标准文献知识图谱构建与应用研究*

杨跃翔 涂新雨 刘文玲
（中国矿业大学（北京）管理学院，北京 100083）

摘要：促进标准文献知识的开发和利用，需要研究标准文献的知识组织模式和方法，推动标准数字化转型。本文通过分析标准文献的结构特征，构建标准文献的本体框架，涵盖标准文献中共性要素的概念和关系；并通过XML标准标签集拓展，构建适用于我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，并以实例挖掘标准文献知识图谱的应用价值。本研究聚焦标准文献，提出标准文献知识图谱构建方法，实现标准知识的交叉关联和共享重用，助力标准文献知识服务和智能应用。

关键词：标准文献；知识组织；标准数字化；知识图谱；知识服务
中图分类号：TP182；G307　DOI：10.3772/j.issn.1673-2286.2022.06.004
引文格式：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.

标准是为了在一定范围内获得最佳秩序，经协商一致制定并由公认机构批准，共同使用的和重复使用的一种规范性文件[1]。随着时代的发展，标准在经济社会发展中所起的作用越来越突出，标准文献在数量上呈现增长趋势。

目前，我国标准文献是以PDF格式或纸质版进行发布和存储，不能实现机器可读，标准尚处于以文献为基本单元的方式进行知识表示的阶段，无法实现标准文献间知识交叉关联；现有标准文献的存储方式不利于标准体系构建和知识梳理，无法实现标准文献知识的快速检索和精确匹配，难以发挥标准文献知识辅助决策的作用。同时，标准文献之间存在对同一术语进行不同定义等知识冲突现象，易造成用户理解上的歧义，不利于标准的规范化，影响使用，亟需进行标准的数字化转型。标准的数字化转型是利用数字技术对标准化工作的全流程及标准本身的全生命周期赋能，实现灵活高效可交互的标准研制过程，创建标准结构和内容机器可读的新型标准模式，拓展标准使用的数字化、智能化服务[2]。

知识图谱技术的出现能够很好地赋能标准的数字化转型。知识图谱是一种解释实体之间关系的语义网络，可以实现对事物及其相互关系进行形式化描述，同时也提供了一种全新的信息检索模式。借助知识图谱技术对标准文献进行知识组织可以较好地展示标准的知识语义关联，实现标准文献内容知识的可关联、可对比、可追溯和可分析。知识图谱通过实体、属性和关系来定义标准知识，支持标准知识的共享和重用，同时采用语义相似度计算和实体关系匹配等方法，可以对标准文献中的使用范围、术语、关键技术指标等知识进行比较分析，为标准立项、审核、使用和修订等工作提供标准知识的辅助决策，服务于标准全生命周期。

本研究通过对标准文献的结构特征进行解析，得到标准文献中共性要素的概念和关系，构建标准文献知识图谱本体层，并参照国外标准标签集对我国标准标签集进行拓展和细化，完成标准文献XML格式转换，实现机器可读，同时进行知识抽取，从而构建标准文献知识图谱，并以自然灾害应急标准进行实证研究，构建自然灾害应急标准文献知识图谱，探讨标准文献知识图谱的实际应用场景。

* 本研究得到国家重点研发计划基金项目“突发事件应对标准数字化通用方法与关键标准研究”（编号：2021YFF0600401）资助。

1 相关研究

国际上已经开展标准数字化转型的相关研究，但都处于初步探索阶段。目前，三大国际标准组织（ISO、IEC、ITU）、欧洲标准化组织（CEN/CENELEC），以及美国、德国、俄罗斯等均已启动机器可读标准的研制和实施工作[3]。其中，国际标准化组织（ISO）开发了标准标签集ISOSTS（ISO Standards Tag Set）[4]，用于描述标准全文内容和元数据，提供可用于发布和交换标准内容的通用格式。美国国家信息标准组织（NISO）在ISOSTS的基础上进行丰富和优化，形成了标准标签集NISOSTS（NISO Standards Tag Set）[5]。我国于2019年发布国家标准《基于XML的国家标准结构化置标框架》（GB/T 37967—2019），规定了标准文本结构的XML标签集，但标签集相对简略，仅实现了标准结构层面的标注。标准标签集用于对标准文献结构和技术内容要素进行标记和分析，可以通过标准标签集拓展和细化以丰富机器可读标准内容[6]。借助XML建模语言技术转化标准文献为机器可读，可以实现标准文献内容信息的直接提取和查询[7]。但经过转换的标准文献XML格式，只能实现机器可读，对语义关系的表示有限，不能实现不同标准文献内容知识的语义交叉关联和共享重用[8]，难以达到标准知识智能化服务的效果。对此，ISO定义了SMART（Standard Machine Applicable，Readable，Transferable）标准的概念[9]，认为构建机器可用、可读、可解析标准是标准数字化的发展方向。Loibl等[10]提出要实现标准的机器可操作需要将标准文献信息建模为机器可操作的形式，并从语义关联可见性、易扩展性和数据调用速度等角度将传统关系型数据库与图数据库进行对比，认为图数据库更适合机器可操作标准的存储和应用。刘曦泽等[11]提出利用知识图谱等技术进行标准内容知识的提取、分类与表达，将标准文本转化为可自由使用的动态知识网络，进而实现“人机交互”，这是标准数字化的发展趋势。

