

引用格式：谭笑,刘冰,崔秀国,等.基于油气管道标准知识图谱的标准草案辅助编写技术与系统研发[J].标准科学,2025(1):36-43.

TAN Xiao, LIU Bing, CUI Xiu-guo, et al. Research on Technology and System Assisting in Writing Standard Draft Based on Standard Knowledge Graph for Oil and Gas Pipeline[J]. Standard Science, 2025(1):36-43.

基于油气管道标准知识图谱的标准草案辅助编写 技术与系统研发

谭笑¹ 刘冰¹ 崔秀国¹ 于巧燕¹ 彭泽恒¹ 曹燕¹ 郭德华²

(1.国家石油天然气管网集团有限公司科学技术研究总院分公司; 2.中国标准化研究院)

摘要:【目的】针对标准研制过程中的共性需求,开展标准草案辅助编写技术与系统研发,促进油气管道标准数字化工作的发展。【方法】以油气管道标准知识图谱为基础,研究提出标准草案辅助编写技术方案,使标准草案可直接成为机器可读标准;开展油气管道标准草案辅助编写信息系统研发,实现标准结构推荐、标准内容推荐、草案文本编辑与导出等功能。【结果】油气管道标准草案辅助编写系统可以为标准研制人员提供标准草案的自动生成服务,进一步拓展油气管道标准知识图谱的智能化应用。【结论】本文提供了一种基于油气管道标准知识图谱的标准草案辅助编写的解决思路、技术方案和实际应用,可为其他行业标准数字化发展提供借鉴参考。

关键词: 油气管道; 标准; 知识图谱; 辅助编写; 数字化

DOI编码: 10.3969/j.issn.1674-5698.2025.01.006

Research on Technology and System Assisting in Writing Standard Draft Based on Standard Knowledge Graph for Oil and Gas Pipeline

TAN Xiao¹ LIU Bing¹ CUI Xiu-guo¹ YU Qiao-yan¹ PENG Ze-heng¹
CAO Yan¹ GUO De-hua²

(1. PipeChina Institute of Science and Technology; 2. China National Institute of Standardization)

Abstract: [Objective] In response to the common requirements in the process of developing standards, this paper focus on the research on the technology and system that can assist in writing standard drafts, to promote the development of digitalization of standards for oil and gas pipeline. [Method] This paper proposes a technical solution for facilitating the writing of standard draft automatically based on the standard knowledge graph for oil and gas pipeline, which enables standard draft to become machine-readable standards; the information system for assisting in writing standard draft has been developed to achieve functions such as standard structure and content recommendation, and edit and export of draft. [Result] The system can provide automatic generation services of draft standards for standard developers, which will further expand the intelligent application of standard knowledge graphs for oil and gas pipeline. [Conclusion] This paper provides the idea, technical solution, and practical application for assisting in writing standard draft based on standard knowledge graph for oil and gas pipeline, providing reference for the digitalization of standards for other fields.

Keywords: oil and gas pipelines, standard, knowledge graph, facilitating preparation, digitization

基金项目: 本文受国家石油天然气管网集团有限公司科技项目“大口径、高压、长距离天然气管道标准产业化研究与海外应用”(项目编号: JCGL202202)资助。

作者简介: 谭笑, 本科, 高级工程师, 主要从事油气储运领域标准研究工作。

0 引言

党的十八大以来,以习近平同志为核心的党中央高度重视数字化发展,明确提出“十四五”时期要加快数字经济发展,以数字化转型整体驱动生产方式、生活方式和治理方式变革。2021年10月,中共中央、国务院印发《国家标准化发展纲要》,将“标准数字化水平不断提高”作为战略目标之一,提出“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”。国家管网集团坚决贯彻党中央决策部署,把数字化战略作为公司发展的“四大战略”之一,把数字化转型作为事关全局的系统性变革、事关长远发展的关键举措、事关员工福祉的重点任务来谋划和推动,以数字化转型支撑国家管网集团建设中国特色世界一流能源基础设施运营商,陆续部署开展了油气管道机器可读标准相关项目的研究和实践应用,研究开发了机器可读标准辅助工具,以实现机器可读标准制修订过程的数字化支撑和基于标准要素提取的智能化应用服务^[1]。同时,油气管道标准知识图谱的研究与实践应用等项目也在同步开展。

