

引用格式: 金彦名,余晓,黎燕.产业链视角下标准知识图谱的构建与应用研究[J].标准科学,2025(5):95-100.

JIN Yan-ming,YU Xiao,LI Yan.Research on the Construction and Application of Standard Knowledge Graphs from an Industry Chain Perspective [J].Standard Science,2025(5):95-100.

产业链视角下标准知识图谱的构建与应用研究

金彦名¹ 余晓^{1*} 黎燕²

(1.中国计量大学 经济与管理学院; 2.中国质量标准出版传媒有限公司)

摘要:【目的】为推动标准数字化转型,解决传统标准管理中信息分散、应用效率低等问题,本文提出并探讨标准知识图谱的构建方法及其应用。【方法】提出标准知识图谱的构建原则,设计了数据收集与预处理、本体设计、知识抽取、知识存储及可视化展示的五步构建流程。以石油产业为例,基于Python和Neo4j构建标准知识图谱。【结果】构建了石油产业标准知识图谱,展示了标准间的层次结构和关联关系。用户可通过标准知识图谱快速查询特定标准及其关联信息。【结论】从产业链视角出发,构建标准知识图谱,可全面反映标准体系在整个产业链中的集成性,促进产业链各环节的协同发展。以石油行业为例,验证了所提出方法的有效性,为相关领域标准知识图谱的研究提供了新的思路。

关键词: 标准; 知识图谱; 标准数字化; 石油产业链

DOI编码: 10.3969/j.issn.1674-5698.2025.05.014

Research on the Construction and Application of Standard Knowledge Graphs from an Industry Chain Perspective

JIN Yan-ming¹ YU Xiao^{1*} LI Yan²

(1.College of Economics and Management, China Jiliang University; 2.Standards Press of China)

Abstract: [Objective] In order to promote the digital transformation of standards and solve the problems of information dispersion and low application efficiency in traditional standards management, this paper proposes and discusses the construction method and application of standard knowledge graph. [Methods] The construction principles of standard knowledge graphs are proposed, and a five-step construction process of data collection and preprocessing, ontology design, knowledge extraction, knowledge storage and visual display are designed. Taking the oil industry as an example, a standard knowledge graph is built based on Python and Neo4j. [Results] A petroleum industry standard knowledge graph is constructed, showing the hierarchical structure and correlation between standards. Users can quickly query specific standards and their associated information through the standard knowledge graph.[Conclusion] This paper constructs a standard knowledge graph from the industry chain perspective, highlighting system integration and promoting collaboration across its segments. The method is validated in the petroleum industry, providing new insights for related research.

Keywords: standard, knowledge graph, standard digitization, the petroleum industry chain

基金项目: 本文受国家社会科学基金项目“新的国际背景下我国制造业突破技术标准锁定的策略研究”(项目编号: 24BJY030)资助。

作者简介: 金彦名,硕士研究生,研究方向为标准化。

余晓,通信作者,博士,教授,研究方向为标准与创新管理。

黎燕,硕士,副编审,研究方向为标准信息化。

0 引言

随着信息技术的快速发展,传统的纸质标准、PDF格式标准已无法满足新兴数字经济的需求。标准体系的不断扩展、制修订频率的提高及标准检索效率低下等问题给标准领域带来诸多挑战。因此,为了提高标准的应用效率和响应速度,适应现代科技发展的新需求,标准的数字化转型已成了迫切需要解决的问题。

中共中央、国务院印发的《国家标准化发展纲要》中明确提出要发展机器可读标准、开源标准,推动标准化工作向数字化、网络化、智能化转型^[1]。此外,许多国际和区域标准化组织已经开始推进标准数字化相关工作。例如:国际标准化组织(ISO)提出“SMART”标准的概念,倡导将标准从传统的纸质形式向“机器可读标准”(Machine Applicable, Readable and Transferable)转型,推动标准数字化进程^[2]。SMART标准的提出标志着标准化工作不仅仅限于文本规范,更向着能够自动处理、智能分析和跨系统应用的数字化目标迈进。欧洲标准化委员会(CEN)、欧洲电工标准化委员会(CENELEC)发布的《CEN-CENELEC数字化转型战略计划》希望通过标准数字化来确保工业领域数字化转型的标准化需求得到满足^[3]。由此可见,标准数字化已成为国内外标准领域关注的焦点。目前,我国正处于发展机器可读内容阶段,这一阶段主要是将标准组织成知识图谱,进行标准剪裁,形成语义网络。