关于采用知识图谱技术对标准文献进行知识组织，相关学者从不同角度论证了其可行性。Luttmer等[12]以公式为例将标准内容从XML格式转换为基于图形表示的知识图谱，验证知识图谱适合于表示机器可操作的标准内容。Sana等[13]分析了基于XML数据进行知识图谱建模、存储和处理的可能性。XML标签可以自定义携带语义信息，可以通过XML解析实现批量知识抽取，辅助知识图谱的构建[8]。刘慧琳等[14]提出在各种文献信息资源中，标准文献的自身特点可以较好适配知识图谱结构。目前学术界对标准文献知识图谱构建方法的研究比较缺乏，部分学者只是选取标准文献中的部分结构要素进行知识抽取。Ren等[15]提出了标准文献知识图谱构建和应用的结构框架。张慧等[16]采用基于规则的知识抽取方法，抽取标准文献的前言部分和规范性引用文件部分，构建了描述标准文献与组织机构关联关系的知识图谱。张鹏飞等[17]通过大量人工标注，采用BERT-TCNN-BiLSTM模型对绿色标准中的部分共性结构要素进行实体抽取，搭建绿色标准知识图谱。基于规则的方法可以保证知识抽取的准确率，但需严格限制文本语言格式，只能局限于部分知识的识别和抽取，而采用深度学习等方式进行知识抽取需要大量人工标注，且其实验准确率欠佳，难以满足标准文献规范度的要求。郝文建等[18]提出标准文献要素抽取的思路，认为可以采用基于规则的方法与自然语言处理技术相结合的方式进行要素抽取。秦丽等[19]采用基于规则和人工参与相结合的方式，对标准文献中引用关系和标准中的部分内容进行知识抽取，构建国家食品安全标准知识图谱。Jiang等[20]通过分析建筑安全标准体系，设计了由五个层次概念和八种类型关系组成的概念层，构建了建筑安全标准知识图谱。

综上，目前标准文献知识图谱构建的研究大多停留在仅选取标准文献的部分结构性内容（如标准文献引用关系、标准文献与组织单位之间的关系等），缺少从标准文献整体结构内容出发，对标准文献进行知识拆解的研究，对于标准文献知识抽取方法的研究也处于探索阶段，目前尚未形成适用于标准文献知识抽取的较为成熟的方法。因此，本文从标准文献整体结构内容出发，采用拓展XML标准标签集，基于XML标注进行知识抽取的方法，构建标准文献知识图谱，实现标准文献整体结构内容的知识切片和重组，可以更全面地挖掘、分析和展示标准文献之间知识的关联关系，解决现有研究对标准知识加工不充分、知识关联不全面，难以有效支持标准文献知识的实际应用的问题，更好地服务于标准的应用。

2 标准文献的结构特征分析

标准是为各项活动及其结果提供规则、指南或特性，共同使用和重复使用的文件，标准的起草和编写需要按照统一的规则和规范，以便于起草者编订适用性更

好的标准，更好地服务于标准使用者。为此，我国先后制定了多项标准编写规范类的标准。通过分析标准文献的结构和内容，可以发现标准具有文本结构规范、层次清晰和词义表述明确、言简意赅的特点。标准文献的知识单元和知识关联模式是识别、研究和应用标准知识的基本出发点。构建标准文献知识图谱，需要对标准文献的组成要素、层次和知识关联逻辑进行分析，进而确定标准文献文本特征的知识切片和重组方法。因此，标准文献的结构解析是采用知识图谱对其进行表达的基础。

《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）明确规定了标准文献的组成要素，包括封面、目次、前言、引言、范围、规范性引用文件、术语和定义、符号与缩略语、分类与编码/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、参考文献和索引。此外，按照要素存在的状态将要素分为必备要素和可选要素，其中，封面、前言、范围和核心技术要素是必备要素，规范性引用文件、术语和定义既属于必备要素也属于可选要素，其他要素属于可选要素。封面主要包括标准中文名称、标准英文名称、标准号、ICS分类号、CCS分类号、发布日期、实施日期和发布单位；前言包括提出单位、归口单位、起草单位、起草人信息；范围是标准文献的摘要信息，主要介绍标准所规定的内容和适用界限；规范性引用文件主要包括标准所引用的文件和文件代码；术语和定义是对标准中所涉及的专业术语进行定义；核心技术要素是标准的主体内容部分，主要以章、条标题和内容形式呈现。将标准文献中必备的组成要素定义为标准文献的共性结构要素，可得标准文献的共性结构要素如图1所示。