国际标准化组织以及英国、美国、德国、俄罗斯等国家,也将实现标准数字化转型纳入其标准化战略并开展了相关的研究。ISO于2019年提出了SMART 标准的新型标准概念,确定了SMART标准的发展阶段及特征,依据标准内容与机器的交互程度将标准的演进划分为“纸质文本”“开放数据格式”“机器可读文档”“机器可读内容”“机器可交互内容”共5个阶段,各阶段对数字技术的使用逐步深入,如:文本识别、可扩展标记语言(XML)、本体建模、知识图谱等^[2]。ISO和IEC联合开发的在线标准制定(OSD)平台将于2025年1月起投入使用,实现了从准备阶段至发布阶段的在线协同标准制定。俄罗斯在《俄罗斯标准化战略(2019–2027)》中明确提出制定“机器可读标准”的要求,将国家标准转换为“机器可读格式”,通过自动化系统提供标准文本的创建、编辑和应用,以及在不同系统间交换文本的能力^[3]。

在对油气管道领域技术人员进行的机器可读

标准典型需求调研中,在标准研制阶段的需求包括:标准编写需要满足严格要求,审核工作量大;需要参考大量其他标准,查询与引用难。在德国标准学会(DIN)和德国电工委员会(DKE)于2022年5月31日发布的《SMART标准用例白皮书》中,涉及标准创建者的用户需求包括:希望能够“掌握所有引用的标准和标准内容”,以便“一目了然地查看所有标准内容中的必要要求”;希望能够“知道为什么该标准机构要对标准进行修改”,以便“理解修改的原因”^[4]。从这些调研中可见,参考和引用标准问题是标准研制过程中的共性需求,标准编写格式要求也是困扰标准制定人员的普遍问题。

针对标准研制过程中的这些问题,本文以研究构建的油气管道标准知识图谱知识库为基础数据支撑,提出标准草案辅助编写技术方案,并进行系统研发,为标准研制人员提供标准草案的自动生成服务,促进标准研制的效率提高和标准的协调一致性,提高标准质量,推进油气管道标准数字化工作的深入开展。

1 油气管道标准辅助编写技术方案

1.1 标准草案自动生成思维导图

根据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》^[5],按照标准内容的功能,标准可分为术语标准、符号标准、分类标准、试验标准、规范标准、规程标准、指南标准;标准的结构包括: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、分类和编码/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。

按照各功能类型标准的要求和结构,归纳各类标准的核心技术要求和结构,作为标准草案内容生成的数据指导,标准草案内容的生成思路,如图1所示。

首先需要将某形式的标准文档中的内容提取出来并转化为能够导入Neo4j的数据格式。

其次需要一个统一的模型,检测到某标准中出现某类标准的核心技术要素时,就生成一个该类标



图1 标准草案内容生成相关思维导图

准核心要素类型的节点，挂载到该标准节点上去。

在解决上述两个问题之后，就可以通过数据导入流程将数据导入进Neo4j，之后再通过合适的算法对数据进行应用（根据关键词查询标准或是提供关键词来生成标准）。

对于每一个标准，都将成为图数据库中的一个标准类型节点，标准中的内容，例如：术语条目、范围、前言、引言、单位、其他文件，将通过包含或是引用的关系挂载到该节点上，从而表达出，该标准包含/引用/起草了这些术语条目、范围、其他文件的信息。

1.2 标准草案的XML文件标签化处理

为了使标准草案直接成为机器可读标准，在设计时直接使用已经研究的油气管道标准标签集作为模板。我们在草案编制过程中引入了标签化、结构化和数据格式规范化的原则。这意味着将每个标准条目、定义和技术规范都进行标签化处理，使其符合统一的数据模型，以便后续在数字化平台上进行自动解析和应用。

油气管道标准标签集分为两大类，通用标签集和扩展标签集。通用标签集又分为全文结构化标签集和技术指标标签集。

(1) 全文结构化标签基于标准结构元素进行标准信息单元划分，包括单位、术语和定义、封面、引言、前言、范围、与其他文件的关系等，封面实体中包括文件名称、文件英文译名、采用国际文件、采用文件编号、采用文件英文名称、一致性程度标识、文件编号、等同采用文件编号、被替代文件等属性。

