# 基于知识图谱的标准知识管理研究

杨德相　李剑锋

（中国计量大学）

**摘　要：**行业标准化体系构建过程中，随着标准的种类与数量不断更新，人工加载以及查询的方式已经难以满足标准查新跟进，知识服务手段较为单一。知识图谱技术为整合标准知识提供了一种全新的知识互联思路，为完善标准化建设路径、标准文件结构化查询提供了全新的方向。本文分析了标准体系构建现存的问题，以食品安全国家标准为例搭建了以食品产品标准为中心的标准引用知识图谱，基于该图谱展示了其可视化检索、标准重要性等应用，进一步分析知识图谱这一知识管理形式在标准体系建设上的优势，促进标准知识智能服务与发展。标准领域知识图谱强调对标准知识管理，有效集成各类标准文本知识、梳理标准信息，同时结构化知识有益于精确标准查询和关联标准挖掘从而助力推动标准的数字化发展。

**关键词：**标准，知识图谱，知识管理

**DOI编码：**10.3969/j.issn.1674-5698.2023.04.005

# Standards Knowledge Management Based on Knowledge Graph

YANG De-xiang　LI Jian-feng

（China Jiliang University）

**Abstract:** The types and quantities of standards are constantly increasing in the process of industry standardization system construction, and manual loading and querying have been difficult to meet the demands of rapid checking and following up of standards. The means of knowledge information service is relatively single. Knowledge graph technology provides a new knowledge interconnection idea for integrating standards knowledge, and provides a new direction for improving the standardization construction path and the structured query of standards documents. This paper analyzes the existing problems in the construction of the standards system, builds a standards citation knowledge graph centered on food product standards as an example, demonstrates its applications such as visual retrieval and standard importance based on this graph, further analyzes the advantages of knowledge management form of knowledge graph in the construction of the standards system, and promotes the standards knowledge intelligent service and development. The standards domain knowledge graph emphasizes the intelligent management of the guiding normative text, effectively integrating various types of standard text knowledge and sorting out standards information, while the structured form can effectively improve the accuracy and efficiency of retrieval for

**基金项目：**本文为浙江省大学生科研创新活动计划资助项目“基于知识图谱的食品安全标准智能问答机器人研究”（项目编号：2021R409047）的研究成果。

**作者简介：**杨德相，硕士研究生，研究方向为知识图谱、知识管理、金融欺诈防范。

李剑锋，副教授，研究生导师，研究方向为商务数据分析、人工智能（机器学习）、大数据技术、管理信息系统。

the subsequent provision of accurate queries and intelligent mining of associated standards, thus helping to promote the digital development of standards.

**Keywords:** standards, knowledge graph, knowledge management

## 1　引言

标准是在一定范围内获得最佳秩序，对活动或其结果规定共同的和重复使用的规则、导则或特性的文件。起到规范和约束行为的功能，在推动行业和社会稳定发展向前方面有着不可或缺的作用。随着信息技术、人工智能以及大数据技术的持续发展和不断变革，新兴技术赋能更丰富的新应用使数据呈现规模式增长[1]。新兴行业、新兴技术需要新标准规范约束，原标准也需要顺应发展技术等因素不断更新完善，因而标准智能化知识管理更加需要与时俱进跟上行业飞速发展的步伐。知识图谱提供了一种全新的知识互联思路，为整合标准体系与完善标准化建设提供实现标准联结、梳理标准框架和标准动态更新的新方向。

知识图谱本质上是一种语义网络知识库，旨在描述客观世界的概念、实体、事件及其间的关系，提供了一种让用户快速获取相关知识及其逻辑关系的渠道。其核心要义是以图形方式向用户返回经过加工和推理的知识，揭示实体之间关系的语义网络[2]。知识图谱分为未聚焦于特定领域的开放知识图谱和聚焦特定领域的垂直领域知识图谱，前者追求知识广泛度，深度较浅，后者则主要面向专业领域，追求知识深度与准确度。在垂直领域知识图谱的研究中，知识图谱通过表示领域内部的知识联系用以辅助复杂的分析，在生物医学领域的智能问诊[3-4]与金融领域的风险评估[5]、防欺诈[6]以及电商领域[7]等有较多的研究发展。在数据时代，知识图谱通过对数据的整合与规范，向人们提供有价值的结构化信息，已被广泛应用于信息搜索、自动问答、决策分析等领域，是推动数据价值挖掘和支撑智能信息服务的重要基础技术[8]。

