



开放科学  
(资源服务)  
标识码  
(OSID)

# 融合大语言模型的标准知识图谱构建与应用 ——以医疗器械为例

周贞云<sup>1,2</sup> 金江威<sup>1,2</sup>

1. 杭州电子科技大学图技术与数字化联合实验室 杭州 310018;
2. 杭州电子科技大学中国科教评价研究院 杭州 310018

**摘要:** [目的/意义] 标准数字化浪潮席卷各行业, 针对现有标准知识图谱的技术手段已无法满足日益增长的知识处理需求, 亟须探索标准文献知识图谱与大模型等新兴技术结合的路径。[方法/过程] 以医疗器械为例, 通过对比各类型元数据、分析标准结构完成模式层构建, 融合规则与大语言模型完成数据层构建, 并结合大语言模型探究知识图谱应用。[局限] 在知识图谱的构建过程中, 对于重复知识以及需要进行消歧处理的部分, 缺乏高效的知识工作机制。[结果/结论] 提出含 29 类实体、19 类关系的本体, 提出融合规则与大语言模型的知识抽取方法, 构建了含 354548 个实体、434484 条关系的知识图谱, 并探究验证图谱的标准知识检索与分析功能, 展现出大语言模型融合标准知识图谱在提升标准数字化水平方面的独特优势。

**关键词:** 标准管理; 知识图谱; 知识组织; 大语言模型; 医疗器械

**中图分类号:** G353.1

## Standard Knowledge Graph Construction and Application Integrated with LLM: A Case Study of Medical Devices

ZHOU Zhenyun<sup>1,2</sup> JIN Jiangwei<sup>1,2</sup>

1. Joint Laboratory of Graph and Digital technologies, Hangzhou Dianzi University, Hangzhou 310018, China
2. Chinese Academy of Science and Education Evaluation, Hangzhou Dianzi University, Hangzhou 310018, China

**Abstract:** [Objective/Significance] As standard digitalization expands across industries, existing standard knowledge graph techniques struggle to meet growing demands, necessitating integration with emerging technologies such as the large language model (LLM). [Methods/Processes] Using medical devices as an example, this study constructs the pattern layer by comparing various types of metadata and standard structures, then combines rules and LLMs to build the data layer, exploring LLM-enhanced knowledge graph applications. [Limitations] The construction process lacks efficient mechanisms for handling

**基金项目** 浙江省科学技术厅技术开发项目“基于图技术的科教评价信息资源开发与管理”(KYH263123010M)。

**作者简介** 周贞云(1979-), 通信作者, 博士研究生, 副教授, 主要研究方向为图数据库与知识图谱, E-mail: zhouzhenyun@hdu.edu.cn; 金江威(1998-), 硕士研究生, 主要研究方向为图数据库与知识图谱。

**引用格式** 周贞云, 金江威. 融合大语言模型的标准知识图谱构建与应用——以医疗器械为例[J]. 情报工程, 2025, 11(5): 13-21.

repetitive or ambiguous knowledge. [Results/Conclusions] The proposed ontology includes 29 entity types and 19 relationship types. A hybrid rule-LLM extraction method builds a knowledge graph with 354548 entities and 434484 relationships. The functionality of the graph for standard knowledge retrieval and analysis was investigated and verified. Experiments confirm the unique advantages of combining LLMs with standard knowledge graphs in improving standard digitalization.

**Keywords:** Standards Management; Knowledge Graph; Knowledge Organization; Large Language Model; Medical Device

## 引言

标准是一组为维持各方利益与维持秩序而得到各方认可的关于产品、技术和工艺的特性及参数的规范<sup>[1]</sup>,是维持产品质量,保护用户,推动行业发展的必备因素,已成为规范和指导各领域健康发展的重要支撑<sup>[2]</sup>。当前标准化领域正经历以数字化为核心的深刻变革。作为标准数字化的优秀实践,标准知识图谱已成为标准数字化进程中的中坚力量,但其核心技术仍需持续优化<sup>[3]</sup>,在应对大语言模型等新技术应用时,存在显著的提升空间与优化需求<sup>[4]</sup>。

大语言模型的出现为标准知识图谱构建与应用提供了新的思路,得益于其出色的自然语言处理能力,语义分析能力和其独特的问答任务形式,提升了知识图谱构建与应用水平<sup>[5]</sup>。然而目前大语言模型在标准知识图谱领域的探索十分有限,尤其是本体与知识抽取框架都尚未针对大语言模型特性进行优化,严重阻碍了标准数字化的进程。