(2) 技术指标标签在结构化单元基础上进行更细颗粒度的识别，标准信息单元细化到标准指标对象，包括指标名称、指标内容、指标影响因素、指标性质-定量/定性等属性。

扩展标签集根据对油气管道标准领域主要技术内容、标准应用场景等特征，分为对象和业务两个维度，对象维度和业务维度的扩展标签集示例分别见表1、表2。

表1 对象维度的扩展标签集示例

中文名	英文名	说明	数据类型
管道线路	pipeline	用管子、管件、阀门等连接管道起点站、中间站和终点站，所构成的管道运输线路	复合型
油气管道线路	oil and gas pipeline	将石油、天然气等从起点直接运输到终点的管道线路	复合型
输气管道线路	gas pipeline	将天然气等从起点直接运输到终点的管道线路	复合型
输油管道线路	oil pipeline	将石油及石油产品等从起点直接运输到终点的管道线路	复合型

表2 业务维度的扩展标签集示例

中文名	英文名	说明	数据类型
设计施工	design and construction	从项目前期设计开始到竣工，完全由一个总承包单位完成的管理模式	复合型
结构设计	structural design	将抽象的原理方案具体转化为实物，以体现其所要求的外观和功用	复合型
勘察测量	reconnaissance survey	施工前对实地进行调查测量	复合型

2 油气管道标准辅助编写系统需求分析

2.1 功能点分析

油气管道标准辅助编写系统的具体功能点见表3。

2.2 智能推荐功能

智能推荐用例中，根据用户所要生成的标准类别、关键词等信息，智能推荐相关标准，并可选择性地推荐的相关标准信息（范围、术语、实验步骤等）加入标准草案。表4详细描述了智能推荐用例。

表3 油气管道标准辅助编写系统功能点说明		
功能	子功能	功能点说明
标准草案自动生成	智能推荐	根据用户所要生成的标准类别、关键词等信息，智能推荐相关标准，并可选择性地将相关标准信息加入标准草案
	生成草案	导出草案内容，生成word文件
	人工编辑	可人工搜索并选择将标准大纲、术语、试验步骤等加入标准草案

表4 智能推荐用例描述表	
用例名称	智能推荐用例
主要参与者	用户
涉众及其关注点	用户：根据用户选择的信息自动生成推荐内容
前置条件	登录系统
后置条件	生成标准草案
主成功场景	1.用户根据需要选择需要推荐的标准类型、参考标准集合、术语名、关键词等 2.用户选择系统需要智能推荐的内容，包括大纲、术语、实验步骤等等 3.系统自动生成推荐内容
扩展（可选）	1.用户操作过程中网络断开 （1）重新连接网络 （2）若用户信息未过期，用户继续使用系统 （3）若用户信息已过期，用户重新登录后使用系统 2.用户未登录直接进入标准化对象加属性检索结果页 （1）提示用户需要登录后才能访问 （2）提供给用户登录页面的链接

2.3 标准草案自动生成功能

生成草案用例中，向用户自动推荐要生成的草案信息，并根据用户的选择生成草案，用户可导出草案内容，生成Word或XML文件保存。表5详细描述了草案自动生成用例。

3 程序描述

3.1 基于结构的推荐

方法名: recommendStructure
参数: 关键词、协作关系列表
功能描述: 该方法根据业务人员的关键词查询或选择，利用协作关系列表中的节点信息，得到知识图谱中的结构信息，从知识图谱中推荐出与用户需求相关的知识，并将结果返回给用户。

表5 草案自动生成用例描述表	
用例名称	草案自动生成
主要参与者	用户
涉众及其关注点	用户：根据系统推荐的信息自动生成草案并导出草案。
前置条件	登录系统
后置条件	生成标准草案
主成功场景	1.用户根据需要选择需要推荐的标准类型、参考标准集合、术语名、关键词等 2.系统自动生成一份草案 3.循环步骤，直至用户确认结束 （1）用户选择草案的一部分，输入更多关键词，让系统推荐更多； （2）用户选中某一个推荐，替换现有部分； 4.用户确认草案符合要求 5.选择导出文件（word、XML）
扩展（可选）	1.用户操作过程中网络断开 （1）重新连接网络 （2）若用户信息未过期，用户继续使用系统 （3）若用户信息已过期，用户重新登录后使用系统 2.用户未登录直接进入标准化对象加属性检索结果页 （1）提示用户需要登录后才能访问 （2）提供给用户登录页面的链接