随着社会、行业的进步与发展，标准体系在不断壮大，各种数字共享标准平台层出不穷，但是检索方式本质上仍是单条目人工检索，最终呈现的是单一的标准，缺少标准之间的关联和分析。使用知识图谱技术管理标准知识，一方面可以整合标准知识，对于指定的信息给予精确查询和关联标准的链接呈现，提高检索的准确性和效率，另一方面知识图谱将文档层次的粗粒度知识拆分为细粒度的切片化知识，更有益于针对行业标准体系的构建与完善。以标准知识图谱作为知识库为智能查询等提供业务支持，为标准起草人员分析标准信息、检索标准关联、排查标准的重复等漏洞问题，也给各行业相关企业提供标准研读与制定的信息参考。

本文分析了标准知识管理存在的问题，提出构建标准领域知识图谱实现标准知识管理与智能应用。在食品安全国家标准上进行实证，构建了基于食品产品标准知识图谱，实现了知识查询和关联分析。证明了知识图谱这一知识管理形式在促进标准体系智能化建设与知识服务上的优势。

## 2 标准知识管理现状

### 2.1 标准制定存在信息差异

标准本身的分类中，国家标准、行业标准和企业标准涉及的标准制定方不同。我国标准化工作开展较晚，不同标准委员会的信息不完全共享等情况会导致对于标准术语的定义、量度等可能有所不同,进而导致在进行追责时出现负责部门权责模糊、推诿懒政的现象。以食品安全标准为例：肖有明等[11]提出食品安全标准因涉及制定部门较多而导致追责困难，于航宇等[10]指出食品安全标准中对于尚无权威定义的食品品类，后续的标准制定工作无法高效开展。标准制定中的信息差异阻碍了标准化进程与发展，不利于标准的知识整合与管理应用。

### 2.2 标准资源获取效率低下

标准覆盖范围广、分类依据多。每个行业中涉及的标准数量庞大，近几年标准文本进入数字化管理时代，市面上已有较成熟完备的标准文本数据平台，对标准进行存储并实现简单单条目的查询和下载。但各级标准化管理部门在进行标准化工作时通

过该方式获取的标准相对分散，企业实际获取标准过程中往往需要多渠道多次获取，费时费力。部分标准词汇并不局限于某一行业，因此当使用标准中的词汇查询时无法避免其他无关行业对于标准查询的干扰，人工检索的效率较低。

### 2.3 标准知识管理智能化受限

当前新兴产业和新兴技术发展迅速，对应的各级各类标准更新与维护会愈发频繁。现实情况是标准的编制单位和各专业标准化技术委员会分布于多家企业，受到管理的局限性，往往不能做到实时更新，标准的发布相对滞后[12]。在标准的更新过程中，靠人工筛查重复或冲突的标准效率低、准确性也难以保证。并且在该过程中，标准制定部门主要采取的手段仍是人工上载，在数据信息爆炸增长的时代，面对大量的标准维护工作，非智能的信息维护手段给标准化建设和标准体系的构建造成了较大的阻力。

另一方面如今标准化行业发展呈现多行业、多维度的全新局面。由于各专业标准化技术委员会相对独立，有的专业划分界线并不十分明确，导致部分专业交叉、工作重复、标准多头归口、体系交叉重复、技术指标不一致等问题依然存在[13]。现有标准数据共享平台建设过程中，其主要查询方式缺陷在于无法获得标准与标准的关联，无法直接获取关联标准的相关信息。在标准体系中标准与标准之间并非独立，其标准建立过程存在清晰的逻辑思路，标准文档直接堆叠整理并不能体现出标准体系搭建过程中的整体逻辑，当前我国标准知识管理智能化有待深化。

## 3 标准知识图谱架构

### 3.1 标准知识图谱整体框架

通过标准知识图谱实现标准文本知识管理，即按照一定的规则对标准进行知识重组和知识管理，以图数据库形式对标准进行结构化的整合与可视呈现，挖掘标准与标准之间的关联性，以实现标准的深层次信息处理和挖掘。标准知识图谱构建流程如图1所示，从原始数据层逐步深入，本体概念层涉及知识的规则制定，实体数据层包含知识深加工与知识动态更新需求，最后以此作为知识库实现各类与标准知识相关的智能应用。