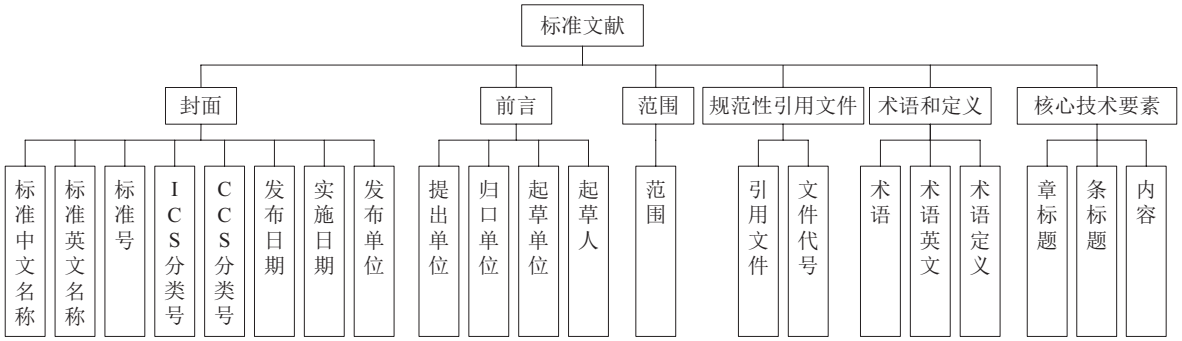

图1 标准文献共性结构要素

由于标准文献具有共性结构要素，可以对其进行结构化处理。XML格式文档可以实现对固定文本结构文档的结构化处理，通过对标准文献中的共性要素进行自定义标签标记，将其转化为XML文档，可以更便捷地被计算机识别、读取，进而实现标准文献的批量解析，从中获取标准文献共性要素知识信息。

标准按照功能类型可划分为术语标准、符号标准、分类标准、试验标准、规范标准、规程标准和指南标准。不同功能类型的标准文献的核心技术要素存在描述逻辑和内容的不同。例如，术语标准主要包含术语条目，试验标准主要包括试验步骤和试验数据处理。不同功能类型的标准文献所对应的核心技术要素部分文本内容如表1所示。

对于不同功能类型的标准文献核心技术要素，可以参照其结构内容进行结构化处理。其中，术语标准的核心技术要素主要以术语条目形式逐条展示，可直接进行XML标签标记实现结构化处理；符号标准和分类标准的核心技术要素多以表格形式和描述性文本形式呈现，表格属于结构化程度较高的内容展现形式，便于处理和表示；试验标准、规范标准、规程标准和指南标准中的内容多以列项、段落等形式展示，其中试验步骤和指标类内容等多以列项展示，便于进行结构化处理，但对于要求、指示等多以段落性文本描述存在的文本，需在保留语言描述的完整性和准确性的基础上做进一步知识加工。

表1 标准文献核心技术要素

不同功能类型的标准	核心技术要素
术语标准	术语条目
符号标准	符号/标志及其含义
分类标准	分类和/或编码
试验标准	试验步骤、试验数据处理
规范标准	要求、证实方法
规程标准	程序确立、程序指示、追溯/证实方法
指南标准	需考虑的因素

3 本体层构建

构建标准文献知识图谱就是对标准文献内容进行知识粒度细加工，深入到标准文献内部的语义知识单元，挖掘标准文献知识单元之间的关联关系，进而对标准文献内容进行语义组织，实现标准文献内容的细粒度表达和知识语义关联。通过解析标准文献的结构特征，选取自顶向下的方式进行标准文献知识图谱构建，首先应进行本体知识建模，主要包括概念集和属性的确定以及概念间关系的定义，用于约束数据层实体、关系和属性的抽取及语义关联。

3.1 概念集和属性的确定

在构建标准文献本体层的概念知识体系时，要充分考虑标准文献的结构层级、知识单元和用户需求，因此，标准文献本体层的构建应遵循标准文献内部的结构层次逻辑，将相同或者相近语义的知识片段归纳概括为具有普适性和通用性的概念，以标准文献中共性必要要素构建模式层的概念体系。另外，概念的定义应充分考虑对应实例知识单元细分程度，既要尽可能细化以保证标准文献的所有语义可展示，又要恰当切分以避免语义关系缺少和错乱。

从标准文献的内容和结构层次来看，标准文献中存在共性的必要元素，这些元素可以定义为概念，如标准中文名称、标准号、提出单位、归口单位、起草单位、起草人、章标题和条标题是标准文献的共性必备要素，应分别设立为本体层的概念。范围是标准文献核心内容的提取，用于界定标准文献的规定内容和适用界限，根据标准文献的范围部分结构内容，可将范围中的“规定内容”和“适用界限”分别设为两个概念。标准文献中对于术语进行定义是为了避免引起误解或对技术内容的理解产生歧义，术语对标准规范化意义重大，可将“术语”设为一个概念。由于术语存在一词多义等现象，如果将术语定义设为数据属性则难以挖掘术语定义的不同来源情况，因此将“术语定义”单设为一个概念。由此，通过对标准文献进行知识梳理和整合，最终建立包含12个本体概念的标准文献概念集，具体如表2所示。