实现思路分析：业务人员输入要推荐的源数据（如：标题、类别、摘要等），以及要参考引用的标准，然后利用Java去调用Python实现推荐算法。算法思路是使用Neo4j的GDS库中的FastRP算法来为每个节点生成一个低维嵌入向量，用于表示节点在图中的位置和邻居信息，然后使用GDS库中的K最近邻居算法（K-Nearest Neighbor，简称KNN），来为每个节点找到最相似的K个节点，并在它们之间建立相似关系（SIMILAR），从而实现协作过滤（Collaborative Filtering），为当前关键词推荐最相关的或者最感“兴趣”的标准。最后利用Neo4j去查询这些标准获得包含此关键词的内容或条目，并按相似度降序，从而实现基于结构的top10推荐算法。如图2所示为基于结构的推荐流程图。

按照各类标准编写的规则^[6]，基于结构推荐所涵盖的结构元素示例见表6。

3.2 基于内容的推荐

方法名: recommendSemantic
参数: 关键词

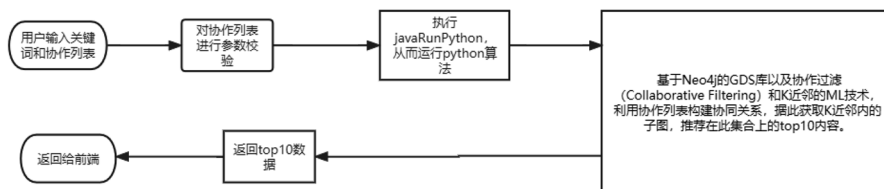


图2 基于结构的推荐流程图

表6 基于结构推荐的结构元素示例

结构元素	标准类型	推荐内容描述
1 范围	术语标准	本文件界定了[某领域]-----的术语和定义，确立了[某领域]-----的概念和体系
	符号标准	本文件界定了[某领域]的-----符号/标志，给出了对应的含义/名称和说明-----
	分类标准	本文件确立了[标准化对象/领域]-----，给出了-----的分类和代码表-----
	试验方法标准	本文件描述了-----测定/测试[标准化对象]-----的方法。 本文件描述了用-----法测定[标准化对象]-----
	规范标准	本文件规定了-----[产品、过程或服务]的-----要求/通用要求，描述了对应的证实方法，-----
	规程标准	本文件确立了-----程序，规定了-----阶段/步骤的（操作、管理等）指示，以及-----阶段/步骤之间的转换条件，描述了-----追溯/证实方法
	指南标准	本文件提供/给出了-----[某主题]的指导/建议/信息，-----
2 规范性引用文件	各类标准	下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。 本文件没有规范性引用文件
3 术语和定义	术语标准	术语条目
	其他类型标准 （选择合适的引导语）	下列术语和定义适用于本文件。 ……界定的术语和定义适用于本文件。 ……界定的以及下列术语和定义适用于本文件。 本文件没有需要界定的术语和定义

功能描述：该方法根据业务人员的关键词查询或选择，结合知识图谱中的内容信息，利用知识图谱全文索引的特性，从知识图谱中搜索出与用户需求相关的知识，并将推荐返回给用户。

实现思路：使用db.index.fulltext.createNodeIndex()过程创建一个全文索引，指定要索引的节点标签和属性。然后使用db.index.fulltext.queryNodes()过程查询全文索引，传入索引名称和查询字符串，返回匹配的节点及其Lucene查询分数，并按其相似度降序，从而实现基于内容的top10推荐算法。