知识图谱的概念最早由谷歌提出^[4],其作为一种新型的知识表示和管理工具,在近年来得到了广泛的应用。它通过图结构表示实体及其关系,能够有效地将知识进行组织和存储。在金融风险^[5]、医疗智能问诊^[6]等领域,知识图谱技术已经取得了显著的应用成果。由于传统的标准文档通常以纸质或PDF文本的形式存在,这种信息碎片化导致标准之间难以有效关联,无法全面展示标准之间的逻辑关系和适用场景,而知识图谱技术恰能解决这一问题。目前相关研究已在食

品安全^[7]、应急管理^[8]等领域构建相关标准知识图谱并采取应用。然而,现有研究主要针对局部视角,如油田^[9]、油气管道^[10]等具体领域的相关研究缺乏系统性的视角,也难以全面反映标准体系在整个产业链中的集成性和协调性。因此,本文从产业链视角出发,以石油产业为例,构建标准知识图谱,采用Python编程语言及其相关工具包,利用Neo4j图数据库对标准知识图谱的存储和管理,并通过Neo4j实现图谱的可视化展示及关联查询功能,为推动行业标准的数字化转型提供理论支撑,进而促进产业链环节的协同发展。

1 标准知识图谱构建

1.1 构建原则

为确保标准知识图谱的质量和可用性,在构建标准知识图谱的各个环节中应遵循3个核心原则:准确性原则、语义明确性原则、可扩展性原则^[11]。遵循准确性原则是标准知识图谱构建的基本要求,直接影响知识图谱的可靠性和用户信任度。它可以确保图谱中所有数据和知识关系的真实性与精确性。因此,在数据收集阶段,需要严格筛选数据源,以确保所采集的原始信息权威且真实。对于标准知识图谱而言,应优先选择国际标准化组织(ISO)、国家标准化机构及其他权威机构发布的正式文件信息作为数据来源。在数据预处理过程中,应对数据内容进行全面的校验,剔除冗余、错误的数据,以保证数据的准确性,从而提高数据质量。其次,在本体结构设计时,需精确定义领域内的核心概念、属性及其关系,从而避免概念间可能产生的误解或错误关系。语义明确性原则强调知识图谱中节点和关系的语义表达应清晰、无歧义。因为构建标准知识图谱的目的不仅仅是呈现特定领域知识的概念结构,还需对概念间的语义关系进行直观和明确的表达,所以遵循语义明确原则对用户理解和应用标准知识图谱尤为重要。可扩展性原则着眼于标准知识图谱的动态更新能力,确保其能够适应未来数据的扩充与变化。标准知识图

谱并非静态的系统,而是一个需要持续更新、动态维护的知识库。随着相关标准的修订、废止和新标准的发布,图谱中的数据和关系也需及时调整以反映最新的领域知识。

1.2 构建方法

构建标准知识图谱的流程一般包括:数据收集与预处理、本体设计、知识抽取、知识存储及结果可视化展示^[12]。

数据收集是构建标准知识图谱的基础,其主要目的是明确目标领域的用户需求并收集与标准相关的数据信息。对于标准知识图谱,主要涉及国际和国家标准、行业标准,以及企业标准等相关数据。这些数据来源多样,既包含结构化数据,如从数据库中导出的数据、表格信息等,又包含非结构化数据,如PDF文档等其他文本信息。结构化数据是指内容统一、格式整齐的数据,这类数据通常可直接用于构建标准知识图谱。然而在实际构建过程中,收集到的大部分数据为非结构化数据,而目前标准文件多以PDF格式为主,这就是一种非结构化的数据,这种数据形式不易被计算机直接识别和理解。因此需要对其进一步处理和结构化转化,使其成为结构化数据。在数据预处理阶段,需对原始数据进行清洗和规范化操作,包括但不限于去除重复数据、纠正错误数据、填补缺失信息及过滤无关内容。通过这一阶段可以提高数据质量,为后续环节提供准确和完整的基础数据。此外,预处理阶段还可能涉及自然语言处理(NLP)技术,以进一步规范和组织非结构化数据。

本体设计是标准知识图谱构建的核心环节,本体是一种用于描述特定领域中概念、属性及其关系的形式化模型,能够为知识图谱提供结构化的逻辑框架。首先,明确领域中的基本概念,例如标准文件中的必要要素:标准名称、标准号、起草单位、归口单位等。属性是概念的特征体现,包括数据属性和对象属性。数据属性是指概念本身具有的属性,如标准的发布时间、实施时间、标准的状态等;对象属性是用来描述概念之间的具体关系,如起草人与起草单位之间的从属关系“起草人