就概念的属性而言，标准英文名称、ICS分类号、CCS分类号、发布日期、实施日期和状态可设置为概念“标准中文名称”的属性，用于描述标准文献的基本信息；术语英文可设为概念“术语”的属性；章标题和条标题下对应的段落性文本则可分别设为概念“章标题”和“条标题”的属性。

表2 标准文献概念集

概 念	注 释
标准中文名称	标准文献封面中的标准中文名称
标准号	标准文献的标准号，包括国标号、行标号等
提出单位	标准文献前言部分的提出单位
归口单位	标准文献前言部分的归口单位
起草单位	标准文献前言部分的起草单位
起草人	标准文献前言部分的起草人
规定内容	标准文献范围部分所规定的内容
适用界限	标准文献范围部分的适用界限
术语	标准文献中所定义的术语
术语定义	标准文献中对所涉及术语的定义
章标题	标准文献主体内容部分的一级标题
条标题	标准文献主体内容部分一级标题下的子标题

3.2 概念间关系的定义

知识图谱本体层的概念之间的关系包含层次关系和非层次关系，其中，层次关系为概念间上下位关系，非层次关系主要基于概念所属范围和类型进行定义。标准文献本体概念间的关系是根据标准文献结构和内容知识关联关系进行定义的，概念间关系以非层次关系为主。通过分析不同概念在标准文献内部和标准文献间的语义关联关系，可以对标准文献本体概念间的关系进行定义。

基于已建立的标准文献概念集，参照标准文献的结构特征，依据各本体概念在标准文献文本中的位置，可初步建立标准文献内部本体概念间的关系。其中，“标准中文名称”与“标准号”概念之间的关系为“标准号”；标准文献的前言部分包含“提出单位”“归口单位”“起草单位”和“起草人”四个概念，“标准中文名称”与其关系可分别定义为“提出于”“归口于”“起草于”和“起草人”；标准文献中术语与术语定义部分包含“术语”和“术语定义”两个概念，“标准中文名称”与“术语”之间的关系可定义为“涉及术语”，“术语”和“术语定义”之间关系为“定义”；标准文献的核心技术要素以章、条标题及内容进行展开，“标准中文名称”与“章标题”两个概念之间的关系可定义为“包含”，“章标题”与“条标题”之间关系定义为“包含”。

除了上述标准文献内部各本体概念之间的关系之外，标准文献间还存在大量的知识交叉关联，需要进一步深入挖掘标准文献间的知识关联关系，进而补充和丰富标准文献本体概念之间的关系。①不同标准文献间存在引用现象。由于标准文献的规范性引用文件也多为标准文献，即标准文献之间的引用关系是在概念“标准中文名称”下的实例之间产生，因此可以在“标准中文名称”与规范性引用文件的“标准中文名称”之间建立引用关系。②标准文献存在不定期的更新修订。由于标准文献之间存在对于已作废标准文献的引用，为准确地追踪溯源，所建立的标准文献知识图谱中应保留部分已作废但被引用的标准文献。对于不同状态的标准文献，应建立关系为“更新”。③对于标准文献中的术语与术语定义部分，既存在由于术语在不同标准文献中应用场景等的不同而对术语定义进行改写的现象，也存在不同标准文献之间同一术语和定义引用的现象。由于术语的改写主要是术语定义发生改变，即术语定义的改写关系是在概念“术语定义”下的实例之间产生，因此，部分“术语定义”实例间存在“改写”关系。为了清晰表明相同术语不同定义的来源情况，可在“术语定义”与“标准中文名称”之间建立关系为“来源于”。对于同一术语不同标准文献引用的现象，由于在构建标准文献知识图谱时，会自动将重复节点进行合并，难以直接展示所引用术语的初始位置，可通过“标准中文名称”之间的引用关系来表明术语的来源。④标准文献核心技术要素部分也存在知识交叉关联现象，由于不同标准文献之间存在章标题或条标题相互重复现象，构建标准文献知识图谱可以实现重复内容的共享重用，但需要标明标题来源，因此在“条标题”与“标准中文名称”之间建立关系为“来源于”。通过分析和挖掘标准文献内部和标准文献之间的知识关联，得到标准文献本体概念间的关系（见表3）。

表3 标准文献本体概念间关系

概 念	关 系	概 念
标准中文名称	标准号	标准号
	提出于	提出单位
	归口于	归口单位
	起草于	起草单位
	起草人	起草人
	规定了	规定内容
	适用于	适用界限
	涉及术语	术语
	包含	章标题
	引用了/更新	标准中文名称
术语	定义	术语定义
术语定义	改写	术语定义
	来源于	标准中文名称
章标题	包含	条标题
条标题	来源于	标准中文名称