基于内容的推荐时序图如图3所示。

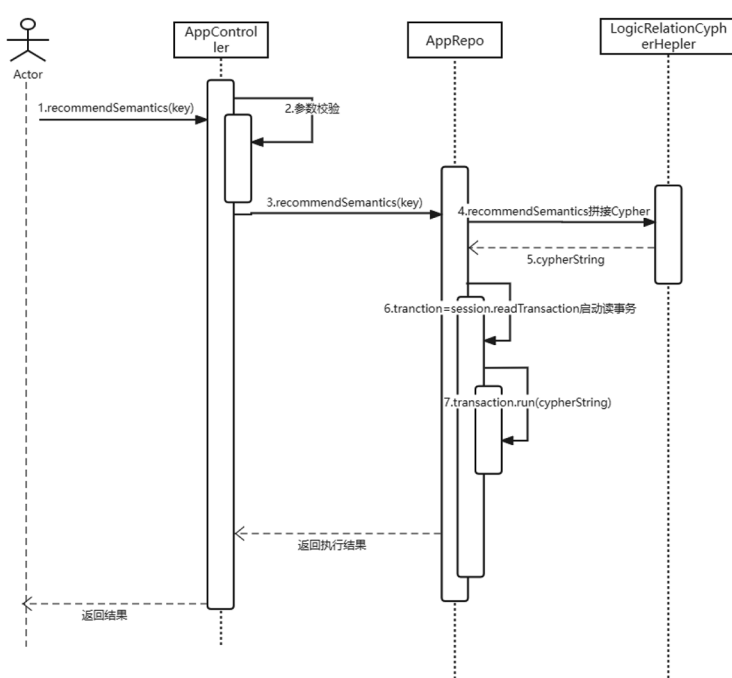


图3 基于内容的推荐时序图

3.3 获取草案模板

方法名: showDraftTemplate
参数: 标准类型
功能描述: 该方法根据业务人员选择的草案类型, 获取服务端的模板文件, 并将文件返回给用户。

草案生成实现思路: 因为标准文件具有类型效应, 即每一标准类型的标准核心技术要素是一致的, 有共同的框架结构, 因此可以基于该框架模板进行草案生成, 并结合智能推荐实现对草案内容进行不同粒度的推荐, 由此辅助生成标准草案。获取草案模板的时序图如图4所示。