为丰富标准管理手段,提升标准数字化水平,本文以医疗器械标准为例,提出适配大语言模型的标准本体,融合大语言模型的知识抽取框架,结合大语言模型的应用范例。通过知识图谱与大模型融合研究,为各标准管理领域

提供参考,提升各行业标准数字化水平。

## 1 相关研究

### 1.1 标准数字化

标准数字化的本质在于运用云计算、大数据、区块链、物联网、人工智能等前沿数字技术,全方位赋能标准自身及其全生命周期<sup>[6]</sup>,旨在将标准所蕴含的规则与特性转化为数字设备能够高效读取、精准传输与有效应用的形式,进而实现标准化工作效率、透明度、协同性及智能化水平的全方位跃升<sup>[7]</sup>。标准数字化探索目前已积累了丰富且极具价值的典型案例。初步形成了技术工具、组织领导、标准数据、管理制度、人员能力五大能力域<sup>[8]</sup>。

虽然我国标准数字化进程已取得阶段性成效,但技术工具这一能力域的研发力度仍需持续强化<sup>[3]</sup>,尤其在应对人工智能、大数据等新技术应用时的适应性与稳定性方面,存在显著的提升空间与优化需求。

### 1.2 标准知识图谱与大语言模型

知识图谱概念于2012年由谷歌正式提出<sup>[9]</sup>。知识图谱作为一种知识表示形式,是一种包含实体、概念及其之间的各种语义关系的大规模

语义网络<sup>[10]</sup>。当前研究者们通过知识图谱有效细化了标准文献的知识结构,降低了学习难度,推动了标准文献的广泛应用,通过技术创新构建方法<sup>[11]</sup>、本体框架<sup>[12]</sup>、知识单元五元组<sup>[13]</sup>等,为标准数字化转型提供了重要的技术支撑和实践参考。然而目前方法在完成实体识别、知识抽取等任务时效果仍不理想<sup>[14]</sup>。大语言模型的出现,给相关问题带来了新的解法<sup>[15]</sup>。

在标准知识图谱中引入大语言模型,可有效赋能标准数字化。大语言模型的语义理解能力能自动解析标准文本中的术语、条款及隐含逻辑,而知识图谱则通过知识关联能力将知识转化为结构化语义网络,从而解决传统标准数字化方法中局限于文本数字化,却忽略知识数字化的桎梏。并借助大语言模型实现标准知识的智能应用,提升数字化知识的应用效率。知识图谱为大模型提供精准的标准知识约束,避免生成内容的随意性。而大语言模型则通过语义理解扩展图谱的覆盖范围,自动发现标准间的新关联<sup>[5,16]</sup>,两者融合有望提升标准数字化效率。

## 2 医疗器械标准模式层构建

构建模式层是利用模型和方法来表示、组织和处理知识的过程<sup>[17]</sup>。构建模式层的核心是创建一个能够清晰、准确地描述和表达知识的本体<sup>[18]</sup>,在知识建模过程中,本体提供了强大的概念表示能力和语义关系描述能力<sup>[19]</sup>,有助于建立结构化、层次化的知识模型。

### 2.1 标准本体要素

标准文献作为十大文献情报资源之一<sup>[20]</sup>,

具备其他文献情报资源通用的共性,也拥有自身的独特性。本文通过横向对比与采纳相关规范总结出标准本体的基本信息,通过分析标准文献特殊性,总结出标准的技术要素,综合考虑两者选取构建本体所需要素。

#### 2.1.1 基本信息

针对论文、书籍等各类科技文献数据,涌现了以 BIBO、FRBR 和 BIBFRAME 为代表的多种元数据或本体。三者均定义了作者、出版者、时间等核心元数据,核心公用属性节点围绕作品、实例、创作者、标题、出版信息、标识符等基础元数据展开并通过关系链接实体<sup>[21-23]</sup>。

《GB/T 22373-2021 标准文献元数据》在规定了类似核心元数据的基础上,附加了涵盖更丰富的描述信息的公共元数据,如中国标准分类号、国际标准分类号、起草单位、主要起草人、摘要、关键词、语种等。其中定义的元数据与其他文献元数据相比,增加了标准文献特有的要素,也保留了科技文献通用特性。考虑本文所采用的医疗器械标准文献的本体构建需紧密结合专业特性,依据我国监管要求,引入医疗器械基本信息,包括分类等级、预期用途和监管类型等,以关联标准与具体器械。