标准知识图谱的架构主要包括概念层与数据层两部分。概念层存储的是概念化的结构知识，又称为本体，这一部分是知识图谱的概念基础框架，所有存储的数据以该层面定义的知识结构来存储。数据层则是根据概念层规则，在原始数据中提炼出的知识。知识应区别于原始文本，是对标准文本进行拆分细化后形成的“碎化”信息。高质量的数据知识对于标准知识图谱以及后续的智能应用效果至关重要，因此原始数据的知识抽取与加工转换为结构化的知识元是搭建标准知识图谱的关键步骤。

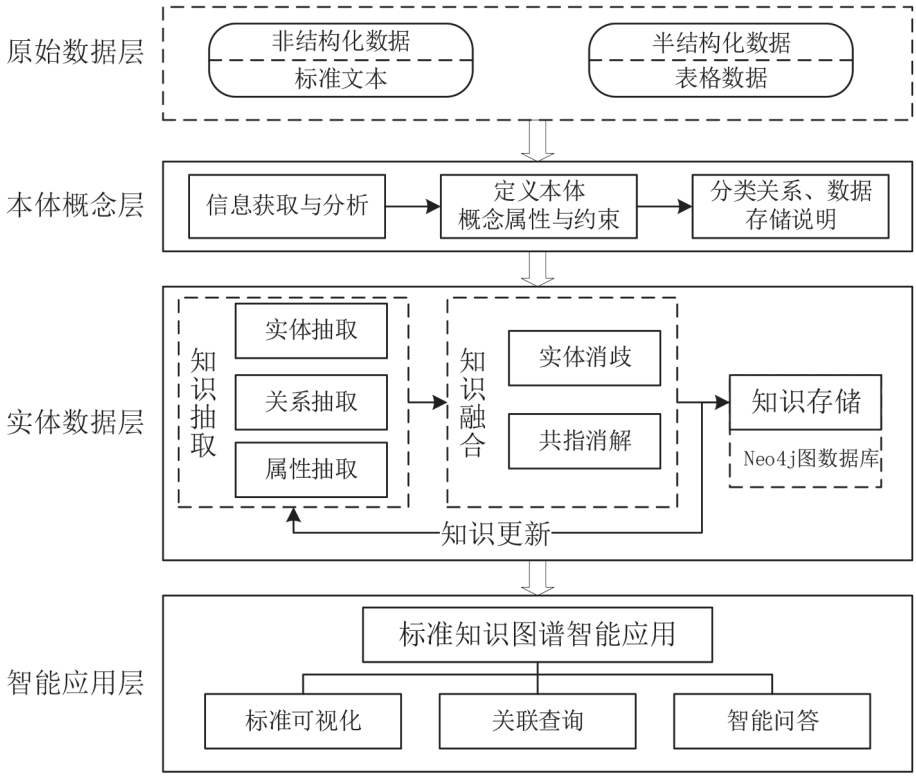


图1 标准知识图谱构建流程

### 3.2 本体概念层

概念层设计就是本体设计，是对最终呈现知识结构的整体把控。本体的构建应以具体的项目领域和任务作为起点，以便于进行本体功能的描述和知识的获取。本体构建技术分为人工[14]、自动[15]和半自动[16-17]3类，在自动构建本体方面，目前还极少有方法能够得到覆盖率和准确率都表现良好的本体，并且没有专门针对中文文档知识的成熟方法。大多构建本体过程都需要人工参与，考虑到标准的结构一致性等特征，采取人工构建本体中的七步法[18]作为标准领域知识图谱本体构建的主要方法。七步法本体构建方法的流程包括：确定标准本体构建领域及范围、获取并分析领域信息、定义本体概念和概念层次、定义概念的属性和属性约束、本体更新评估、本体实例化、文档化说明。该过程中充分结合标准起草人的起草逻辑等专家知识，参照标准编写规则，对标准的内容结构以及特点进行分析，借助工具方法定义本体概念以及属性约束，并对处理后的标准本体进行文档化说明。

标准按照要素的类型和位置共分为4类：资料性概述要素、资料性补充要素、规范性一般要素和规范性技术要素。资料性概述要素包括标准封面、目次、引言以及前言中的内容；资料性补充要素包括标准资料性附录、参考文献以及索引中的内容；规范性一般要素包括标准的名称、范围和规范性引用文件中的内容；规范性技术要素包括术语和定义、符号、代号和缩略语以及规范性附录等内容。结合标准编排要求，标准的一般内容组成如图2所示。