通过整合标准文献本体概念、属性和关系，得到标准文献知识图谱本体模型（见图2）。其中，“组织机构”指标准文献前言部分所包含的“提出单位”“归口单位”“起草单位”和“起草人”。

该本体模型涵盖了标准文献中所有必备要素，但其概念和关系的定义主要针对于标准文献中共性要素。

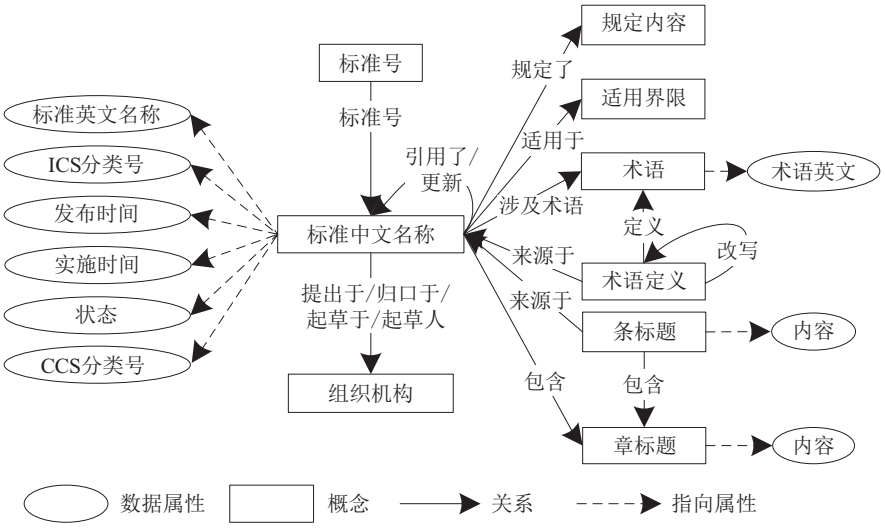

图2 标准文献知识图谱本体模型

例如，对于标准文献核心技术要素部分内容，将各级标题归纳为概念，标题下的内容归为各级标题对应的数据属性。此建模方式适用于术语标准；对于符号标准和分类标准，可以通过进一步对表格进行知识抽取，细化对应核心技术要素部分的知识，实现细粒度标准知识组织；对于试验标准、规范标准、规程标准和指南标准，此建模方式可以保证语义完整性和准确性，但进一步知识细粒度加工需结合领域知识搭建知识层级关系进行知识建模。

4 基于XML标准标签集的知识抽取

根据本体层所定义的概念、属性和关系，通过标准文献XML格式转化，使得标准文献的内容片断包含语义标签，将标准文献的知识组织方式从文献粒度的树形分类结构向知识粒度的网络结构转变，同时便于进行文档解析，获取相关实体、关系和属性，从而构建标准文献知识图谱。

4.1 XML标准标签集拓展

对于标准文献结构化处理，我国制定了国家标准《基于XML的国家标准结构化置标框架》，其中定义了适用于我国标准格式内容的标准标签集，涵盖标准的封面、前言、引言、术语和标题等内容标签，共包含元素56个、属性2个。该标准所定义的标准标签集相对粗略，仅实现了对标准文献整体结构框架的标签标记，不能覆盖标准文献本体层所定义的概念、属性和关系，需要进行标准标签集拓展。

目前，国际上具有代表性的标准标签集包括ISOSTS和NISOSTS。其中，NISO发布的对应于美国标准的标准标签集内容较为详细，除了对标准结构进行标签标记外，还包括样式和表格等具体内容的标记。因此，在对我国标准文献进行结构化处理时，可以在标准《基于XML的国家标准结构化置标框架》所定义的标准标签集基础上，参照NISO标准标签集，拓展和细化我国标准标签集，从而增加标准标签集的语义信息。通过分析NISO标准标签集，对我国标准标签集进行拓展，在原有标准标签集的基础上，针对标准前言、范围、规范性引用文件、核心技术要素中所包含的标准标签集进行拓展和细化，共拓展了19个元素，目前标准标签集共包含75个元素，拓展后的核心标准标签集涵盖了对于标准文献封面信息、前言部分信息、范围、规范性引用文件、术语、章条标题和段落文本的标签标记，基于拓展后的标准标签集进行标准文献XML转化可以实现机器可读，同时便于对标准文献知识进行细粒度加工，为标准文献知识切片和重组奠定基础。