属于起草单位”,标准文件之间的引用关系、替代关系等。目前本体设计的方法主要有3种:手工设计、半自动设计和自动设计^[13]。手工设计适合于规模较小、结构较清晰的领域;半自动化设计结合了人工与机器的优点,进而对本体进行设计;自动化设计则依赖于机器学习和语义分析技术,适合于处理海量数据,但这种设计方法对于算法要求较高。

知识抽取是将非结构化数据转化为计算机能够识别并理解的结构化数据的关键过程,主要包括实体抽取、关系抽取和属性抽取。在这一阶段可以结合机器学习相关方法来提高抽取效率。

将结构化数据存储于图数据库中是构建标准知识图谱的关键步骤。当前,主流的图数据库有Neo4j、GraphDB等。它们能够有效管理和查询以节点和边形式表示的知识数据。选用图数据库时需要综合考虑数据规模、查询性能及系统扩展性。将数据存储后对相应结果进行展示,标准知识图谱的可视化展示旨在直观地呈现出各标准之间的关联关系,从而帮助用户对标准体系有更好地了解 and 认识。

2 石油标准知识图谱的构建与应用

本文以石油产业为例构建标准知识图谱。首先,将石油产业链划分为上、中、下游环节。其中,上游环节包括油气的勘探、开发和生产,涉及的标准有钻采、石油工程设计、石油地质、石油管材、采油采气、油气田开发、石油工程施工、石油钻井工程等。中游环节主要集中在原油、天然气的加工和炼化,涉及的标准有油气储运、石油安全、石油物资采购、石油节能等。下游环节是指石油化工产品的深加工和最终消费,涉及的标准有石油天然气计量、专业计量器具检定规程和校准方法、油田化学等。具体划分情况如图1所示。

本文所用数据来源于全国标准信息公共服务平台、标准网,在平台及网站中检索与石油产业相关标准构成原始数据集,本文共收集相关标准13,656条。再依据石油产业链环节设计本体结构。

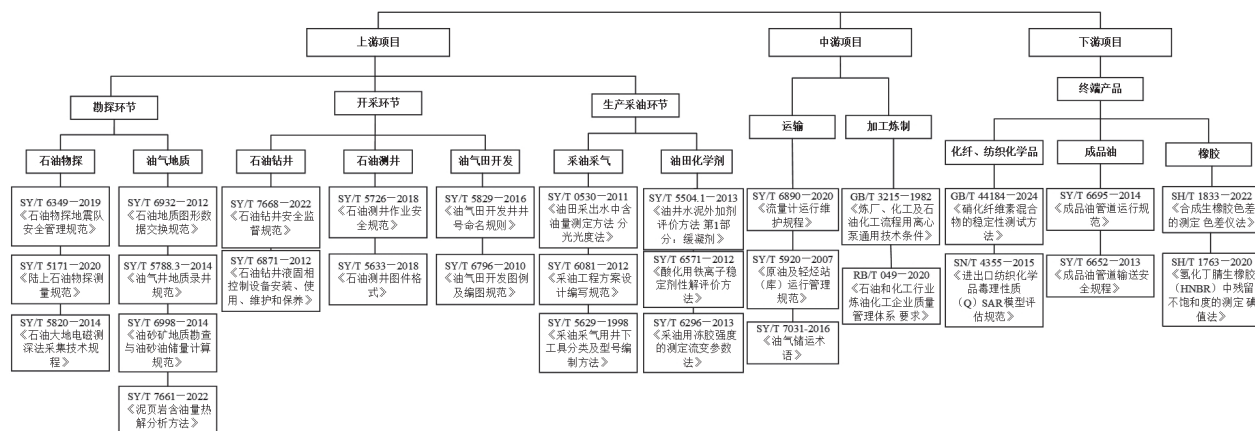


图1 石油标准分类

本体结构设计需结合行业特点和数据需求,确保图谱中的实体和关系能够准确描述标准知识的语义。本文依照行业惯例将石油产业链分为上游、中游和下游三大项目环节,各部分又包含多个细分环节。上游主要涉及勘探、开采和生产采油;中游涵盖运输和加工炼制;下游则聚焦于石油化工产品的深加工和市场消费。基于这一划分,明确以下核心概念:定义“上游”“中游”“下游”3个一级概念,并基于其细分环节进一步扩展,如“勘探”“开采”“生产采油”“运输”等二级概念。关系的定义主要包括从属关系(如某标准适用于勘探环节,属于上游项目标准),以及时间关系(如某标准何时发布)。基于以上设计本体结构图如图2所示。