4 系统实现

4.1 系统功能概述

基于油气管道标准知识图谱, 实现油气管道标准草案的辅助编写, 用户在进行草案生成工作中, 可选择两种方式。

(1) 用户可以选择标准的类型, 系统根据相关标准自动生成标准草案的XML结构。

(2) 用户也可以选择内容推荐, 通过输入关

键词、关联标准等信息告知系统所需推荐的信息, 系统显示相关推荐结果。推荐结果支持用户拷贝, 系统也支持在加工区域编辑导出。系统主页如图5所示。

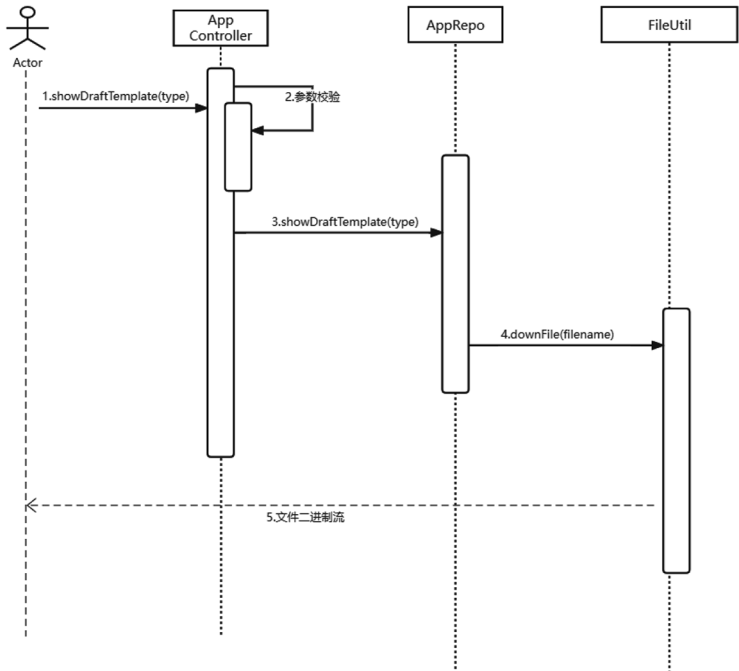


图4 获取草案模板时序图

4.2 结构推荐

选择“结构推荐”, 输入关键词, 结构推荐处

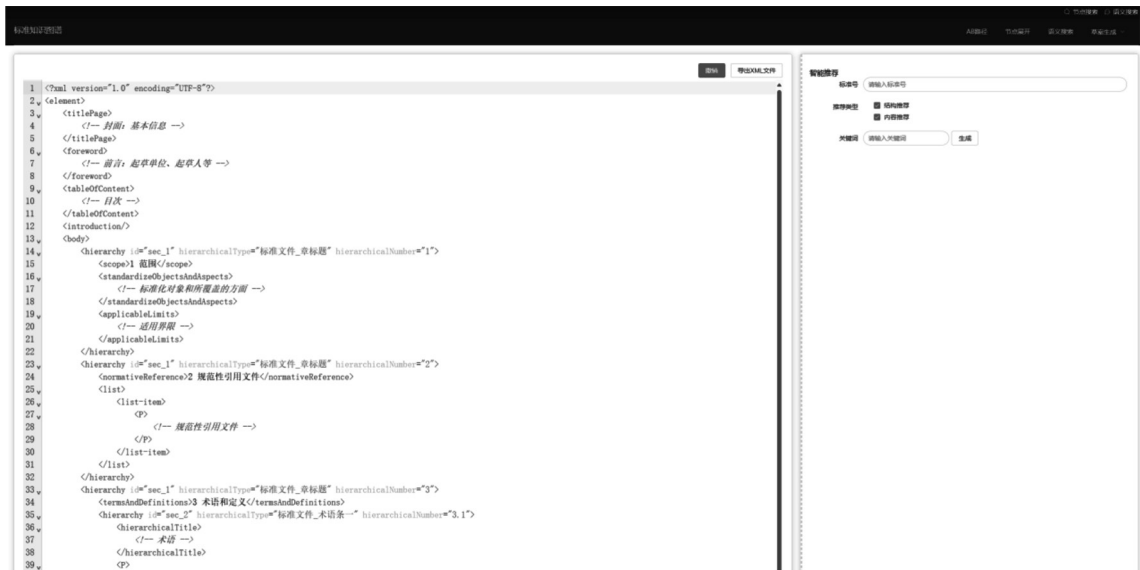


图5 油气管道标准辅助编写系统主页图

的关键词是左侧编辑区域中提供的XML内容的标签名。(例如: titlePage、foreword、tableOfContent、scope等), 点击生成按钮, 显示推荐结果列表, 选中列表中某一项, 可查看具体内容, 如图6所示。输入关键词“titlePage”, 生成4条推荐内容, 点击每条推荐内容查看具体信息, 如推荐1内容为“<titlePage><ChineseDocumentName>油气管道计量及交接技术规范 第1部分: 输气管道</ChineseDocumentName>…</titlePage>”。

4.3 内容推荐

选择“内容推荐”, 输入关键词, 点击生成按钮, 显示推荐结果列表, 选中列表中某一项, 可查看具体内容, 如图7所示。输入关键词“故障”, 生成2条推荐内容, 查看推荐1具体内容为“<P>计量回路应设置故障保护, 当运行计量回路出现通讯或电力中断等故障导致不能计量时…线色谱分析仪数据时, 应根据相邻站场天然气实际组分, 修改组分替代值。</P>”。