4.2 实体、关系和属性获取

根据拓展后的标准标签集对标准文献进行XML转化，实现标准文献的结构化处理。首先需要对拓展后的标准标签集进行定义，标签定义方式有DTD和XML Schema两种，由于XML Schema是基于XML语法，且对DTD的数据类型进行了扩充，可选取XML Schema对拓展后的标准标签集所包含的元素、属性和嵌套关系进行定义，同时对标准文献中的必备要素和可选要素进行定义。在标准文献XML转化时引入所构建的XML Schema文件，实现标准标签集自动生成。同时，对于PDF格式的标准文献，采用OCR文字识别技术，提取标准文献文本内容，将标准文本内容与标准标签进行关联匹配，实现标准文献XML转化，完成标准文献的结构化处理。

对于转化后的标准文献XML文件，采用Dom4j和XPath两种解析方式相结合编写Java代码实现XML文档解析，批量获取相关实体、关系和属性。具体流程如图3所示。

通过XML文档解析，可以获取标准文献本体层所对应的实体、关系和属性，并将其以三元组形式导入Neo4j中，完成标准文献知识图谱的构建和可视化，实现标准文献知识关联，从而更好地服务于标准文献的应用。

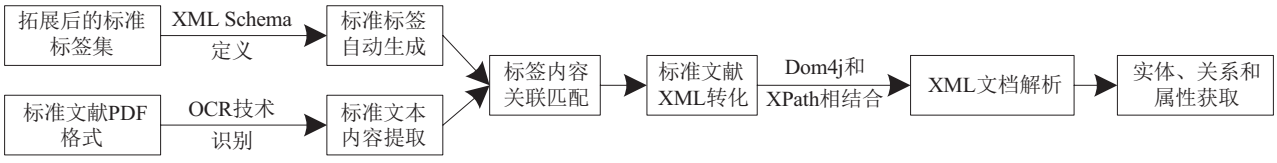

图3 实体、关系和属性获取流程

5 实证分析

本文选取自然灾害应急国家标准为实例，基于上述标准文献知识图谱构建方法构建自然灾害应急国家标准文献知识图谱，对标准文献知识图谱构建思路的可行性进行验证。

5.1 自然灾害应急标准文献知识图谱构建

本研究以自然灾害事件预防准备、监测预警、处置救援和恢复重建的应急管理流程为主线，以城市、社区、企业、应急物资等为主要对象，梳理自然灾害应急国家标准文献，其文本主要来源于国家标准全文公开系统、全国标准信息公共服务平台、中国应急信息网、中国地震局官网、中国气象局官网等，通过下载获取标准文献PDF格式共106份，其中，自然灾害基础通用标准12份、地质地震灾害应急标准31份、气象水文灾害应急标准37份、海洋灾害应急标准7份、生态环境灾害应急标准8份、生物灾害应急标准11份。所搜集到的自然灾害应急标准文献多为术语标准、分类标准、规范标准、规程标准和指南标准，试验类标准较少。

基于拓展的标准标签集，对自然灾害应急标准进行XML格式转换和文档解析，可以获取自然灾害应急标准中文名称、标准英文名称、ICS分类号、CCS分类号、发布时间、实施时间、标准号、提出单位、归口单位、起草单位、起草人、范围、术语、术语英文、术语定义、章标题、条标题和内容等信息。由此，可以完全解析自然灾害应急术语标准；分类标准可以通过进一步表格知识提取实现完全解析；对于规范标准、规程标准和指南标准，通过提取其各级标题和段落列项信息，可以实现标准知识初步解析。最终共获取实体5 039个，关系7 600个，属性值1 954个，将所得的实体、关系和属性导入Neo4j中进行存储和可视化。

5.2 知识图谱应用价值分析

（1）基于标准文献知识图谱构建方法得到的自然灾害应急标准文献知识图谱可以清晰地展示出标准文献与各组织机构间的关系，通过分析自然灾害应急标准文献与各组织机构的关系，可以辅助挖掘领域权威，为领域相关研究提供指导。例如，通过分析地震应急领域标准文献的起草人，挖掘出孙柏涛和张令心共同参与了多个地震应急标准的起草，由此可以得出两位专家在地震应急领域具有一定的权威性，同时可以推理出同时存在两位起草人的标准文献内容具有相关性；除此之外，可以进一步根据此类标准文献的引用文件情况，推断出标准文献间的相关性。

（2）通过分析自然灾害应急标准文献知识图谱，可以检测不同标准文献之间是否存在不一致等知识冲突现象。现行自然灾害应急标准文献中存在同一术语不同定义的现象。例如，共有7份标准文献中涉及“有害生物”这一术语（见图4），但对于“有害生物”这一术

