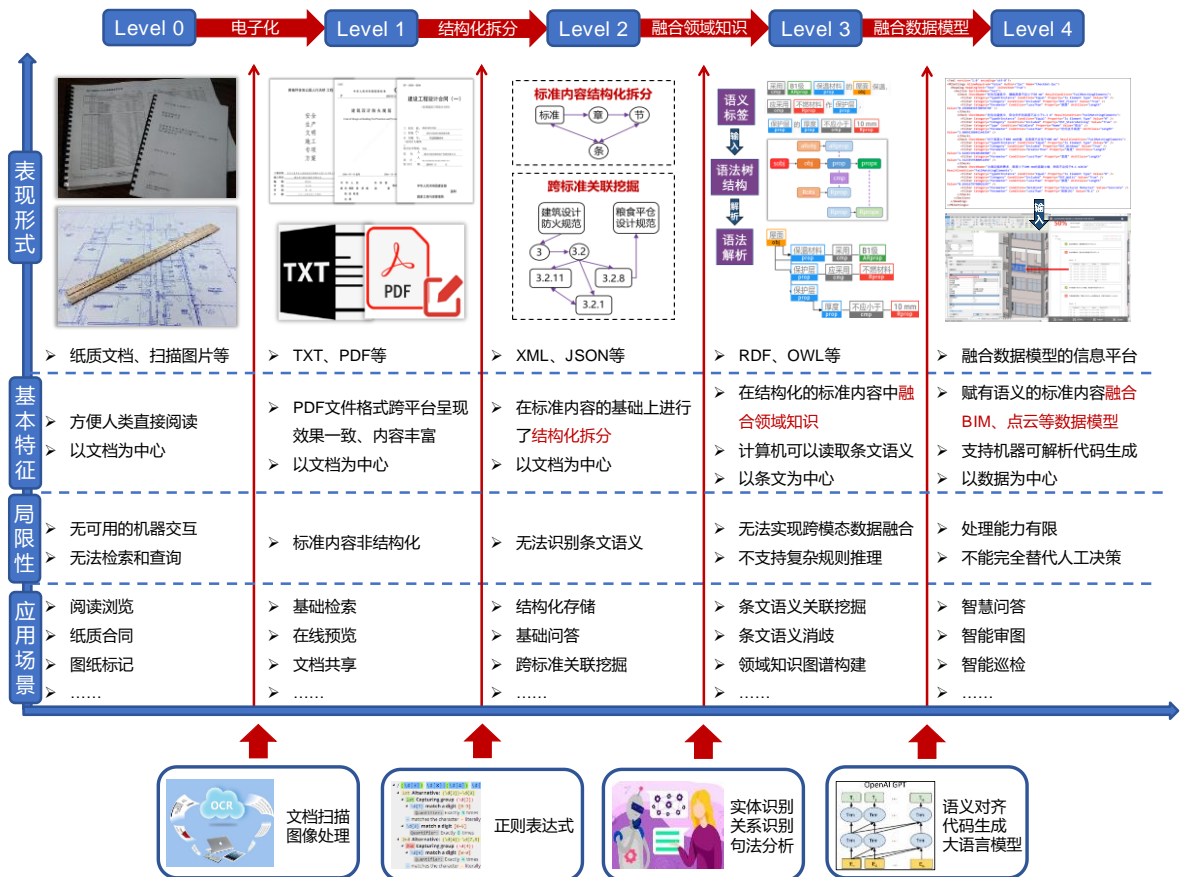


图2 建筑业标准数字化发展的分级模型

Fig.2 Maturity model for standard digitalization in the building and construction industry

0 级阶段主要表现形式是纸质文档和扫描图片, 这种形式方便人类阅读, 但是无法被计算机检索和查询; 1 级阶段主要表现形式是 TXT、CSV 等纯文本格式和 PDF、Docx 等可移植文档格式, 这种形式可以实现基础的内容检索, 方便跨平台的文件预览; 2 级阶段主要表现形式是结构化的标准内容, 利用 XML、轻量级的数据交换格式 (JavaScript object notation, JSON) 等文档结构化描述语言在 1 级阶段的基础上进行了标准内容章、节、条、款的拆分, 能够实现文档级别的基础问答和标准间的关联挖掘, 例如《建筑设计防火规范》与《粮食平房仓设计规范》标准间引用关系的挖掘, 但是该阶段无法识别条文语义; 3 级阶段主要表现形式是赋有语义的标准内容, 在 2 级阶段的基础上融合了领域知识, 实现条文内容的语义标注和语句解析, 进而应用于条文之间的语义关联关系识别和消歧, 例如《人民防空工程设计防火规范》与《建筑设计防火规范》中对舞台与观众厅之间防火隔墙耐火极限的规定, 但是该阶段无法实现复杂的规则推理; 4 级阶段主要表现形式是融合 BIM、点云数据以及其他相关数据模型的信息平台, 可以生成机器可解析的代码, 实现标准语义和不同数据模型之间的对齐, 进而应用于面向机器的标准推理和校核等场景, 但还不能完全替代人工决策, 复杂场景推理仍需人工干预。

3 标准数字化关键技术

3.1 整体框架和思路

2018 年全国标准信息公共服务平台与中国标准信息服务网同步升级, 基本标志着国际标准信息服务推广体系在我国的初步建成^[10], 到 2020 年, 国家标准委在制造业和航空航天领域启动了机器可读标准的试点项目, 为工程标准数字化提供了可借鉴的经验。在工程建设领域, 对于已制定并投入使用的工程标准, 其电子化呈现形式的探索, 即 1 级阶段已经较为成熟, 正在逐步过渡到条文之间关系的建立以及规则的语义标签化和表达的形式化, 即 2 级和 3 级阶段^[11]。当前研究重点在于基于章节条拆解、条文分类、知识抽取、句法分析、知识融合等标准内容的数字解译技术, 以及如何构建领域标准条文库、语义常识库、机器规则库, 并实现条文的高级检索应用。同时, 在工程标准制定与修订环节, 研究融合人工智能的标准新型编制技术, 并结合既有标准条文库、语义常识库实现相关标准内容的检索、查重、智能辅助编写与标准知识库更新, 也是未来标准数字化的重要方向。进而, 面向计算机与机器人等机器推理与应用需求, 研究机器规则语言的自动生成与高效推理校核技术, 实现标准的机器可读与高效推理。

结合以上分析和梳理, 本研究提出如图 3 所示的标准数字化关键技术体系框架, 覆盖了存量标准智能解译、增量标准智能编制与标准多场景应用 3 个方面。具体包括标准数字模型、标准内容的数字解译技术、新型编制技术、机器语言标准生成技术、标准推理校核技术及应用 5 个部分。其中, 标准多场景应用包括机器语言标准生成和标准推理校核。标准数字模型是标准内容的数字解译、新型标准编制、标准多场景应用的基础, 标准内容的数字解译结果、新编的数字化标准、生成的机器语言又不断扩充标准数字模型。标准内容的数字解译是其生成机器语言的基础, 生成的机器语言和新编的数字化标准又被直接应用于标准的推理校核, 标准推理校核的过程中也可以不断完善生成的机器语言, 服务于新型标准的编制。以《建筑设计防火规范》(GB 50016—2014)^[12]的数字化转换为例, 首先, 在标准数字模型阶段, 需构建领域本体模型, 定义防火隔墙、安全出口、防火分区等核心概念及其属性关系。其次, 在标准内容数字解译阶段, 需要通过文档扫描、图像处理技术将标准规范转换为可处理的文本, 通过章节条的拆解将其转化为结构化文本, 进而识别“6.2.1 剧场等建筑的舞台与观众厅之间的隔墙应采用耐火极限不低于 3.00 h 的防火隔墙。”等具体条文; 再次, 通过条文分类将该条文作为可解译的简单句, 利用自动化知识抽取技术提取“舞台”、“观众厅”、“防火隔墙”等实体及其关系, 识别“3.00 h”这一具体数值要求; 最后将条文解译为“IF exists(舞台) AND exists(观众厅) AND connected(舞台,观众厅) THEN requires(防火隔墙.耐火极限≤3.00 h)”等机器可理解的逻辑表达式。在新型编制技术阶段, 基于已构建的知识库, 系统可智能推荐相关条文, 辅助

