

油气管道标准知识图谱服务模式研究与系统研发

彭泽恒¹ 刘冰¹ 谭笑¹ 崔秀国¹ 曹燕¹ 郭德华² 于巧燕¹

(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: 知识图谱作为人工智能的重要分支,成为知识服务领域的一个新热点,而关于标准知识图谱方面的相关研究和探索也正在开展,对推动标准数字化工作具有重要意义。本文针对油气管道标准知识图谱服务模式开展理论和实践应用研究,研究提出油气管道标准之间、机构合作之间、技术内容之间等多种知识节点的标准知识关联与知识内容的可视化浏览和查询服务以及语义搜索服务模式,研究开发相应的标准知识图谱服务系统,实现油气管道标准的智能化服务,为油气管道标准数字化发展奠定了理论和实践基础。

关键词: 油气管道, 标准, 知识图谱, 服务模式, 数字化

DOI编码: 10.3969/j.issn.1674-5698.2024.12.010

Research and System Development of the Service Mode of Standard Knowledge Graph for Oil and Gas Pipeline

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

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

Abstract: As an important branch of artificial intelligence, knowledge graph has become a new hotspot in the field of knowledge services. Related research on standard knowledge graph are also being carried out, which is of great significance for promoting the digitalization of standards. This paper focuses on the theoretical and practical application research on the service model of standard knowledge graph for oil and gas pipeline. It proposes the visual browsing and discovering service model of knowledge connection and knowledge content of standards for oil and gas pipeline among various knowledge nodes such as standards, institutional cooperation, and technical content. It also proposes the semantic search service model of standards for oil and gas pipeline. The service system of standard knowledge graph is developed and intelligent services of standards for oil and gas pipeline is achieved. It will lay a theoretical and practical foundation for the digitalization of standards for oil and gas pipeline.

Keywords: oil and gas pipelines, standard, knowledge graph, service mode, 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: 彭泽恒, 博士, 工程师, 主要从事标准数字化和数据分析方面的研究。

0 引言

我国于2021年10月发布的《国家标准化发展纲要》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。”多年来，我国油气管道标准化领域在标准信息系统建设、标准内容揭示、可视化检索等方面都进行了研究探索和实践应用，虽然可以进行标准技术指标的检索查询，但还没有深入到标准文本的内容挖掘。通过对油气管道领域标准服务需求的调研，主要的痛点问题包括：标准嵌套引用造成查询与阅读困难；针对某一具体业务场景，需要查找该场景所需的标准；交叉引用及相近标准需要索引知识可追溯困难；信息孤岛，跨段落/跨标准的相关知识很难有效串联理解；引用标准版本滞后，难以识别与最新版标准的区别；标准对象所需满足的指标获取困难造成应用不便等。因此，迫切需要运用现代技术方法构建和提供标准知识发现和标准知识重组等标准智能化服务，支撑油气管道建设与运行。

知识图谱作为人工智能的重要分支，是人工智能发展的核心驱动力之一，在行业落地、产业智能化当中发挥着重要作用，成为知识服务领域的一个新热点。知识图谱的概念是在2012年由谷歌首先提出，是一种描绘实体之间关系的语义网络，国内知识图谱研究发展迅速，取得了一些突破和进展，但在行业知识图谱方面还存在很多空白。在突发事件应对^[1]、金融^[2]、油田^[3]、煤矿^[4]等行业领域，我国近年来开展了标准知识图谱方面的相关研究和探索，但在系统化服务模式研究和可实际应用的标准知识图谱服务系统研发方面尚待提升，在油气管道标准知识图谱研究与实践方面，国内外更是一片空白。基于油气管道标准知识图谱构建的研究和建设的知识库，本文针对油气管道标准知识图谱服务模式开展研究，研究开发相应的服务系统，实现为相关技术人员提供油气管道标准的智能化服务。