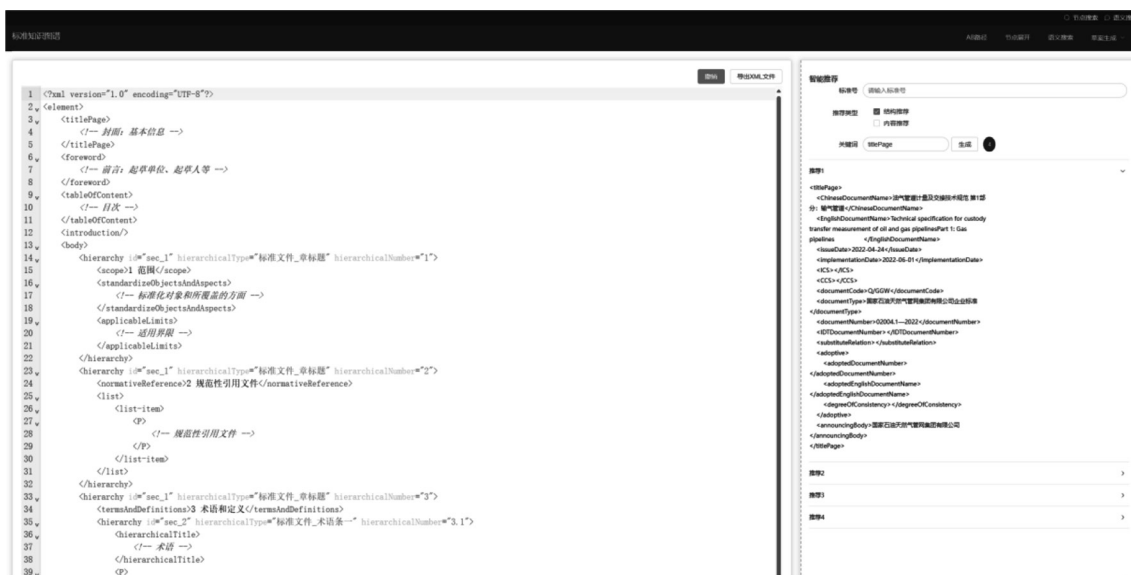


图6 结构推荐

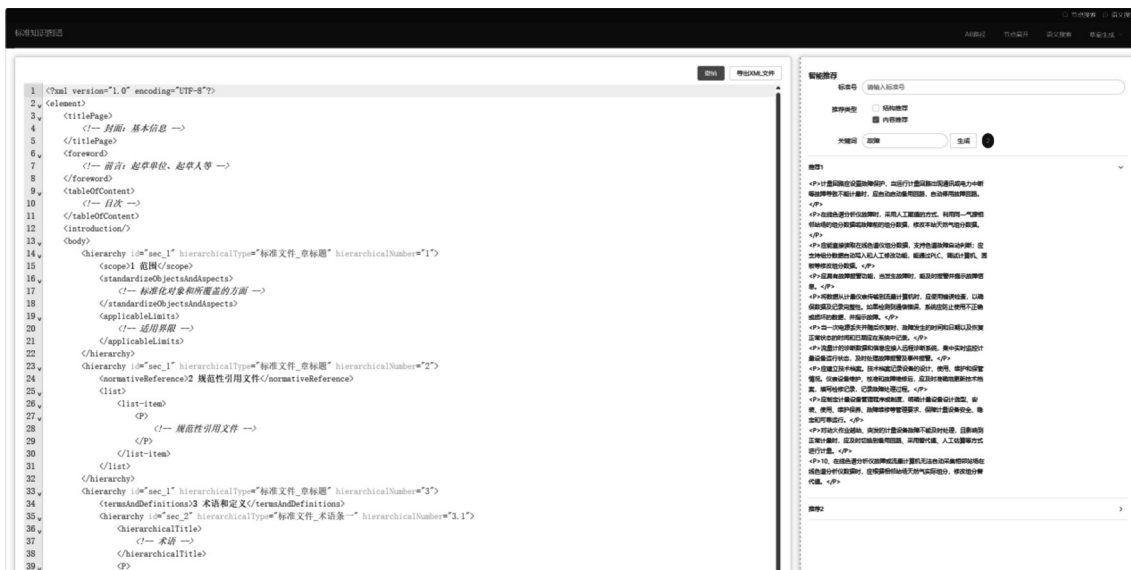


图7 内容推荐



图8 编辑修改和导出XML文件

4.4 编辑并导出

根据推荐内容,可复制到左侧编辑区域,并进行修改,点击导出按钮可将当前编辑区域的内容,以XML格式导出文件,如图8所示。将生成的内容/结构推荐的内容粘贴到左侧特定位置,如:“目次”结构下。点击“导出XML文件”按钮可将编辑好的文本内容保存到本地XML文件中。

5 结语

本文提供了一种基于油气管道标准知识图谱的标准草案辅助编写的解决思路、技术方案和实际应用,目前应用重点在于支持标准立项、标准预研等工作,可直接辅助生成油气管道标准草案。该系统目前作为油气管道标准知识图谱系统的服务功能之一,未来应用将与油气管道机器可读标准辅助系统有效结合并融合使用,在标准制定全过程中支持标准制修订,进一步拓展油气管道标准知识图谱的智能化应用,实现油气管道标准的数字化转型。该技术方案也可为其他行业提供借鉴参考。

参考文献

- [1] 刘冰,惠泉,谭笑,等. 油气管网领域机器可读标准研究与实践[J]. 标准科学, 2023 (S2): 34-41.
- [2] 刘曦泽,王益滋,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
- [3] 汪烁,段非凡,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 (03): 1-3+14.
- [4] 中国标准化研究院标准化理论战略研究所. SMART

- 标准用例分析: 基于《SMART标准用例白皮书》[EB/OL]. [2006-04-10]. https://www.cnis.ac.cn/ynbm/llyzlyjs/kydt/202210/t20221026_54091.html.
- [5] 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[S].
- [6] 白殿一, 刘慎斋, 等. 标准化文件的起草[M]. 北京: 中国标准出版社, 2020.