编制人员进行条文的查重与关联分析。在机器语言生成阶段，将逻辑表达式进一步转化为可执行的程序代码。最后在标准推理校核阶段，基于生成的代码规则，可在 BIM 模型中自动检查防火隔墙的布置是否符合规范要求，实现标准的智能应用。标准数字化各环节之间存在紧密的逻辑关联，共同构成了一个完整的技术链条。本文将分类总结各部分的相关研究应用进展。

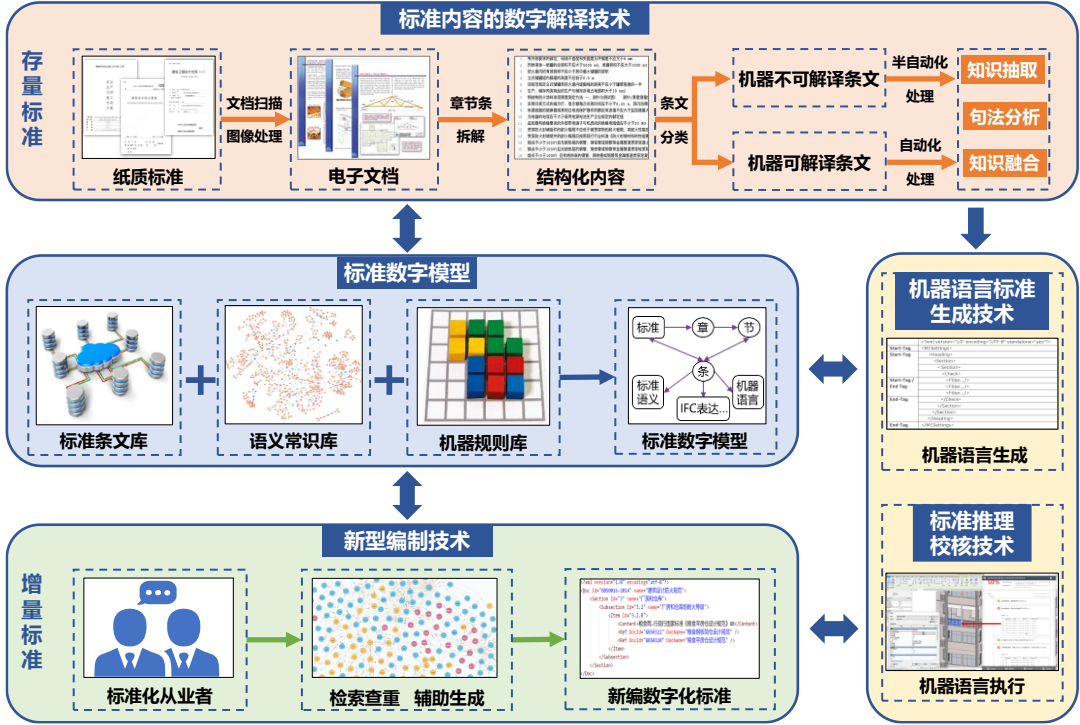


图3 标准数字化关键技术体系框架

Fig.3 Overall technical framework for standard digitalization

3.2 标准数字模型

在工程建设领域，标准数字模型的构建是实现标准数字化转型的基础。标准数字模型的核心是实现自然语言文本、工程语义特征及机器可读规则语言的结构化及形式化表示，并提供与 BIM、点云、图像等多模态数据的融合及语义映射机制，以形成一个完备、统一、通用的标准数字化框架，兼顾人类便捷阅读使用需求与机器高效读取推理需求。

数字标准模型的发展始于 Level 2 阶段，Level 2 阶段的标准模型主要基于正则表达式匹配，采用 XML、JSON、RASE（requirement, applicability, selection, exception）^[13-14] 等标记语言，对标准文档章、节、条构成及层级关系进行结构化描述，相关技术已经相对成熟。Level 3 阶段的标准模型则进一步融合了领域知识，结合语义网技术形成领域知识图谱，目前研究的探索主要集中在这个阶段，本节重点介绍该阶段的相关研究进展。而 Level 4 阶段的数字标准模型究竟如何构建，仍有待深入研究和探索。各阶段对应的数字标准模型主要表现形式如图 4 所示。

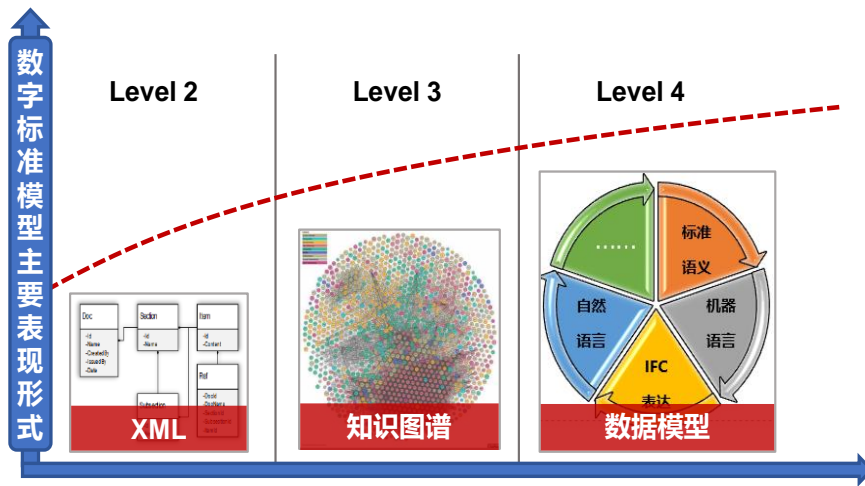


图 4 数字标准模型主要表现形式

Fig.4 Main representation models for digital standards

本体及语义网技术通过机器可理解的统一信息描述来共享领域知识^[15]，基于 RDF（resource description framework）框架、OWL（web ontology language）本体语言等可以定义和表示建筑工程标准概念之间的类别、属性和关系，以结构化的形式存储和组织大量领域知识。例如，本体可以定义防火门这一概念，并描述其具有耐火等级、开启方向等属性，以及与防火分区、安全出口等其他概念的关系。RDF 框架可以通过主语（subject）、谓语（predicate）、宾语（object）的三元组形式来描述标准知识。例如，“防火门-具有-耐火等级”就是一个典型的 RDF 三元组。OWL 则可以在 RDF 的基础上表达更复杂的约束和推理规则，例如，“Restriction: 耐火等级 \geq （可否改为 \geq ）‘3.0 h’”可以表示防火门的耐火极限。以本体作为模式层扩展实例层领域知识，以节点表示实体或概念，边表示实体和实体之间的关系以及实体和属性的关系，则可形成领域知识图谱，将碎片化的领域知识融合形成高度关联的结构化知识^[16]。因此，语义网及知识图谱等技术被广泛用于解决建筑业数字化转型升级过程中面临的知识表示难题。不同标准数字模型的特点与适用场景如表 1 所示。

表 1 不同标准数字模型的特点与适用场景

Tab.1 Characteristics of different representations of standards and their applicable scenarios				
模型类型	主要特点	优势	局限性	适用场景
XML/JSON	采用标记语言描述文档结构	通用性强	缺乏语义表达能力	文档结构化存储
RASE	采用 4 种运算符描述文档内容	逻辑结构清晰	表达能力有限	简单规则提取
本体	定义领域概念及关系	语义表达丰富	构建成本高，需要专业知识支撑	知识形式化表达
知识图谱	基于节点和关系构建实体关系网络	知识关联完整	维护更新较为复杂	知识推理应用