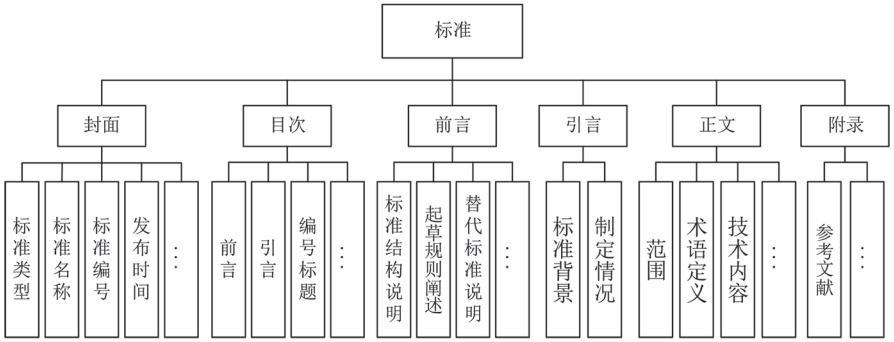


图2　标准的内容组成

参照标准的一般结构，标准实体的基本属性来源于资料性概述要素、资料性补充要素，包含标准的类型、名称、发行时间、起草单位等信息，可以以此直接定义其基本属性概念。规范性一般要素和规范性技术要素中涉及与行业紧密相关的术语、适用范围以及细化的行业技术要素，并不适合直接使用其属性概念，故而参照同行业的标准文件中存在必要的共同元素，以共同元素作为参考进行本体设计。比如：在食品安全标准中，技术内容包含：理化指标要求、污染物限量等共同要素，则“污染物限量”可以作为一个关系概念用以指向该标准与引用标准之间的关系属性。标准文件的专业特性与已有编著逻辑性，决定了其本体建模主要结合专业性知识，以语义判断为根本原理施行[19]。

### 3.3 实体数据层

标准知识管理应注重标准的知识完整性、准确性，唐爽等[13]提出标准知识库应具有信息时效性，赵丹[20]构建企业标准体系时强调系统需保证标准体系的动态更新确保标准体系对于企业的准确指导，均强调了对于标准知识的完整性与准确性要求。因此标准实体数据层的知识质量也决定了标准知识图谱的整体质量以及采信度。标准实体数据层主要包括知识抽取、知识融合、知识存储以及更新。其中知识抽取是将标准文本中的必要关键信息进行格式转换后结合自然语言处理技术得到实体、关系属性等信息，初步获得结构化知识实现。从而实体间语义链接。知识融合需要对冗余的知识进一步处理简化，对三元组在统一框架标准下进行整合、消歧，简化知识体系，形成标准知识网络。知识存储环节采取开源Neo4j图数据库作为工具，导入精简化后的结构化知识，通过标准节点之间的引用关联将标准知识组合成可以系统查询与更新的知识网络。

## 4 食品标准知识图谱实证

### 4.1 食品标准知识图谱构建

食品安全标准是相关权威机构依照程序制定的规范性文件，对推动食品安全发展起到至关重要的作用。我国已有食品、食品添加剂、食品相关产品国

家标准1,300余项，行业标准2,900余项，地方标准1,200余项，形成了相对完善的食品标准体系。但是食品安全标准种类多、层次丰富，一定程度上给标准系统知识管理造成了一定的困难。其配套法规政策不足，制定范围、定位不明确 ，内容庞杂，并且相互引用形式多样，在信息公开性上仍有欠缺，这些都对于消费者合理维权，企业有效生产经营产生不良影响。

食品安全标准是众多的食品标准中唯一强制执行的标准，因此本文以现行食品安全国家标准作为研究对象。根据食品安全标准与监测评估司发布的食品安全国家标准目录显示，食品安全国家标准分为通用标准、食品产品标准、特殊膳食食品标准、食品添加剂质量规格及相关标准等共计12类。由于食品安全标准制定底层逻辑是围绕食品的生产制造过程进行的，故而选择食品产品和特殊膳食食品标准共计80份标准文件用作实证，以下将上述国家安全标准统称为食品产品标准。