此类信息在《数据论文出版元数据》《GB/T 37967-2019 基于 XML 的国家标准结构化置标框架》等标准中被称为基本信息,本文沿用该称呼,引入相关节点作为本体中的标准基本信息,确立为本体建模的基础架构,为后续的知识建模与语义关联奠定了规范化基础。

#### 2.1.2 技术要素

当前各类文献元数据已通过编码语言对文献资源进行了基本的形式化处理,能为后续应

用提供部分语义基础。然而完全套用当前元数据却无法充分发挥知识图谱灵活性与可扩展性等能力<sup>[24]</sup>，也不能进一步挖掘标准中的知识，限制了标准知识数字化发展。

本文结合知识图谱特点和应用需求，进一步分析标准文献特征，归类出文献层次、核心技术要素、其他技术要素三类实体，以扩充标准文献本体内容。标准文献的文献层次主要分为章和条两级，便于精准定位内容。核心技术要素是表述标准特定功能的要素，按照《GB/T 1.1-2020 标准化工作导则 第1部分：标准化文件的结构和起草规则》划分为术语条目、试验步骤、要求等九类。其他技

术要素包括术语、范围等标准规定内容。各类技术要素蕴含在文本中，难以简单提取，往往不纳入元数据或本体中。然而随着数字技术发展，得益于大模型的深度语义理解能力，可通过其精准识别并提取此类要素，进而丰富标准知识图谱内容，提升知识图谱的实用性与智能化水平。

## 2.2 标准本体构建

参考各类文献元数据，综合考虑标准特殊性与知识图谱应用实际，本文从基本信息和技术要素出发，并补充实体间关系，构建医疗器械标准本体，如表1所示。

表1 医疗器械标准本体

类别	小类	实体名称	实体名称	
基本信息	标准基本信息	标准	时间	
		人员	国际标准文献分类号	
		单位	中国标准文献分类号	
		摘要	状态	
		医疗器械	医疗器械一级分类	
	医疗器械基本信息	产品描述	产品大类	
		预期用途	标准类型	
		医疗器械二级分类	监管类型	
		文献层次	章节	条款
		技术要素	核心技术要素	分类/编码
要求/证实方法	术语条目			
符号/标志及其含义	追溯/证实方法			
试验步骤/数据处理	程序指示			
需考虑的因素				
其他技术要素	中文术语	范围		

该本体构建了一个全面且深入的知识体系，提炼了医疗器械标准中的多维知识精髓，能够精准捕捉标准中的基本信息与技术要素，以及

其特有的引用、替代、采用等关系。通过细致的分类与汇总，明确了知识类型，为后续的知识抽取工作奠定了坚实基础，也为大语言模型

的介入提供了快速定位和深入理解医疗器械标准的有效途径。

### 3 医疗器械标准数据层构建

数据层构建涉及对文献内容的深度理解与结构化处理，必须确保知识在数据层中的准确性和完整性，减少过度分割以及抽取层次不够等问题。

#### 3.1 基于规则的显性知识抽取

基本信息实体与文献层次实体可从标准封面页等来源，使用简单方法直接提取，此类信息提取较为容易，通常具有固定的格式和位置，可通过对标准内相关信息进行定位提取。

本文针对显性信息的抽取，采用了基于正则表达式的规则匹配方法。正则表达式是一种文本匹配工具，能够通过定义特定的模式，从文本中快速定位和提取目标信息。标准编号通常以“GB/T”或“YY”开头，章节标题通常以“第x章”或“x.x”的形式出现，本文通过re模块，编写正则表达式“GB/T\d+-\d+”和“YY\d+-\d+”，能够从文本中快速提取出符合格式的标准编号。对于章节标题和条款内容，本文通过正则表达式“第x章”和“x.x”进行定位，并提取相应的文本内容。其他显性知识通过类似方式进行处理提取。

#### 3.2 基于大语言模型的隐性知识抽取

以技术要素为代表的隐性知识以句或段作为最小单位保存，无法直接通过标准内容提取。本文提出的隐性知识抽取方法采用分阶段提示词工程框架，通过角色定位、格式约束、动态调整三