为了明确标准规范相关知识的概念层次结构，增强其表达推理能力，本体的概念逐渐被应用到标准规范中的知识组织。面向工程规范中有关工程概念、属性等知识的表示问题，刘江波^[17]针对盾构施工规范构建了结合规范条文层与条文概念层的两级知识图谱架构；刘广宇等^[18]建立了公路工程安全领域的知识本体架构，提出了可以精准捕捉知识间逻辑关系的多维度层状分类知识体系；Jiang 等^[19]将知识图谱引入建筑安

全标准管理，构建了建筑安全标准知识图谱，实现了建筑安全标准知识的系统化整合，并在此基础上进行知识推理和检索应用。面向标准规范条文知识及约束规则的表达和解析，Xu 等^[20]针对地下基础设施，建立了领域特定本体以描述地下设施空间关系等知识并支持相关规范条文解析；Zheng 等^[21]则构建了建筑消防安全领域本体 (fire protection for building ontology, FPBO)，整合了建筑信息领域和消防领域的特定概念、属性、关系和描述，为消防法规的自动解译提供了重要支撑。但目前相关知识建模的研究都针对特定领域展开，如何构建覆盖多领域的通用、大规模本体和知识图谱仍有待探索。

此外，一些领域特定的形式化语言和语义标签也被提出，用来表达规范文本的语义结构特征。Dimyadi 等^[22]将法律领域新兴的开放标准 LegalDocML 和 LegalRuleML 引入建筑法规条文的表示中，说明其具有一定的应用潜力。Zhang 等^[23]等针对建筑法规文本的特点提出了关键的语义标签和概念，可以很好地适用于规范中定量要求的表示。Zhou 等^[24]面向智能审图，提出了用于规则检查的建筑结构设计规范语义标签，实现了条文约束要求信息的准确表达，并且具有良好的可扩展性。Solihin 等^[25]则利用结构化的概念图 (conceptual graph, CG) 来表示知识，其中节点表示概念或对象，关系表示对象之间的关系，函数则表示可以执行的操作。鉴于本体及知识图谱技术仍以一阶逻辑表示为主，为了解决规范中复杂高阶逻辑表达的难题，Zheng 等^[26]针对建筑工程规范的复杂约束和逻辑表示，提出了基元函数的概念，并构建了初步的基元函数库，探索了将复杂逻辑封装为基元函数的复杂逻辑表示方法。同时，中国建筑标准设计研究院已经联合多家业内公司、高校开展了“CN-IFC”标准体系的探索，将建立系列领域语义标准，为工程建设领域标准文本的语义解析和数字化转换技术提供基础^[27]。

尽管有关学者及业界已做了诸多探索，但目前工程建设领域尚未形成统一的大规模标准数字模型，缺乏一个面向整个建筑工程领域通用的知识表示模型。相关研究仍存在规范条文覆盖范围有限、数据集规模较小以及难以处理标准中的图片、公式、表格等问题，难以实现规范内容的完全机器推理与高效应用需求。

3.3 标准内容的数字解译技术

工程建设领域存量纸质标准的数字化转换是当前领域标准数字化的迫切需求之一。本研究将该过程定义为标准内容的数字解译技术，整体流程如图 5 所示。在文档内容识别阶段，将纸质或 PDF 格式的标准文档转换为可处理的电子文本；在章节解构阶段，利用正则表达式等技术识别文档的层次结构，生成结构化文档；在条文分类阶段，基于定义的条文分类标准对条文进行智能分类；在知识抽取阶段，用于识别标准中的实体概念及其关系；在规则解译阶段，将抽取的知识转换为形式化规则；在知识融合阶段，对抽取的知识中存在的 inconsistency 进行冲突消解。本节将重点介绍这些技术在国内外取得的研究成果，以及在实际应用过程中存在的问题。

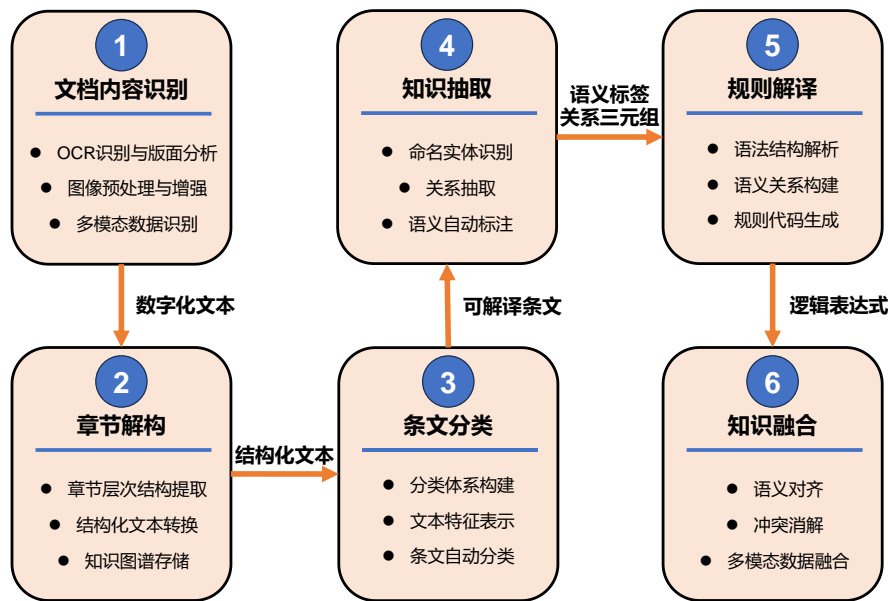


图 5 标准内容的数字解译流程

Fig.5 Process for digital interpretation of standards

3.3.1 文档内容识别技术

标准数字化的第一步是需要对纸质标准进行文档扫描和图像处理，对其内容进行识别和处理，并转换为电子文档。图像处理过程中，常采用光学字符识别（optical character recognition, OCR）技术，OCR 技术已经被广泛应用于书籍和其他非结构化文档的数字化。

处理流程通常包括 3 个主要阶段：预处理、文字区域检测和文字识别。预处理包括对图片的降噪、倾斜校正等。传统的 OCR 技术常使用灰度化、二值化的方法对文本图片进行预处理，而基于深度学习的方法主要聚焦于图像增强^[28]。文字区域检测通过版面分析可以将文档图像分为纯文本、标题、表格、图片等^[29]。传统的文字识别大多使用模板匹配的方法，运算效率较低，利用卷积循环神经网络（convolutional recurrent neural network, CRNN）^[30]等基于深度学习的字符识别模型可以实现较高精度的文字识别。

OCR 识别仍然面临重叠、模糊等各类成像问题，而标准规范涉及到文本、表格、公式乃至图集等多模态数据，其内容的完整识别将更加复杂和困难，未来需要进一步发展相关文档识别技术以实现高质量的纸质标准内容识别和电子文档生成。

3.3.2 章节解构技术

章节解构的重点在于从文档中提取出层次化结构。标准正文按照章、节、条、款、项划分层次，章是标准的分类单元，节是标准的分组单元，条是标准的基本单元，当条的层次较多时，可细分为款，款也可再分成项^[31]。标准规范编制过程对前述各层次的编号往往具有明确统一的要求，这种规范性、统一化的编排形式有利于通过编写正则表达式的方式快速实现章节解构，大幅提高处理效率。

正则表达式由简单的元字符与运算符组成，可以灵活匹配复杂的文本模式，是一种快速有效的文本处理方法。通过正则表达式可以快速匹配有关信息^[32]，这种方法已经被广泛应用于标准规范文本的章节解构中，将非结构化文本转换成 XML、CSV 等格式，并基于知识图谱实聚合存储^[33]，但是正则表达式的编写主要依赖人工，适用范围还有待进一步扩大。针对上述模式匹配方法存在的依赖人工、可扩展性差等问题，Bartoli 等^[34]探索了基于用户提供的示例自动合成正则表达式的系统，能够自动化处理和生成复杂的正则表达式，但是系统的性能、效率和泛化能力还有待进一步提升。