图2 本体结构框架

通过知识抽取技术将所需的实体和关系进行提取,并以图数据库的形式进行存储,本文采用Python中的Py2neo工具包把数据存储在Neo4j中,存储的核心结构为“实体-关系-实体”的三元组

形式。通过Neo4j对石油产业链上游和中游环节进行可视化展示。上游环节共创建25,001个节点,100,034条关系,可视化结果如图3所示。中游环节标准知识图谱可视化结果如图4所示。下游环节标准知识图谱可视化结果如图5所示。



图3 上游环节标准知识图谱可视化展示

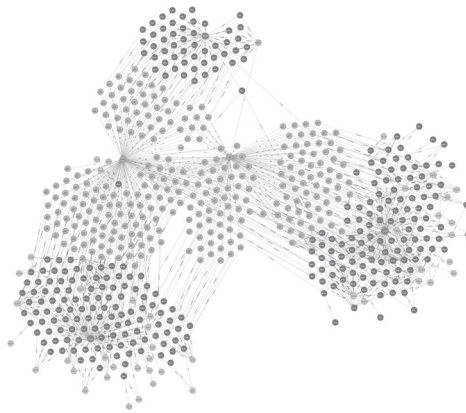


图4 中游环节标准知识图谱可视化展示

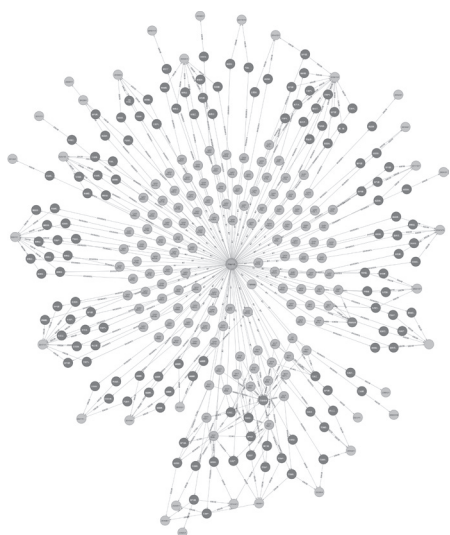


图5 下游环节标准知识图谱可视化展示

标准知识图谱还可提供相关查询功能。主要包括查询某个节点、查询某个节点的所有关系和查询指定类型的关系这3类功能。接下来分别举例说明这3类查询功能的具体应用：

(1) 查询某个节点：查询《石油天然气工程总图设计规范》这个标准，在Neo4j中输入“MATCH (n:标准名称 {name: '石油天然气工程总图设计规范'}) RETURN n”代码，查询结果如图6所示。基于这个查询结果还可查看当前节点的相关信息。



图6 《石油天然气工程总图设计规范》查询结果

(2) 查询某个节点的所有关系：查询与石油管材相关的所有关系，在Neo4j中输入“1、MATCH (n:三级产业链环节 {name: '石油管材'})-[r]-(related) 2、RETURN n, r, related”代码，即可展示与石油管材相关的所有标准，如图7所示。还可在操作栏左侧切换Text模式即可显示文字版本，如图8所示。

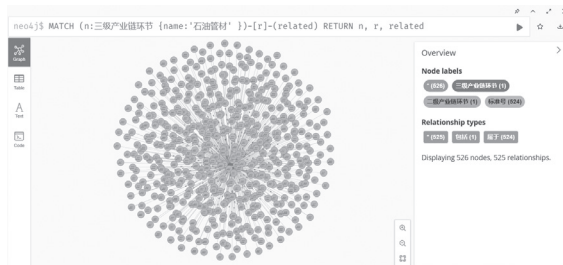


图7 石油管材的所有关系查询结果

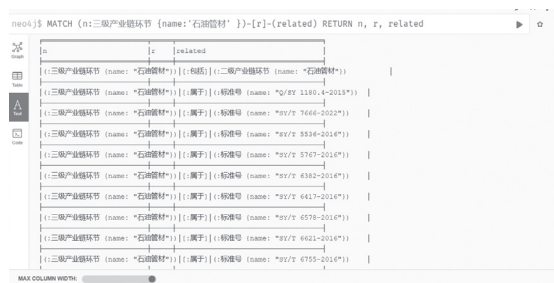


图8 石油管材的所有关系Text模式查询结果

(3) 查询指定类型的关系：查询二级产业链环节中勘探环节包括哪些关系节点，在Neo4j中输入“1、MATCH (n:二级产业链环节 {name: '勘探环节'})-[r:包括]->(related) 2、RETURN n, r, related”代码，查询结果如图9所示。