对食品产品文件分析，以标准作为实体，对其主要共有元素进行分析设计本体。其封面中包含的标准名称、编号以及发行时间作为标准的基础属性。以食品产品标准GB 5420为例，标准名称为《干酪》，标准编号为GB 5420，发行时间为“2021”。食品产品标准的内容属性包括适用范围、相关术语以及术语定义，位于文件正文部分“1 范围”以及“2 术语和定义”。参照“3 技术要求”部分定义食品产品标准与其他标准的关系属性，食品产品技术要求包含原料要求、感官要求、理化指标、污染物限量和真菌毒素限量、微生物限量、食品添加剂和食品营养强化剂，其中微生物限量常细分为致病菌限量和微生物限量。文件“4 其他”包含食品外包装等其他相关要求，综合上述内容结合食品安全国家标准的12个大类别，定义食品产品标准同其他标准的关系属性见表1，食品产品标准知识图谱本体模型如图3所示。根据本体设计逻辑，基于规则对标准原始数据的进行知识抽取与加工，并将简化后的结构化三元组进行存储。

表1 食品产品标准关系属性

| 关系 | 解释 |
|---|---|
| 感官要求检验方法 | 该产品的感官要求检验方法应符合该标准 |
| 理化指标检验方法 | 该产品的理化指标检验方法应符合该标准 |
| 污染物 | 该产品污染物限量应采用该标准要求 |
| 真菌毒素 | 该产品真菌毒素限量应采用该标准要求 |
| 致病菌 | 该产品致病菌限量应采用该标准要求 |
| 微生物检验方法 | 该产品微生物检验方法应符合该标准 |
| 微生物分析处理方法 | 该产品微生物分析处理方法应符合该标准 |
| 农药残留 | 该产品农药残留限量应采用该标准要求 |
| 食品添加剂 | 该产品食品添加剂的使用应采用该标准要求 |
| 食品营养强化剂 | 该产品食品营养强化剂的使用应符合该标准要求 |
| 包装 | 该产品包装应符合该标准要求 |

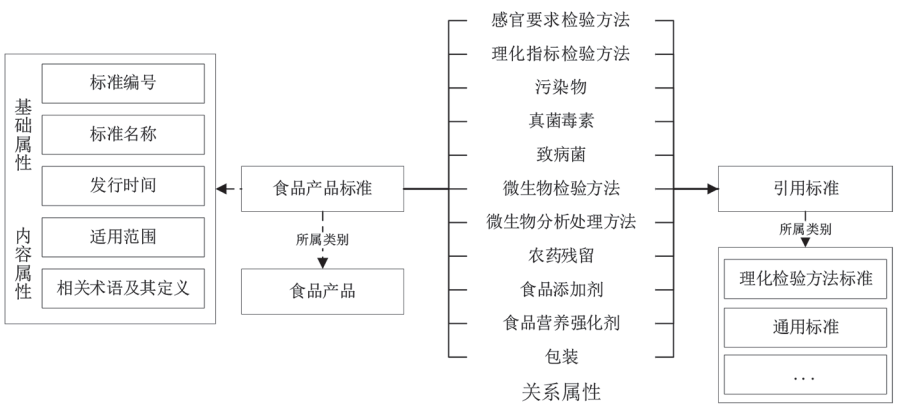


图3 关联知识图谱本体设计

## 4.2 食品产品标准知识图谱可视化

食品产品知识图谱可视化通过Neo4j图数据库实现，Neo4j图数据库可以清晰地展示出节点之间的依赖关系以及显性关系属性。对食品产品标准以其本体模型进行知识抽取与知识融合后，形成食品产品标准与其他食品安全标准的关联数据资源导入图数据库中，实现食品产品标准资源的可视化存储与访问。通过py2neo工具包可通过Python应用程序内部和命令行直接使用Neo4j，实现批量结构化知识导入。数据导入后在Neo4j图数据库中可使用Cypher查询语言检索食品产品标准知识图谱中节点及其关联关系，相关Cypher语句示例见表2。