3.3.3 条文分类技术

标准条文涉及到不同的专业领域和应用场景，对标准条文进行分类可以为标准知识抽取奠定基础。条文分类体系是条文分类的基础，我国学者已经对面向建筑工程领域的标准分类体系展开了系列研究。基于规范的约束类别，胡海盟^[35]把质量验收规范条文分为属性约束、工序约束、检查约束和参照约束 4 类，建立了不同的知识表达模型；李璨^[36]将规范条文分为存在类规范条文、比较类规范条文、组合类规范条文和暂不可转译的规范条文 4 类；舒赛^[37]把消防设计规范分为存在约束类规范、类别约束类规范、数量约束类规范、距离约束类规范、位置约束类规范、属性约束类规范 6 类。面向智能审图应用过程中条文解译的难度，Zheng 等^[3]将其分为直接属性约束类条文、简介属性约束类条文、方法约束类条文、外部参考性条文、通用设计条文、术语定义类条文、其他条文 7 类，并提出了分类优先级，以避免条文分类过程中的歧义。然而，目前尚缺乏一个面向建筑工程领域通用的分类体系框架。

采用自动化条文分类技术可显著提升处理大量条文的效率^[3]。早期的文本分类方法多依赖于人工定义规则和统计学习，这要求在特征工程上投入巨大的精力。基于深度学习的方法实现条文分类首先需要将自然语言文本表示为向量，在这个过程中词袋模型和词嵌入模型^[38]被广泛使用。然后，基于卷积神经网络（convolutional neural network, CNN）和循环神经网络（recurrent neural network, RNN）等模型，利用大量人工准备的训练数据集^[20]对模型进行训练。近年来，预训练的深度学习模型，如 BERT（bidirectional encoder representations from transformers），因其所需的训练数据集相对较小，效果优于传统的深度学习模型，已经在各个领域得到广泛应用^[3]。此外，LLM 在少样本训练下的文本分类任务中展现出了卓越的性能^[39]，具有良好的泛化能力，可以处理表达方式多样的条文内容，但将其应用于建筑工程领域标准条文的自动分类任务的研究工作还相对匮乏，分类结果的可靠性以及对于不同类别标准的适用性仍然需要进一步的专业验证，特别是涉及专业术语和复杂逻辑的条文。

3.3.4 知识抽取技术

知识抽取是基于给定的形式从不同来源、不同结构的文档内容中抽取有关知识实体、特征的过程，是实现标准知识结构化的重要步骤^[40]。命名实体识别（named entity recognition, NER）和关系抽取（relation extraction, RE）是知识抽取中的 2 个关键任务。NER 的目的是从非结构化文本中识别出预定义类别的实体及概念，例如标准规范条文：“矩形截面框架柱的边长不应小于 300 mm”中包含构件“矩形截面框架柱”、属性“边长”、尺寸“300 mm”等。RE 则是在此基础上，进一步抽取各个实体之间的语义关系，构建结构化的三元组知识^[41]，如属性“边长”和尺寸“300mm”之间存在的比较关系“不应小于”等。

传统的 NER 方法主要基于预定义的规则或字典^[42]。尽管基于规则的方法简单易用、可解释性强，但是需要大量人工工作，泛化性能差。而基于机器学习的方法可以通过训练好的模型来预测命名实体，在处理大规模数据时具有一定的优势，经典的方法依赖于条件随机场（conditional random field, CRF）^[43]等序列标记模型。然而，基于机器学习的方法可能会导致特征提取误差传播问题^[44]。近年来，深度学习技术被用于提高 NER 的准确性^[45]。Lample 等^[43]提出了长短期记忆网络（LSTM）和 CRF 的混合架构，可以不依赖任何语言学特征和资源，在多语言命名实体识别任务上取得非常好的效果。Zhou 等^[24]通过预训练模型 BERT 微调的方式实现了规范条文的高效语义标注，获得了较高的准确率。

关系的识别与实体类似，也主要基于定义的规则、机器学习和深度学习方法^[42]。CNN 和其他基于自注意力机制^[46]的深度学习模型被广泛用于关系提取。为了确保模型能够精确地处理建筑工程领域的特定任务，必须人工构建特定的领域数据集进行训练，但目前仍然缺乏面向建筑工程标准知识抽取的大规模开放数据集。这限制了用于特定领域知识抽取的模型训练，实现知识抽取过程的全流程自动化还较为困难。

随着对深度学习研究的不断深入,研究者发现 LLM 在文本生成、少样本学习、零样本学习、推理任务等方面展现了优异的性能,为 NER 和 RE 任务提供了新的解决方案。Zheng 等^[47]训练了针对建筑业的领域语言模型,在 NER 任务中性能显著优于谷歌、百度等通用模型。Wang 等^[48]通过提示词工程将 GPT-3 模型用于 NER 任务,在资源受限和少样本学习场景下获得了良好的性能。但 LLM 在建筑工程领域的 NER 和 RE 任务中的应用尚处于起步阶段,知识抽取准确率仍然较低,值得进一步探索和推广应用,需要根据具体的任务对 LLM 进行微调或提示,探索提高准确率的技术路径。

3.3.5 规则解译技术

规则解译的最终目的是将标准条文的自然语言表达转换为规则代码,规则解译的过程中需要通过句法解析确定复杂条文的语法结构,将其转换为便于计算机程序进行自动化解译和推理的格式。现有的研究主要集中在基于正则表达式的模式匹配方法^[49]。但这种方法中正则表达式的构建和维护成本很高,且其通用性较低,表达能力有限,难以处理复杂的条文。为了克服这些限制,相关研究开始探索更为强大的工具。上下文无关文法(context free grammar, CFG)有更强的表达能力,可以根据知识抽取结果生成具有层次关系的树结构,从而更准确地捕捉规范文本中的复杂语法结构。Zhou 等^[24]提出了一种表达语义元素之间的关系语法树结构,基于 CFG 实现了知识表达形式的转换。未来,仍需要进一步研究和探索涉及复杂数学运算或高阶逻辑的条文的解译方法。

3.3.6 知识融合技术

在知识抽取的过程中可能会存在语义的歧义,因此需要对抽取的知识进行融合。这是由于标准规范中的概念和关系与设计模型中的概念和关系可能存在不一致^[50],并且不同的标准文件可能使用不同的自然语言词汇来描述相同的概念^[21]。

用于知识融合的语义对齐方法主要被分为基于规则的方法、监督学习方法、无监督学习方法 3 种。基于规则的方法通过基于字典的查找和字符串匹配算法来度量提取的词与本体概念之间的相似性,然后基于人工定义冲突解决规则库来处理语义冲突,实现知识的融合^[49]。此外,构建融合 BIM 等不同数据模型的本体^[51]也可以为语义对齐的过程提供支持。但这些方法往往需要大量的人工投入。

随着深度学习的发展,神经网络模型如 IPTransE 迭代对齐模型也可以用来计算实体间的相似性^[52],从而实现知识的融合。尽管这种方法消除了人工对于字典和冲突解决规则库定义的参与,但仍然需要大量的人工工作来准备训练数据集^[20]。无监督方法如词嵌入模型等在数据准备阶段需要较少的人工工作,可以较好地推广应用,处理大规模标准规范^[21]。尽管如此,当前研究的焦点仍以文本数据的语义对齐为主,如何针对建筑工程特点,实现标准规范文本以及 BIM、点云等多模态数据的语义对齐与知识融合仍需更多探索。未来,需要建立面向建筑工程领域的统一本体框架,探索基于深度学习的多模态数据融合方法,实现标准数字化全流程的无缝对接和互操作。