阶段的协同作用实现知识结构化输出，提高数据层构建自动化程度，提升构建效率和准确性。

在角色定位阶段，通过角色扮演机制模拟领域专家，解决大模型在隐性知识抽取中实用性不足的问题。为确保GLM-4模型准确理解任务需求并输出符合预期的结果，在隐性知识抽取任务中，通过系统提示为模型注入背景，使大语言模型的通用能力与领域专家的认知模式对齐，满足知识抽取的实际任务需求。在格式约束阶段，基于模板约束格式，将非结构化输出转化为方便后续处理的规范化表示，为后续知识图谱构建提供可靠输入。强制模型按照预定模式组织信息，能有效抑制大模型的幻觉生成。在动态调整阶段，手动优化知识抽取策略。首先通过小样本实验数个典型抽取过程，并根据反馈结果，不断在提示词尾端增加要求，不断修正提示词。明确的角色定义和格式要求大幅降低了模型输出结果的随机性，而定制化提示使模型输出更加稳定、可靠，灵活调整提示内容，进而调整抽取的粒度和范围。

通过这种分步骤、分任务的提示词工程方法，本文成功将大模型的通用能力聚焦于特定的知识抽取任务。三个层面的协同作用，形成了具有领域迁移性的框架，对标准文献领域和相关法规具有较好提取效果。

#### 3.3 知识抽取结果

通过针对不同知识制定不同提取方法，对医疗器械标准进行实践后，最终本文成功构建包含354548个实体，434484条关系的医疗器械标准知识图谱。各类型实体与关系数量如表2、表3所示。

表 2 数据层实体抽取结果

实体名称	计数	实体名称	计数
条款	120384	范围	2592
分类 / 编码	90203	单位	2323
要求 / 证实方法	33868	产品描述	1548
符号 / 标志及其含义	25274	预期用途	1461
章节	11766	医疗器械二级分类	1151
试验步骤 / 数据处理	10201	摘要	1006
需考虑的因素	8589	时间	338
中文术语	7921	医疗器械一级分类	205
标准	6512	国际标准文献分类号	157
医疗器械	5948	中国标准文献分类号	42
程序确立	5779	产品大类	22
术语条目	5154	状态	5
人员	5055	标准类型	4
追溯 / 证实方法	4138	监管类型	3
程序指示	3182		

表 3 数据层关系抽取结果

关系名称	计数	关系名称	计数
具备	193165	发布于	2973
含	150053	处于	2973
起草	25405	归口	2731
定义	12128	作用于	2592
属于	11464	替代	1008
具有	11115	强制适用	794
分类	8669	归于	205
收到	3142	废止于	101
引用	2990	采用	3
实施于	2973		

## 4 标准知识图谱应用

随着技术的飞速发展，基于大模型的科技文献检索已取得了显著成果<sup>[25]</sup>，知识图谱的应用方法也面临着新的要求。各种图相关技术的融入，正在引导着标准知识图谱应用产生质的

飞跃。本文总结并探索医疗器械标准知识图谱的知识检索与知识分析两大主要应用。

### 4.1 知识检索

在当今信息爆炸的时代，如何从海量的知识中快速、准确地获取所需信息<sup>[26]</sup>，成为知识

图谱应用领域的关键课题。知识图谱结合大语言模型的知识检索系统，可借助大语言模型检索知识图谱，以实现智能化的标准知识检索，提高知识检索系统的效率和准确性。

本文采用 Py2neo 图形数据库管理工具以及 GLM-4 大语言模型，共同构建了一个高效

的知识检索系统。图 1 展示了向系统提出具体问题后，系统返回的检索结果。图 1 (a) 展示了知识检索系统针对该问题的回答，图 1 (b) 则是直接向大语言模型提问得到的答案，而图 1 (c) 则是在传统搜索引擎上检索得到的结果。