图谱示例如图4所示，实体节点共计212个，关系

数量789。通过进一步点击访问可以查询各节点具体情况基本属性以及关联标准情况。

表2 知识图谱关系语句Cypher语句示例

| Cypher语句 | 功能 |
|---|---|
| CREATE (n1: Nodename{title:"name1"}) | 创建节点 |
| MATCH (n: Nodename)–[r]–() where n.name=' name1' delete n,r | 删除指定节点 |
| CREATE (n1)–[r: relation]–>(n2) | 创建关系 |
| MATCH (n: Nodename) RETURN n | 匹配该标签节点 |
| MATCH (n) RETURN COUNT(*) | 节点总数 |
| MATCH P=()–->() RETURN COUNT(*) AS COUNT | 关系总数 |
| MATCH p=()–[r: relation]–>() RETURN p | 某限制关系所有关联标准 |
| MATCH (a: Nodename1{name:'n1'})–[r:relation ]–>(b) RETURN a, b | 某节点及限制关系关联标准 |
| MATCH (a: Nodename1{name:'n1'})–->(b) RETURN a, b | 标准所有直接关联标准 |

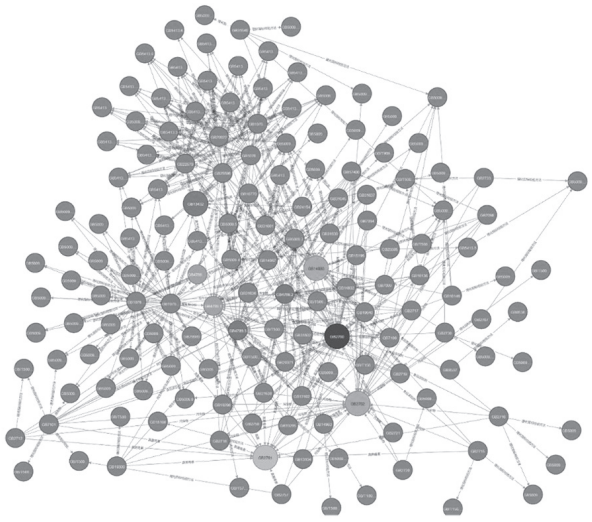

图4 食品产品标准关联知识图谱概览

### 4.3 食品产品标准关联查询

食品产品标准知识图谱的主要优势在于以知识网络直观呈现了食品产品标准与其他标准之间的引用关系，为标准的查找和分析提供便捷的知识管理可视化工具。标准关联查询的首要作用，对于食品生产过程中需要参照标准针对性地进行汇总，对产品涉及的各技术要素以及检验方法实现“一图直达”。标准“GB 25570 辅食营养补充品”的关联图谱，以该产品标准为中心的网状结构直观地整理了该标准存在有28项关联，主要的19项标准关联产生在理化指标检验方法上，对于食品添加剂、营养剂以及外包装均有相关的标准要求（如图5所示）。


图5 “GB 25570”节点关联图谱

其次，关联查询给标准的修订提供了重要性数据参考。食品生产与人民群众生命安全息息相关，知识图谱可以通过节点出度、入度，从数据层面标记标准重要性以及关联程度。以“微生物检验方法标准”为例，在Neo4j中直接查询“MATCH (a)–[r:`微生物检验方法`]–>(b)RETURN b, COUNT(r) ORDER BY COUNT(r) DESC”，图谱如图6所示，数据结果见表3。结果显示与食品产品直接关联的微生物检验方法标准重要性前三分别为GB 4789.3、GB 4789.2、GB 4789.4，关联数目分别为50项、40项以及21项，因此在修订相关标准时对于关联程度较广的标准应更加谨慎、多方考虑。

表3 微生物检验方法标准关联度示例

| 标准 | 出度/入度 |
|---|---|
| GB 4789.3 | 50 |
| GB 4789.2 | 40 |
| GB 4789.4 | 21 |
| GB 4789.15 | 18 |
| GB 4789.10 | 17 |
| GB 4789.26 | 3 |

## 5 结 语

我国标准的领域知识深度广，有效的知识管理对于提高标准文件信息管理水平、促进标准化工作成果具有重要意义。标准知识图谱的核心在于标准知识单元的重组与细化，对标准文档逻辑化的拆分