3.4 新型编制技术

在建筑工程领域,标准规范的制定是一项至关重要的工作。然而,目前我国工程建筑工程标准制定主要依赖领域专家人工编写,表达多样且容易出错,存在标准编制周期较长、难以智能应用等问题^[53]。为了克服这些挑战,必须探索利用机器辅助编写的自动化方式,提高编写效率,减少人为错误,优化整个标准的制定流程。机器辅助编写的优势在于其能够快速处理和检索大量数据和信息,为相关人员提供决策支持。此外,机器辅助编写还可以通过提供编制模板和协同编制平台,减少重复性工作,降低沟通成本,直接编制机器可读标准,提高整个标准的制定和应用效率。对于新增的工程标准,依托协同编制平台,集成智能检索、问答、推荐、辅助生成等功能,可以提高各参与方在标准制定过程中的积极性,优化标准制定流程,直接制定机器可读标准,是实现标准数字化的重要需求。

在标准规范编写过程中，首先需要从大量既有规范条文中检索相似条文，并进行语义挖掘，避免条文冲突。因此，需要研究高效和准确的类似条文检索推荐、条文关联关系提醒、智能推荐和问答技术，这一过程主要基于知识图谱和文本相似语义计算模型实现。传统的基于统计的语义相似计算模型、基于机器学习和深度学习的语义相似计算模型在这一过程中被广泛使用^[54]。

在法律领域，已有平台如中国裁判文书网、中国司法案例网、法信等已经可以基于关键词实现相似案件的智能推送，但是其精确度尚待提高^[55]。在建筑工程领域，住房和城乡建设领域法规标准知识服务平台已经实现了 8000 多本领域标准的在线阅读、版本对比功能^[56]，中建八局的方案人工智能审核平台可以实现相似施工方案的智能检索推荐^[57]。然而，目前面向标准协同编制的相关平台和系统研究仍然停留在基于关键词的智能推荐阶段。检索增强生成（retrieval-augmented generation, RAG）融合了 LLM 强大的语言生成能力与高效的外部信息检索机制，可以有效提升 LLM 性能并应对诸如生成内容不实及缺乏领域知识等挑战。目前，已有研究者开始关注知识图谱与 LLM 之间的协同效应，探索如何将知识图谱中结构化的信息与 LLM 对自然语言的深度理解和创造性生成能力相融合，实现标准规范内容的智能检索应用^[58]。未来，需要进一步探索基于标准数字模型的智能检索通用技术路径。

此外，标准化人员未必胜任编写用 XML 等语言来描述的机器可读标准^[59]，协同编制平台需要为他们提供一系列标准编写模板，以支持机器可读标准条文的自动生成，目前这方面的探索相对较少。国家标准馆自主研发的标准智能编写系统可以根据用户输入的标准要素，生成符合排版规范的标准，但是其在实现标准条文内容的精准自动生成方面仍有局限^[60]。

建筑工程领域的协同设计平台也已经得到一定的发展，提供了全专业、全过程、全参与方的协同设计解决方案^[61]。但在面向领域标准的协同编制平台的研究和应用方面，仍然需要通过持续的探索创新，逐步建立起一套成熟、可复制的新型标准协同编制解决方案。

3.5 机器语言标准生成技术

机器语言即计算机可执行的代码。将数字解译技术处理后的自然语言条文内容转换成机器语言是实现标准的机器自动推理与高效校核的基础。基于 XML 格式的 XMLCheckset^[24]、面向本体的 SWRL 规则语言^[62]、语义网数据查询的 SPARQL 语言^[21]、逻辑编程语言 B-Prolog^[63]等是当前常用的标准条文机器语言表达方式。领域特定语言(domain specific language, DSL) 也可以用来表示规范条文，增强表达的专业性^[64]。但目前尚未形成统一的机器语言表示方法，如何面向不同的应用场景和规范条文实现灵活性的表达仍需进一步探索。

3.6 标准推理校核技术及应用

在工程建设领域，标准校核是确保建筑项目符合规范要求的关键环节，标准校核需要贯穿于建筑全生命周期，包括立项审批阶段建设许可证的颁发，设计阶段的智能审核，施工阶段的进度跟踪等。技术层面，标准的结构化存储和检索主要基于 GraphDB^[21]、Neo4j^[18]等图数据库，相关知识的推理与规则的校核则主要基于 Pellet^[65]、Jena Apache^[66]等推理引擎实现。同时，通过人工编写软件的内嵌规则或计算机语言代码，以实现各类标准校审应用也是当前常用技术方案之一^[67]。

标准校核推理的应用当前主要聚焦于设计、施工 2 个阶段。其中，设计阶段标准校核技术以智能审图或 BIM 模型审查为主，其目的是将建筑标准规则应用到施工方案、CAD 图纸、BIM 模型中，通过提取 BIM 模型、图纸中的信息并运行准解析形成可执行规则代码^[24,68]，以检查有关图纸、BIM 模型是否符合设计规范标准的要求。近年来，在我国各地智能审图政策的引导下，有关软件研发与工程应用进展迅速^[64]。同时，在文档审核方面，可通过有关算法准确识别不同格式、不同书写习惯的施工方案内容，通过抽取识别和抽取关键信息，模拟领域专家对内容完备性、关键表述等进行检查^[69]。施工阶段标准校核技术则往往与 BIM、

三维重建等技术相结合，如利用 BIM 模型和标准规范生成质量验收计划，基于三维点云和标准规范要求对构件几何信息进行自动检查等^[70]，以及通过智能合约自动执行在线仲裁或调解机制规则^[71]。但标准规则校核与 BIM、三维点云等多模态数据的融合尚处于初级阶段，仍以人工关联和映射为主^[72]，如何实现多模态数据融合的推理和计算仍是未来的重要探索方向。

4 未来需求及关键挑战

建筑业标准数字化的主要需求包括标准内容的自动解译、智能化检索、协同化编制与数字化应用等。其目的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，实现标准信息与建筑工程各环节的无缝集成与应用。这是一项复杂的系统工程，涉及技术、管理、应用等多个层面，面临诸多关键挑战。首要挑战在于当前仍缺少系统、完备的标准数字模型，标准数字化相关的语义常识库、标准条文库、机器规则库仍不完善，不同标准之间以及标准与 BIM 等多模态数据之间难以实现语义对齐与知识融合，标准数字模型是实现标准数字化的基础，这直接制约了标准数字化的发展；其次，当前我国建筑结构、给排水、电气、暖通、消防等不同领域的既有规范文本仅有约 1/3 的内容可以转换成机器语言标准，并被计算机自动解析和推理^[3]，实现既有标准内容的完整数字化解译仍面临诸多技术难题，尽管相关研究已取得一定进展，但在复杂条文的语义理解和规则转换方面仍需突破，这是支持多场景机器推理与人工智能应用的关键；最终，从标准生命周期管理的角度，标准数字化亟需优化标准的编写模式，重新审视和优化标准的编制流程与编制工具，实现人类可读自然语言与机器推理规则语言的协同生成，进一步将数字化标准应用于生成式智能设计、智能审图、智慧运维全流程，推动未来智能建造与人机协同发展。当前标准数字化过程中的需求和关键挑战可总结如下：

(1) 标准数字模型的构建与完善。首先，建筑业标准种类繁多，涉及设计、施工、材料、质量、安全等多个专业领域，标准体量大且不同标准在表示方式、组织结构、工程语义表达等方面存在差异，存在交叉引用、语义重叠等问题，如何实现跨专业、跨领域标准知识的统一表示、语义对齐与知识融合是当前关键挑战之一。其次，标准规范内容构成多样，包括文本、表格、公式甚至图集等不同类型的多模态数据，如何高效、准确地提取有关多模态知识，并实现相应知识的计算机统一表示仍需进一步探索。同时，机器人、三维重建、BIM 等技术的推广应用产生了海量的 BIM 模型、点云、图像、图纸等多模态数据，目前对于标准数据融合和对齐的研究仍然集中在文本数据层面，如何实现标准规范数字模型与相应多模态数据模型的语义协调与融合，对促进标准数字化发展有重要意义。