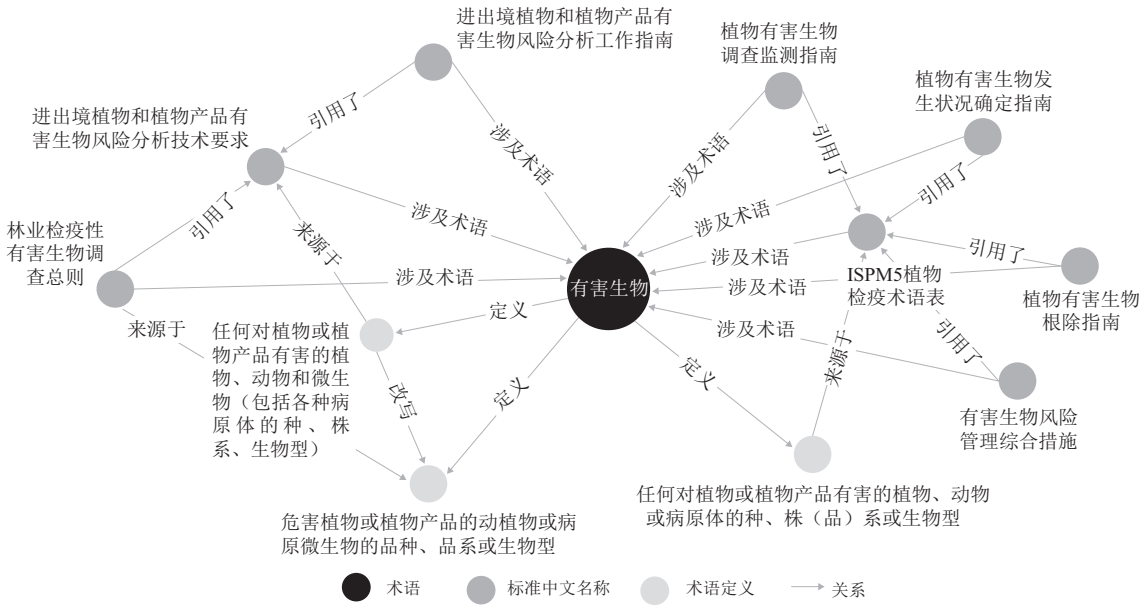

图4 同一术语不同定义和引用示例

语的定义有3种，术语存在定义改写和不同引用现象。当挖掘出标准文献内容在不同标准中存在内容冲突时，借助知识图谱易拓展性的优势，通过对单个节点进行更新修改就可以实现对涉及此内容的所有标准文献自动更新，从而消除不同标准间存在的知识冲突，减少标准更新修订时的工作量，更好地服务于标准制定者和标准使用者。

（3）自然灾害应急标准文献知识图谱实现了标准共性要素的知识关联，同时将标准文献的核心技术要素部分以标题和列项进行了细粒度展示。如图5所示，展示了《自然灾害救助应急响应划分基本要求》（GB/T 29425—2012）的范围和核心技术要素内容。标准文献知识图谱可以服务于标准全生命周期，对于标准制定者，可以通过内容检索获取标准知识现行分布情况，进行知识共享重用；对于标准审核者，可以在标准文献范围内容进行对比的基础上结合标准文献内容进行相似度审查，为内容审核提供参考；对于标准使用者，可以提高用户搜索的深度、广度和精确度，便于标准文献知识的充分应用。