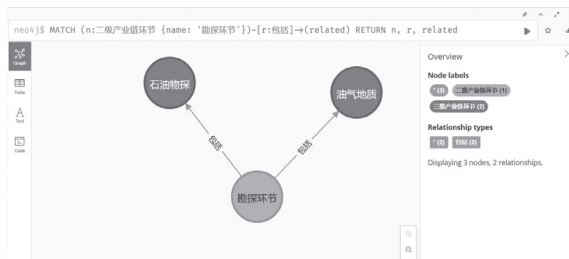


图9 勘探环节包括的关系查询结果

3 结论

本文从产业链视角探讨了标准知识图谱的构建原则与方法，旨在拓展标准数字化发展与智能化应用。在具体应用上，以石油行业为例，将石油标准按照产业链环节进行分类，从而设计本体结构，再依据文中所提出的方法进行标准知识图谱的构建。最后，通过对知识图谱进行可视化展示及相关节点查询发现：目前石油标准主要集中于

上、中游环节,下游环节涉及标准较少。在上、中游重点领域,石油物探与油气地质标准涵盖数据采集、处理、测量及设备使用等多个方面,这些标准为资源评价提供了科学、规范的依据,为后续开发奠定了基础;开采环节则涉及石油钻井、测井及油气田开发等领域的标准,涵盖了设备、作业、数据处理与环境安全等关键因素,旨在提升开采效率与安全性;生产采油环节则集中于采油技术、油田化学剂应用等标准,以实现生产过程的高效、安全

与环保;运输和炼厂环节中的标准,关注流量计量的运行维护、储库管理、原油评价与炼厂气组成测定等方面,确保油品的质量与运输过程的高效与精准。本研究系统梳理了石油产业链上、中、下游各环节的标准,揭示了标准分布呈现“上游密集、中游完善、下游薄弱”的全链条覆盖特征,不仅为促进石油产业链各环节的协调发展提供了重要依据,也为其他行业开展标准知识图谱构建与产业链标准化研究提供了可借鉴的方法参考。

参考文献

- [1] 中共中央、国务院印发《国家标准化发展纲要》[J]. 中华人民共和国国务院公报, 2021 (30): 35-41.
- [2] ISO/TMBG SAG MRS. Questionnaire on ISO TCs' experience of working with SMART standards [Z]. Geneva: ISO, 2019.
- [3] CEN-CENELEC. CEN-CENELEC digital transformation strategic plan [R]. Brussels: CEN-CENELEC Management Centre, 2017.
- [4] LE-PHUOC D, QUOC H N M, QUOC H N, et al. The graph of things: A step towards the live knowledge graph of connected things[J]. Journal of Web Semantics, 2016, 37: 25-35.
- [5] PENGJUN L, QIXIN Z, YINGMIN L, et al. Survey and prospect for applying knowledge graph in enterprise risk management[J]. Computers, Materials & Continua, 2024, 78(3): 3825-3865.
- [6] RUAN J, SU Q, CHEN Z, et al. CPRS: a clinical protocol recommendation system based on LLMs[J]. International Journal of Medical Informatics, 2025, 3(195):105746.
- [7] 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.
- [8] 尤薇佳,孟芸竹,杨跃翔,等. 基于突发事件应对标准的知识图谱构建及应用研究[J]. 情报理论与实践, 2023, 46(11): 164-173.
- [9] 黄珊,牟建荣,王凯月,等. 基于油田领域标准知识图谱的数字化服务平台研究[J]. 标准科学, 2023 (5):71-75.
- [10] 彭泽恒, 刘冰, 谭笑, 等. 油气管道标准知识图谱服务模式研究与系统研发[J]. 标准科学, 2024 (12): 57-66.
- [11] 张天成, 田雪, 孙相会, 等. 知识图谱嵌入技术研究综述[J]. 软件学报, 2023, 34(1): 277-311.
- [12] 周园春,王卫军,乔子越,等. 科技大数据知识图谱构建方法及应用研究综述[J]. 中国科学: 信息科学, 2020, 50(7): 957-987.
- [13] 徐增林,盛泳潘,贺丽荣,等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.