(2) 标准内容数字解译技术的发展与应用。首先，建筑业标准数字化是为了更好地服务于建筑工程的设计、施工、运维等环节，需要与 BIM、物联网、大数据、机器人等技术深度融合，满足不同系统在数据格式、接口标准等方面存在的差异，实现不同应用场景的数据交换和业务协同。其次，标准语义、标准数字模型、标准内容的数字解译技术、新型编制技术、机器语言生成技术、标准推理校核技术之间需要不断集成，实现标准数字化全流程的无缝对接和互操作。再次，亟需探索如何利用新兴 LMMs 技术提升标准规范实体识别、条文分类、语义对齐等环节的性能，甚至探索新型的端到端标准解译范式。此外，未来也需进一步提升计算推理引擎的能力，使其支持处理复杂逻辑推理、数学运算乃至复杂仿真推演，支持更加复杂、多样的工程场景。

(3) 数字化标准编制与管理新模式的探索。首先，建筑业标准化涉及多个部门和机构，存在职责分工不明确、管理流程不顺畅等问题，目前我国建筑业标准化管理体系还有待完善^[59]，因此需要优化现有标准编制流程，研发智能标准编制平台，探索基于生成式人工智能的条文智能解译与编写技术，引入用户交互反馈机制，实现高效的人机协同，不断提升标准内容的严谨性、全面性，建立新型标准编制模式。其次，建筑业标准随着技术进步和实践发展不断更新和完善，需要建立标准的动态更新机制，推广智能标准编制

平台的应用, 实现与传统工作方式之间的有效衔接和协调, 保证数字化标准的时效性。再次, 各种数字技术与建筑业的深度融合, 导致跨学科人才需求剧增, 需要打破学科界限, 完善奖励机制, 不断改革专业人才培养与团队建设模式。最终, 通过利用标准库中集成的机器可读语言, 可以为生成式智能设计^[73]、智能审图、智慧运维等应用场景提供丰富的领域知识和约束, 从而实现更加符合工程特性的建筑工程设计、审查、施工和运维, 促进智能建造发展。

5 总结与展望

建筑业作为国民经济的重要支柱, 面临转型升级的迫切需求。数字化技术的创新与应用为建筑业的高质量发展提供了新的动力。但当前标准规范的表现形式主要面向工程师人工解读与利用, 已难以满足人机协同场景的需求。

推动建筑工程标准数字化与智能化发展对整个行业发展具有重要的意义。因此, 本文系统性地提出了建筑业标准数字化发展阶段的分级模型及其典型应用场景, 为标准数字化与智能化发展提供了清晰路径。同时, 创新性地建立了覆盖存量标准智能解译、增量标准智能编制与标准多场景应用 3 个方面的建筑工程标准数字化关键技术框架, 包括标准数字模型、标准内容的数字解译技术、新型编制技术、机器语言标准生成技术、标准推理校核技术及应用 5 个组成部分。此外, 详细综述了框架各部分关键技术研究应用现状, 据此分 3 个发展阶段提出了标准数字化发展的未来需求及关键挑战。研究表明, 相关研究已在知识抽取、条文解译、协同设计、智能审查等方面已经取得了显著进展, 为标准的数字化解译提供了有力支持。但由于建筑业标准种类繁多, 覆盖领域广泛, 完全实现标准数字化转型尚需时日。

标准数字化的推进是一项系统工程, 实现标准制定到应用全流程的重塑需要多方协同攻关。当前, 标准数字化进程中面临的主要挑战包括标准数字模型构建的基础性挑战、标准内容数字解译技术的创新性挑战、标准编制与管理模式的应用性挑战 3 个方面。

为了克服这些挑战, 需要明确行业的未来发展需求和发展方向。首先, 标准数字化研究领域亟需进一步发展和完善多模态知识的抽取技术和表示方式, 构建统一的标准数字模型, 实现标准规范内容与多模态数据模型的语义深度融合。此外, 研究人员需要加强对标准数字化过程中关键解译技术的研究, 促进 LMMs 的应用, 提升标准的机器可读性和可执行性。最后, 研究人员需要探索数字化标准编制与管理新模式, 在研究过程中引入生成式人工智能技术, 促进人机协同发展, 不断应用和扩充标准库, 推进智能标准编制平台建设和多场景应用, 改革跨学科人才培养和团队建设模式。需要强调的是, 标准数字化不是完全替代人工, 而是通过人机协同提升效率。对于可以明确量化的、简单的指标要求, 计算机可以实现自动审查; 但对于需要工程经验判断、涉及多方案权衡的复杂决策, 仍需要工程师的专业把控。未来标准数字化应致力于充分发挥计算机在数据处理、规则校验方面的优势, 同时保留人工在关键判断环节的主导作用, 共同推动建筑业智能建造与人机协同的高质量发展。

参考文献 (References) (已修改)

- [1] 央视网. 智能建造引领建筑业转型升级 共绘“十四五”高质量发展新蓝图 [EB/OL]. (2022-02-26) [2024-06-25]. <http://finance.people.com.cn/n1/2022/0226/c1004-32360249.html>.
- [2] Craveiro F, Duarte J P, Bartolo H, et al. Additive manufacturing as an enabling technology for digital construction: A perspective on construction 4.0[J]. *Automation in Construction*, 2019, 103: 251-267.
- [3] Zheng Z, Zhou Y C, Chen K Y, et al. A text classification-based approach for evaluating and enhancing the machine interpretability of building codes[J]. *Engineering Applications of Artificial Intelligence*, 2024, 127: 107207.
- [4] 胡云忠. 基于本体的建筑施工质量规范知识建模与应用研究[D]. 武汉: 华中科技大学, 2013.
Hu Y Z. *Research on ontology-based knowledge modeling and application of construction quality specification*[D]. Wuhan: Huazhong

University of Science and Technology, 2013. (in Chinese)