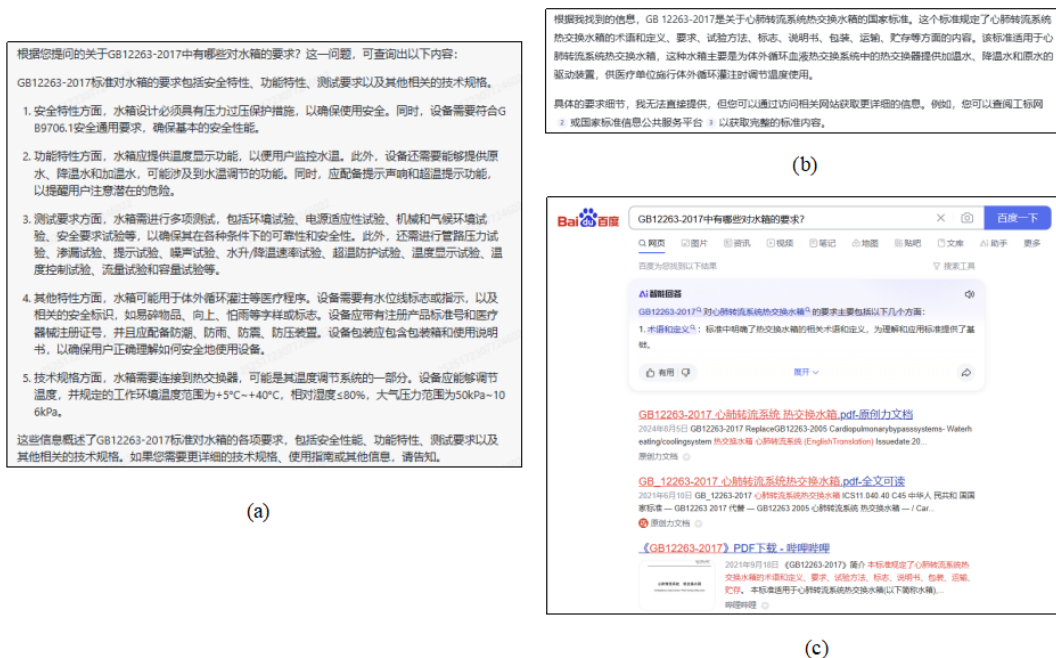


图 1 标准问答系统效果图与人工结果查询对比图

图模互补的知识检索方式在信息处理和呈现上展现出了显著的优势，不仅能够提供精准、全面的知识检索服务，还能够通过知识图谱的视觉化和交互性，提升用户的检索体验，满足用户对于高质量医疗器械标准知识检索的需求，无疑为医疗器械标准文献的查询和研究开辟了新的路径，具有广泛的应用前景。

## 4.2 知识分析

在医疗器械标准知识图谱的应用中，衍生出诸多具有重要学术价值与实践意义的典型应

用，如知识推理<sup>[27]</sup>、跨标准知识融合<sup>[5, 27]</sup>、知识冲突检测等，不仅推动了医疗器械标准知识的系统化与智能化管理，还为标准的制定、修订和实施提供了科学依据<sup>[13]</sup>。

本文以知识冲突检测为例对中文术语制造商进行知识分析功能验证，利用大语言模型进行分析后，得到结果如图 2 所示。

该例中大语言模型结合知识图谱的分析方法在知识分析领域具有显著的优势，相比直接进行人工分析，大模型总结后的内容更为精炼易懂。尤其是针对不同标准中复杂的

术语和定义,能够准确提取关键信息并识别语义差异。除本文展示的术语外,此种方法也可对条款、结构、定义对象等方面进行挖掘,大模型融合知识图谱凭借其强大的数据

处理和分析能力,能够在短时间内完成大量标准知识的分析与展示<sup>[28]</sup>,减轻标准制定者和使用者的工作负担,为知识创新与发展创造了良好条件。

不同标准中制造商的定义存在以下差异:范围和职责细节方面有些定义更侧重于医疗器械上市或投入服务前的相关活动(如YY\_T0316-2008@2.8、YY\_T0316-2016@2.8等定义中提到上市和/或投入服务前对医疗器械的设计、制造、包装或作标记等负有责任),而另一些则进一步详细阐述了包括设计、制造、包装、标记以及系统装配或改装等更广泛的活动范围(如GB\_T29791.1-2013@3.36),还明确指出了制造商对确保符合法规要求的最终法律责任以及上市前和上市后的其他责任(如GB\_T42061-2022@3.10、GB\_T42062-2022@3.9等)。部分定义(如GB\_T20012-2005@3.2.2)只是简单地说制造商是生产设备的组织或个人,相比其他详细阐述职责的定义,范围和职责的明确性较低。引用来源和基础方面不同的定义可能引用了不同的国际标准或指南文件作为基础,例如GB\_T42062-2022@3.9、GB\_T42218-2022@3.9等引用了GHF/SG1/N055:2009、ISO14971:2019等文件,这导致了定义在具体内容和侧重点上可能有所不同。有些定义可能是在其他现有定义的基础上进行了扩展或修改,如GB9706.1-2020@3.55提到从YY\_T0316-2016定义2.8中变化而来。