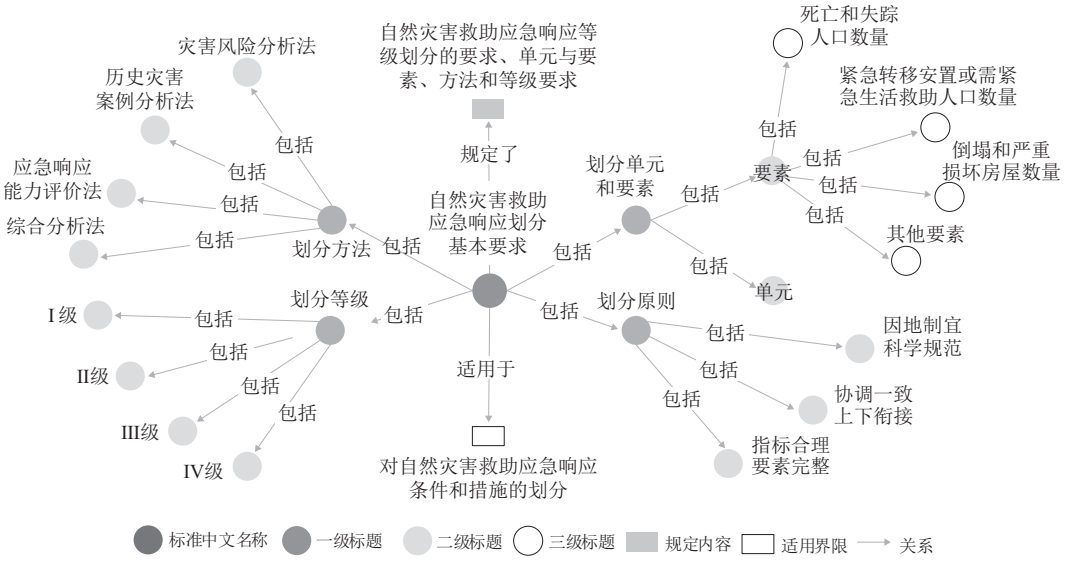

图5 标准文献范围和核心技术要素部分知识组织示例

6 结语

本研究提出了标准文献知识图谱的构建方法，通过剖析标准文献结构和文本特征，将标准文献知识进行“切片”，细化了标准文献知识粒度，建立知识间语义关联关系，进行标准文献知识重组，实现了标准文献从文档单元向知识单元的转化，借助知识图谱从而挖掘出新的知识关联关系，并初步探析了所构建知识图谱的应用价值，为标准文献知识组织和数字化发展提供了思路。现阶段所提出的标准文献知识图谱构建方法实现了标准文献通用知识的细粒度处理，不过标准文献中所包含的段落型长文本涉及的领域知识的组织模式需要进一步研究。

参考文献

[1] 丁恒，陆伟. 标准文献知识服务系统设计与实现[J]. 现代图书情报技术，2016（Z1）：120-128.

[2] 汪烁，段菲凡，林娟. 标准化工作适应全球数字化发展的必然趋势——标准数字化转型[J]. 仪器仪表标准化与计量，2021（3）：1-3，14.

[3] 王春喜，汪烁. 工业自动化领域机器可读标准研究[J]. 中国标准化，2021（S1）：27-31.

[4] 李松丽，曹平，姜盼. 国际标准化组织的标准标签集研究分析[J]. 航空标准化与质量，2018（2）：52-56.

[5] ANSI/NISO Z39.102-2017 STS：Standards Tag Suite[S]. American National Standards Institute.

[6] 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.

[7] 肖英萍，刘悦，何世新，等. 企业标准数字化实现路径初探[J]. 中国标准化，2022（8）：6-10.

[8] TEKLI J，CHARBEL N，CHBEIR R. Building semantic trees from XML documents[J]. Web semantics，2016（37/38）：1-24.

[9] ISO/TMBG SAG MRS. Questionnaire on ISO TCs’ experience

of working with SMART standards [Z]. Geneva: ISO, 2019.

[10] LOIBL A, MANOHARAN T, NAGARAJAH A. Procedure for the transfer of standards into machine-actionability [J]. Journal of Advanced Mechanical Design, Systems, and Manufacturing, 2020, 14 (2): 22.

[11] 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.

[12] LUTTMER J, EHRING D, PLUHNAU R, et al. Representation and application of digital standards using knowledge graphs [J]. Proceedings of the Design Society, 2021, 1: 2551-2560.

[13] SANA A, SUGANTHI G. Modeling and Storage of XML Data as a Graph and Processing with Graph Processor [C] // Computing & Communication Technologies. IEEE, 2017: 16-19.

[14] 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.

[15] REN H, CAI Y, ZHANG M, et al. Standard-Oriented Standard Knowledge Graph Construction and Applications System [M]. Cham: Springer International Publishing, 2021: 452-457.

[16] 张慧, 侯霞. 基于知识图谱的标准文献分析 [J]. 计算机工程与设计, 2017, 38 (4): 1103-1109.

[17] 张鹏飞, 袁志祥, 鲍威, 等. 面向绿色标准的知识图谱构建方法的应用研究 [J]. 标准科学, 2020 (6): 68-73.

[18] 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.

[19] 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.

[20] 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.

作者简介

杨跃翔，男，1976年生，博士，研究员，研究方向：标准数字化、安全与应急管理、质量管理研究。
涂新雨，女，1997年生，硕士研究生，研究方向：标准数字化、安全与应急管理、知识图谱研究，E-mail：tuxinyu010@163.com。
刘文玲，女，1990年生，博士研究生，研究方向：标准数字化、安全与应急管理、知识图谱研究。

Research on the Construction and Application of Standard Documents Knowledge Graph

YANG YueXiang TU XinYu LIU WenLing
(School of Management, China University of Mining and Technology (Beijing), Beijing 100083, P. R. China)

Abstract: In order to promote the development and application of standard documents knowledge, it is necessary to study the knowledge organization mode and method of standard documents, and promote digital transformation of standards. By analyzing the structural characteristics of standard documents, this paper constructs the ontology framework of standard documents, covering the concepts and relationships of common elements in standard documents. Through the expansion of XML standard tag set, a standard tag set suitable for Chinese standard document structure is constructed to realize machine-readable and knowledge extraction of standard documents. Then, with the help of knowledge graph construction technology, the knowledge graph of standard documents is constructed, and the application value of knowledge map of standard documents is mined with examples. This study focuses on the standard documents, and puts forward the method of constructing the knowledge graph of standard documents, which realizes the cross-association and sharing reuse of standard knowledge and helps the knowledge service and intelligent utilization of standard documents.

Keywords: Standard Documents; Knowledge Organization; Standard Digitalization; Knowledge Graph; Knowledge Service

（收稿日期：2022-05-19）