- [5] 夏侯遐迹, 田丰华, 李启明. 智能建造背景下人机协作安全研究综述[J]. 东南大学学报(自然科学版), 2023, 53(6): 1053-1064.
Xiahou X E, Tian F H, Li Q M. Review of research on human-robot collaboration safety in the context of intelligent construction[J]. *Journal of Southeast University (Natural Science Edition)*, 2023, 53(6): 1053-1064. (in Chinese)
- [6] 马悦, 孙智, 武一奇, 等. 工程建设标准数字化应用场景与发展趋势研究[J]. 工程建设标准化, 2023(S1): 102-104.
Ma Y, Sun Z, Wu Y Q, et al. Research on digital application scenarios and development trend of engineering construction standards[J]. *Standardization of Engineering Construction*, 2023(S1): 102-104. (in Chinese)
- [7] 汪烁, 卢铁林, 尚羽佳. 机器可读标准——标准数字化转型的核心[J]. 标准科学, 2021(S1): 6-16.
Wang S, Lu T L, Shang Y J. Machine-readable standards—the core of digital transformation of standards[J]. *Standard Science*, 2021(S1): 6-16. (in Chinese)
- [8] ISO. ISO - IEC/ISO SMART[EB/OL]// (2024-04-22) [2024-06-12]. <https://www.iso.org/smart>.
- [9] 刘曦泽, 王益谊, 杜晓燕, 等. 标准数字化发展现状及趋势研究[J]. 中国工程科学, 2021, 23(6): 147-154.
Liu X Z, Wang Y Y, Du X Y, et al. Development status and trend of standards digitization[J]. *Strategic Study of CAE*, 2021, 23(6): 147-154. (in Chinese)
- [10] 国家市场监督管理总局国家标准技术审评中心. 全国标准信息公共服务平台[EB/OL]. [2024-06-25]. <https://std.samr.gov.cn/>.
- [11] 魏来, 黄爽, 李翔宇, 等. 标准数字化推动工程建设数字化迈上新台阶[J]. 工程建设标准化, 2023(9): 73-81.
Wei L, Huang S, Li X Y, et al. Engineering construction digitalization brought up to a new level by digitalization of standards[J]. *Standardization of Engineering Construction*, 2023(9): 73-81. (in Chinese)
- [12] 中华人民共和国住房和城乡建设部. GB 50016—2014 建筑设计防火规范[S]. 北京: 中国计划出版社, 2015.
- [13] Hjelseth E, Nisbet N. Capturing normative constraints by use of the semantic mark-up RASE methodology[C]//Proceedings of CIB W78-W102 Conference. Sophia Antipolis, France, 2011: 1-10.
- [14] Beach T H, Rezugui Y, Li H, et al. A rule-based semantic approach for automated regulatory compliance in the construction sector[J]. *Expert Systems with Applications*, 2015, 42(12): 5219-5231.
- [15] 刘吉明, 段立平, 林思伟, 等. 语义网赋能建筑信息交付及模型数据模式分析[J]. 土木与环境工程学报(中英文), 2024, 46(1): 244-253.
Liu J M, Duan L P, Lin S W, et al. Building information delivery and model schema analysis empowered by Semantic Web[J]. *Journal of Civil and Environmental Engineering*, 2024, 46(1): 244-253. (in Chinese)
- [16] 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 /Q, Li Y, Duan H, et al. Knowledge graph construction techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600. (in Chinese)
- [17] 刘江波. 地铁盾构施工规范知识图谱构建及应用研究[D]. 西安: 长安大学, 2023.
Liu J B. Construction and application of knowledge map of subway shield construction specification[D]. Xi'an: Chang'an University, 2023. (in Chinese)
- [18] 刘广宇, 安芃, 伍震, 等. 基于本体的公路工程安全领域知识建模和应用[J]. 清华大学学报(自然科学版), 2024, 64(2): 224-234.
Liu G Y, An P, Wu Z, et al. Ontology-based modeling and application of highway engineering safety knowledge[J]. *Journal of Tsinghua University (Science and Technology)*, 2024, 64(2): 224-234. (in Chinese)
- [19] Jiang Y K, Gao X, Su W X, et al. Systematic knowledge management of construction safety standards based on knowledge graphs: A case study in China[J]. *International Journal of Environmental Research and Public Health*, 2021, 18(20): 10692.
- [20] Xu X, Cai H B. Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure[J]. *Advanced Engineering Informatics*, 2021, 48: 101288.
- [21] Zheng Z, Zhou Y C, Lu X Z, et al. Knowledge-informed semantic alignment and rule interpretation for automated compliance checking[J]. *Automation in Construction*, 2022, 142: 104524.
- [22] Dimyadi J, Governatori G, Amor R. Evaluating LegalDocML and LegalRuleML as a standard for sharing normative information in the AEC/FM domain[C]// *Proceedings of the Joint Conference on Computing in Construction*. Heraklion, Greece, 2017: 637-644.

- [23] Zhang J S, El-Gohary N M. Semantic NLP-based information extraction from construction regulatory documents for automated compliance checking[J]. *Journal of Computing in Civil Engineering*, 2016, 30(2): 04015014.
- [24] Zhou Y C, Zheng Z, Lin J R, et al. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking[J]. *Computers in Industry*, 2022, 142: 103746.
- [25] Solihin W, Eastman C M. A knowledge representation approach in BIM rule requirement analysis using the conceptual graph[J]. *J. Inf. Technol. Constr.*, 2016, 21(Jun): 370-401.
- [26] Zheng Z, Chen K Y, Cao X Y, et al. LLM-FuncMapper: Function identification for interpreting complex clauses in building codes via LLM[EB/OL]. (2023-08-17)[2023-10-25]. <https://arxiv.org/abs/2308.08728v1>
- [27] 中国工程建设标准化协会. 协会标准《建筑信息模型工业基础类：通用》等 9 项标准编制组成立暨第一次工作会在京顺利召开 [EB/OL]. (2023-03-14) [2024-06-25]. <http://www.cecs.org.cn/xhzbz/bzdt/13743.html>.
- [28] 王珂, 杨芳, 姜杉. 光学字符识别综述[J]. *计算机应用研究*, 2020, 37(S2): 22-24.
- Wang K, Yang F, Jiang S. Overview of optical character recognition[J]. *Application Research of Computers*, 2020, 37(S2): 22-24.(in Chinese)
- [29] Sarkar M, Aggarwal M, Jain A, et al. Document structure extraction using prior based high resolution hierarchical semantic segmentation[M]//*Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2020: 649-666.
- [30] Shi B G, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(11): 2298-2304.
- [31] 补充作者. 工程建设标准编写规定[J]. *工程建设标准化*, 2008(6): 47-54.
- 补充作者. Provisions for the compilation of engineering construction standards[J]. *Standardization of Engineering Construction*, 2008(6): 47-54.(in Chinese)
- [32] 林佳瑞, 廖盘宇. 面向法规智能的消防规范图谱构建及应用初探[C]//第六届全国 BIM 学术会议论文集.太原, 2020: 5.
- LIN J R, LIAO P Y. Preliminary exploration of the construction and application of a fire protection regulation graph for intelligent regulation [C] // Proceedings of the 6th National BIM Academic Conference. Taiyuan, 2020: 5. (in Chinese)
- [33] Zhou Y C, Lin J R, She Z T. Automatic construction of building code graph for regulation intelligence[C]//*ICCREM 2021*. Beijing, 2021: 248-254.
- [34] Bartoli A, De Lorenzo A, Medvet E, et al. Inference of regular expressions for text extraction from examples[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(5): 1217-1230.
- [35] 胡海盟. 建筑工程质量验收规范知识建模与抽取研究[D]. 武汉: 华中科技大学, 2014.
- Hu H M. Research on knowledge modeling and extraction of construction engineering quality acceptance specification[D]. Wuhan: Huazhong University of Science and Technology, 2014. (in Chinese)
- [36] 李臻. BIM 技术在建筑设计质量检查中规范转译的方法研究[D]. 重庆: 重庆大学, 2019.
- Li C. Research on standard translation method of BIM technology in architectural design quality inspection[D]. Chongqing: Chongqing University, 2019. (in Chinese)
- [37] 舒赛. 支持图审的消防设计规范条文自动结构化方法[D]. 武汉: 华中科技大学, 2019.
- Shu S. Automatic structural method of fire protection design code provisions supporting drawing review[D]. Wuhan: Huazhong University of Science and Technology, 2019.(in Chinese)
- [38] Hassan F U, Le T. Automated requirements identification from construction contract documents using natural language processing[J]. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 2020, 12(2): 04520009.
- [39] Sun X, Li X, Li J, et al. Text classification via large language models[J]. arXiv preprint arXiv:2305.08377, 2023.
- [39] Sun X F, Li X Y, Li J W, et al. Text classification via large language models[EB/OL]. (2023-10-09)[2023-12-18]. <https://arxiv.org/abs/2305.08377v3>.
- [40] 刘鹏博, 车海燕, 陈伟. 知识抽取技术综述[J]. *计算机应用研究*, 2010, 27(9): 3222-3226.
- Liu P B, Che H Y, Chen W. Survey of knowledge extraction technologies[J]. *Application Research of Computers*, 2010, 27(9): 3222-3226. (in Chinese)