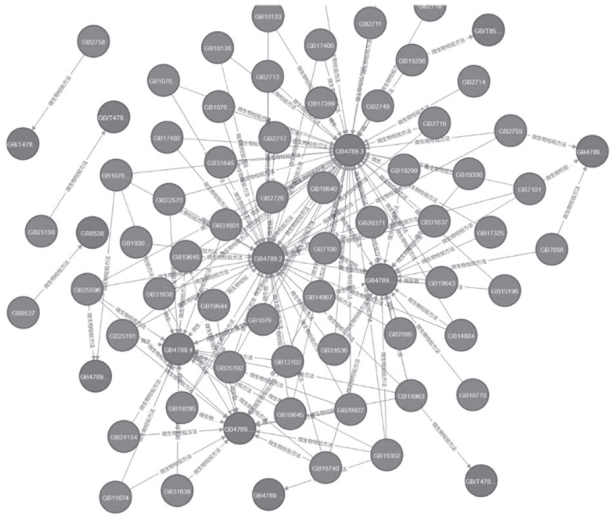

图6 微生物检验方法标准关联图谱

的知识管理优势在于，对于实体不仅囊括其涉及属性的长文本，还能够基于标准的制定逻辑对标准之间引用关系进行存储，对于标准文档做到了知识概括性、引用关联性同时把握。知识图谱的语义网络特性在描述标准的语义关系上充分发挥效能，做到更深层、更高细粒度的知识管理，并为以此作为底层知识库开展的智能应用打下基础。

知识图谱的组织模式提供了标准管理的框架和底层逻辑，后续研究方向聚焦于构建图谱的效率。探索具有通用性的标准本体构建方法、提高标准知识抽取加工过程精确度以及如何深度利用标准知识图谱实现智能应用，例如：精准问答等扩展。

## 参考文献


[1] 杨波,杨美芳. 知识图谱研究综述及其在风险管理领域应用[J]. 小型微型计算机系统, 2021,42(08):1610–1618.

[2] 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(03):582–600.

[3] Cheng B, Zhang J, Liu H, et al. Research on medical knowledge graph for stroke[J]. Journal of Healthcare Engineering, 2021, 2021.

[4] Zhang D, Jia Q, Yang S, et al. Traditional Chinese Medicine Automated Diagnosis Based on Knowledge Graph Reasoning[J]. CMC–COMPUTERS MATERIALS & CONTINUA, 2022, 71(1): 159–170.

[5] Yang B, Liao Y. Research on enterprise risk knowledge graph based on multi–source data fusion[J]. Neural Computing and Applications, 2022, 34(4): 2569–2582.

[6] 袁俊,刘国柱,梁宏涛,等. 知识图谱在商业银行风控领域的研究与应用综述[J/OL]. 计算机工程与应用: 1–16[2022–08–11].

[7] 王思宇,邱江涛,洪川洋,等. 基于知识图谱的在线商品问答研究[J]. 中文信息学报, 2020,34(11):104–112.

[8] 孙佳琛,王金龙,丁国如,等. 频谱知识图谱：面向未来频谱管理的智能引擎[J].通信学报,2021,42(05):1–12.

[9] 胡琳,杨建军,韦莎,等. 工业互联网标准体系构建与实施路径[J]. 中国工程科学, 2021,23(02):88–94.

[10] 黄持伟,吴学科,阳如坤,等. 锂电池智能制造装备标准体系研究[J]. 中国标准化, 2021(07):57–62+93.

[11] 肖有明,赖森森. 我国的食品安全标准体系建设[J]. 现代食品, 2020(17):145–147.

[12] 于航宇,樊永祥,王家祺. 我国现行食品安全地方标准分析[J]. 中国食品卫生杂志, 2019,31(05):485–489.

[13] 唐爽,韩义萍,张玉,等. 标准知识库构建研究[J]. 中国标准化, 2020(S1):46–50.

[14] 赵雪芹,李天娥. 面向数字人文的档案领域本体构建研究——以万里茶道档案资料为例[J/OL].情报理论与实践:1–9[2022–08–09].

[15] 熊励,王成文,王锟. 基于事件本体的疫情知识库构建策略[J]. 图书情报工作, 2021,65(14):138–148.DOI:10.13266/j.issn.0252–3116.2021.14.016.

[16] 刘博,张佳慧,李建强,等. 大气污染领域本体的半自动构建及语义推理[J]. 北京工业大学学报, 2021,47(03):246–259.

[17] 唐琳,郭崇慧,陈静锋,等. 基于中文学术文献的领域本体概念层次关系抽取研究[J].情报学报,2020,39(04):387–398.

[18] Wang P, Mao Y, Song W, et al. A Comprehensive and Scientifically Accurate Pharmaceutical Knowledge Ontology based on Multi–source Data[C]//BIOINFORMATICS. 2022: 168–175.

[19] 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(05):58–65.

[20] 赵丹. 大庆油田标准体系动态管理系统的研究[J]. 中国标准化, 2017(17):114–118.