图2 术语制造商定义大模型分析结果

## 5 结语

本文构建一种含29类实体、19类关系的本体,提出融合规则与大语言模型的知识抽取方法,形成了含354548个实体、434484条关系的知识图谱,并探究验证基于此图谱的标准知识检索与分析功能,提升了标准的可检索性和可读性,挖掘出医疗器械标准中的隐性知识,提升了标准知识图谱与大语言模型的契合度,为其他领域标准知识图谱的构建和应用提供了参考。

当前工作仍存在一定改进空间,在知识图谱的构建过程中,对于重复知识以及需要进行消歧处理的部分,缺乏高效的知识工作机制。一些在表述上存在明显差异但本质指向相同或相近含义的概念,给知识抽取与应用带来了挑战。未来可考虑利用大语言模型建立一套完善的消歧体系,对概念进行有效梳理、挖掘、匹配和合并。

## 参考文献

- [1] 吕铁.论技术标准化与产业标准战略[J].中国工业经济,2005(7):43-49.
- [2] 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.
- [3] 马超.中国标准数字化转型:认知阐释、现实问题及发展路径[J].图书与情报,2023(4):50-63.
- [4] 陈守清.标准文献知识的管理与服务亟待创新[J].标准科学,2015(6):63-66.
- [5] 黄勃,吴申奥,王文广,等.图模互补:知识图谱与大模型融合综述[J].武汉大学学报(理学版),2024(4):397-412.
- [6] 王若雅,王益谊,刘曦泽,等.我国标准数字化发展进程及成效分析[J].信息技术与标准化,2024(8):16-21.
- [7] 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.
- [8] 许海清,王思宁,魏志丰,等.电网企业标准数字化能力成熟度模型构建研究[J].电信科学,2024,40(8):138-148.
- [9] 周贞云,邱均平.一门交叉学科的兴起:论图数据科学的构建[J].图书馆论坛,2023,43(4):97-108.
- [10] 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.
- [11] 郝文建, 魏梅, 张浩, 等. 标准知识图谱的构建与应用[J]. 信息技术与标准化, 2021(8): 44-47.
- [12] 赵青青. 基于实体-关系模型的标准元数据关系研究[J]. 标准科学, 2023(3): 16-20, 66.
- [13] 许平. 标准文献知识链接构建与应用研究[J]. 中国质量与标准导报, 2023(4): 83-87.
- [14] PENG C, XIA F, NASERIPARSA M, et al. Knowledge graphs: opportunities and challenges[J]. Artificial Intelligence Review, 2023, 56: 13071-13102.
- [15] 李阳, 孙建军. 大模型对情报学发展的影响思考[J]. 情报学报, 2025, 44(2): 246-256.
- [16] 张君冬, 刘江峰, 邓景鹏, 等. 图模驱动的在线医疗健康智慧问答服务研究[J]. 现代情报, 2025, 45(1): 164-176.
- [17] 尹亮, 何明利, 谢文波, 等. 装备-标准知识图谱的过程建模研究[J]. 计算机科学, 2018, 45(S1): 502-505.
- [18] 田玲, 张谨川, 张晋豪, 等. 知识图谱综述——表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
- [19] 王向前, 张宝隆, 李慧宗. 本体研究综述[J]. 情报杂志, 2016, 35(6): 163-170.
- [20] 张国华. 标准文献是一种重要情报源[J]. 中国标准化, 2006(7): 66-69.
- [21] 鲜国建, 赵瑞雪, 孟宪学, 等. 基于知识组织体系的多维语义关联数据构建研究[J]. 数字图书馆论坛, 2014(3): 11-18.
- [22] 刘素清. IFLA 书目记录功能需求(FRBR)初探[J]. 大学图书馆学报, 2004(6): 65-69.
- [23] 胡小菁. BIBFRAME 核心类演变分析[J]. 中国图书馆学报, 2016(3): 20-26.
- [24] 李金定. 叙词表、元数据与本体之间关系探究[J]. 图书馆学研究, 2007(8): 61-64.
- [25] 温有奎, 温浩, 乔晓东. 让知识产生智慧——基于人工智能的文本挖掘与问答技术研究[J]. 情报学报, 2019, 38(7): 722-730.
- [26] 姚长青, 程齐凯, 王莉军, 等. 智能情报技术: 内涵、边界与体系[J]. 情报学报, 2025, 44(1): 1-9.
- [27] 徐松林. 大模型与标准文献知识库的融合应用探索[J]. 上海信息化, 2024(7): 31-35.
- [28] 陶玥, 余丽, 吴振新. CoTransH: 科技文献知识图谱中语义关系预测的翻译模型[J]. 情报理论与实践, 2021, 44(11): 187-196.