- [41] 邹佩聂. 面向知识抽取的 ModelOps 平台设计与实现[D]. 北京: 北京大学, 2023.
- Zou P N. Design and implementation of a ModelOps platform for knowledge extraction[D]. Beijing: Peking University, 2023. (in Chinese)
- [42] Hu Z Z, Leng S, Lin J R, et al. Knowledge extraction and discovery based on BIM: A critical review and future directions[J]. *Archives of Computational Methods in Engineering*, 2022, 29(1): 335-356.
- [43] Lample G, Ballesteros M, Subramanian S, et al. Neural architectures for named entity recognition[EB/OL]. (2016-04-07)[2023-06-09]. <https://arxiv.org/abs/1603.01360v3>.
- [44] 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 J X, Zhang X S, Wu C X, et al. Survey of knowledge graph construction techniques[J]. *Computer Engineering*, 2022, 48(3): 23-37. (in Chinese)
- [45] Zhang R C, El-Gohary N. A deep neural network-based method for deep information extraction using transfer learning strategies to support automated compliance checking[J]. *Automation in Construction*, 2021, 132: 103834.
- [46] Wu L T, Lin J R, Leng S, et al. Rule-based information extraction for mechanical-electrical-plumbing-specific semantic web[J]. *Automation in Construction*, 2022, 135: 104108.
- [47] Zheng Z, Lu X Z, Chen K Y, et al. Pretrained domain-specific language model for natural language processing tasks in the AEC domain[J]. *Computers in Industry*, 2022, 142: 103733.
- [48] Wang S H, Sun X F, Li X Y, et al. GPT-NER: Named entity recognition via large language models[EB/OL]. (2023-04-20) [2024-03-25]. <https://www.semanticscholar.org/reader/7c9f69848d28e0a7cbb00942ee83dab9773c23e4>.
- [49] Zhang J S, El-Gohary N M. Automated information transformation for automated regulatory compliance checking in construction[J]. *Journal of Computing in Civil Engineering*, 2015, 29(4): B4015001.
- [50] Zhou P, El-Gohary N. Semantic information alignment of BIMs to computer-interpretable regulations using ontologies and deep learning[J]. *Advanced Engineering Informatics*, 2021, 48: 101239.
- [51] Kim K, Kim H, Kim W, et al. Integration of ifc objects and facility management work information using Semantic Web[J]. *Automation in Construction*, 2018, 87: 173-187.
- [52] Zhu H, Xie R, Liu Z, et al. Iterative entity alignment via joint knowledge embeddings[C]//IJCAI. 2017, 17: 4258-4264.
- [53] 张君. 我国工程建设标准管理制度存在问题及对策研究[D]. 北京: 清华大学, 2010.
- Zhang J. *Study on the problems and countermeasures of China's engineering construction standard management system*[D]. Beijing: Tsinghua University, 2010. (in Chinese)
- [54] 陈观林, 侍晓龙, 周梁, 等. 基于深度强化学习的文本相似语义计算模型[J]. 郑州大学学报(理学版), 2020, 52(3): 1-8.
- Chen G L, Shi X L, Zhou L, et al. A text similarity semantic computing model based on deep reinforcement learning[J]. *Journal of Zhengzhou University (Natural Science Edition)*, 2020, 52(3): 1-8. (in Chinese)
- [55] 吴媛. 类比推理与法律智能案例推送系统研究[D]. 武汉: 华中科技大学, 2021.
- Wu Y. *Research on analogical reasoning and legal intelligence case push system*[D]. Wuhan: Huazhong University of Science and Technology, 2021. (in Chinese)
- [56] 中国建筑工业出版社. 住房城乡建设领域法规标准知识服务平台 [EB/OL]. (2024-04-16) [2024-06-18]. <https://www.kscecs.com/>.
- [57] 澎湃新闻·澎湃号·中建八局建筑工程施工方案人工智能审核系统完成内测 [EB/OL]. (2022-12-28) [2024-04-22]. https://www.thepaper.cn/newsDetail_forward_21347951.
- [58] Pan S R, Luo L H, Wang Y F, et al. Unifying large language models and knowledge graphs: A roadmap[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(7): 3580-3599.
- [59] 张佩玉, 裴继超, 邬贺铨: 标准数字化是大势所趋[J]. 中国标准化, 2022, (07): 24-29.
- Zhang P Y, Cao X X, Wu Hequan: Standards digitalization is the trend of the times[J]. *China standardization*, 2022, (07): 24-29.
- [59] 张佩玉, 裴继超, 邬贺铨: 标准数字化是大势所趋[J]. 中国标准化, 2022(7): 24-29.
- Zhang P Y, Pei J C. *Wu Hequan: Digitalization of standards is the general trend*[J]. *China Standardization*, 2022(7): 24-29.(in Chinese)

- [60] 中国标准化研究院. 国家标准馆[EB/OL]. (2022) [2024-06-25]. <https://write.nssi.org.cn/>.
- [61] 广联达科技股份有限公司. 广联达 BIM 设计协同平台[EB/OL]. [2024-06-25]. <https://gteam.glodon.com/>.
- [62] Fahad M, Bus N, Andrieux F. Towards validation of IFC models with IfcDoc and SWRL[C]//Proceeding of the 12th International Conference on Internet and Web Applications and Services. Venice, Italy, 2017: 25-29.
- [63] Zhang J S, El-Gohary N M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking[J]. *Automation in Construction*, 2017, 73: 45-57.
- [64] 林佳瑞, 周育丞, 郑哲, 等. 自动审图及智能审图研究与应用综述[J]. *工程力学*, 2023, 40(7): 25-38.
- Lin J R, Zhou Y C, Zheng Z, et al. Research and application of intelligent design review[J]. *Engineering Mechanics*, 2023, 40(7): 25-38. (in Chinese)
- [65] Chen Y J. Development of a method for ontology-based empirical knowledge representation and reasoning[J]. *Decision Support Systems*, 2010, 50(1): 1-20.
- [66] Zhang C, Beetz J, de Vries B. BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data[J]. *Semantic Web*, 2018, 9(6): 829-855.
- [67] Sobhkhiz S, Zhou Y C, Lin J R, et al. Framing and evaluating the best practices of IFC-based automated rule checking: A case study[J]. *Buildings*, 2021, 11(10): 456.
- [68] Zhang R C, El-Gohary N. Transformer-based approach for automated context-aware IFC-regulation semantic information alignment[J]. *Automation in Construction*, 2023, 145: 104540.
- [69] 中铁建设集团有限公司. 施工方案智能审核, 领跑行业智慧建造[EB/OL]. (2022-12-30) [2024-06-25]. http://ztjs.crcc.cn/art/2022/12/30/art_2256_4142758.html.
- [70] 马智亮, 蔡诗瑶. 基于 BIM 的建筑施工智能化[J]. *施工技术*, 2018, 47(6): 70-72, 83.
- Ma Z L, Cai S Y. Intelligent building construction based on BIM[J]. *Construction Technology*, 2018, 47(6): 70-72, 83. (in Chinese)
- [71] Mason J. Intelligent contracts and the construction industry[J]. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 2017, 9(3): 04517012.
- [72] 徐照, 康蕊, 孙宁. 基于 IFC 标准的建筑构件点云信息处理方法[J]. *东南大学学报(自然科学版)*, 2018, 48(6): 1068-1075.
- Xu Z, Kang R, Sun N. IFC-based point-cloud information processing method for structural elements[J]. *Journal of Southeast University (Natural Science Edition)*, 2018, 48(6): 1068-1075. (in Chinese)
- [73] 陆新征, 韩进, 韩博, 等. 基于规则学习与编码的剪力墙智能设计优化[J]. *东南大学学报(自然科学版)*, 2023, 53(6): 1199-1208.
- Lu X Z, Han J, Han B, et al. Intelligent structural design optimization for shear wall buildings based on machine learning and rule encoding[J]. *Journal of Southeast University (Natural Science Edition)*, 2023, 53(6): 1199-1208. (in Chinese)