1 油气管道标准知识图谱的服务模式

针对油气管道标准服务需求，标准知识图谱应提供油气管道标准的可视化浏览与查询、语义搜索等智能化服务，支持油气管道领域相关人员进行油气管道标准知识浏览和学习，针对应用场景进行标准的语义搜索，从而消除信息孤岛，全面理解和掌握所需的所有标准知识。

1.1 可视化浏览和查询

在油气管道标准知识图谱中，通过标准之间、机构合作之间、标准技术内容之间、术语之间等多种标准之间的知识节点提供油气管道标准的知识关联关系和知识内容的可视化浏览和查询服务，使用户便捷、直观地了解和直接获得标准的技术内容和知识关联关系。

1.1.1 标准主要内容的可视化浏览和查询

对标准的内容结构和主要内容以可视化知识图谱的方式进行浏览和查询，包括目次、前言、引言、范围、规范性引用文件、术语和定义、符号与缩略语、技术内容、附录、参考文献等，供用户直接、可视化地了解标准的内容架构和主要内容。图1所示为标准内容结构的展示图示例，以GB/T 2820.5-2009《往复式内燃机驱动的交流发电机组第5部分：发电机组》，示例中标准包括：“4其他规定和附加要求”“5频率特性”“6过频率特性”等。图2所示为标准的主要内容展示图示意图，可直接浏览和查询到标准的多级目次以及不同层级目录间的对应关系。图3为标准技术内容的展示示例图，以标准GB/T 34346-2017《基于风险的油气管道安全隐患分级导则》为例，可直接浏览和查询到标准的标准对象、指标对象、指标影响因素、指标要求等内容。

1.1.2 标准的知识关联关系可视化浏览和查询

标准的知识关联关系即包括标准之间的相互采用、相互替代、相互修改、相互引用，也包括标准与专利之间的引用关系^[5]，对于油气管道标准，通过知识图谱的技术方法，建立“标准-采用-标准”“标准-替代-标准”“标准-修改-标准”“标准-引用-标准”“标准-引用-专利”的知识图谱，以可视化方式浏览和查询，就可以直接发现标准的多层嵌套知识关联关系，解决技术人员对标准

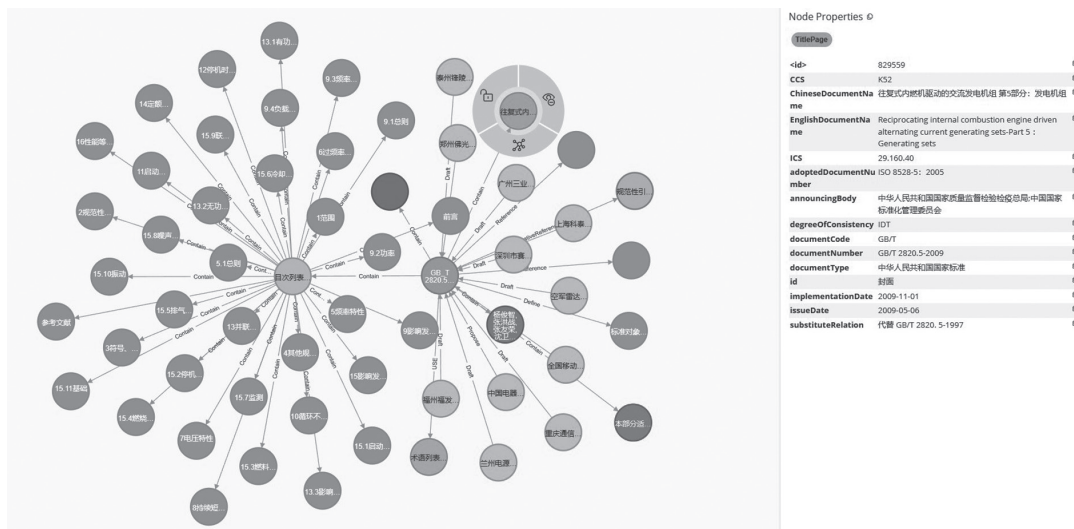


图1 标准内容结构展示图

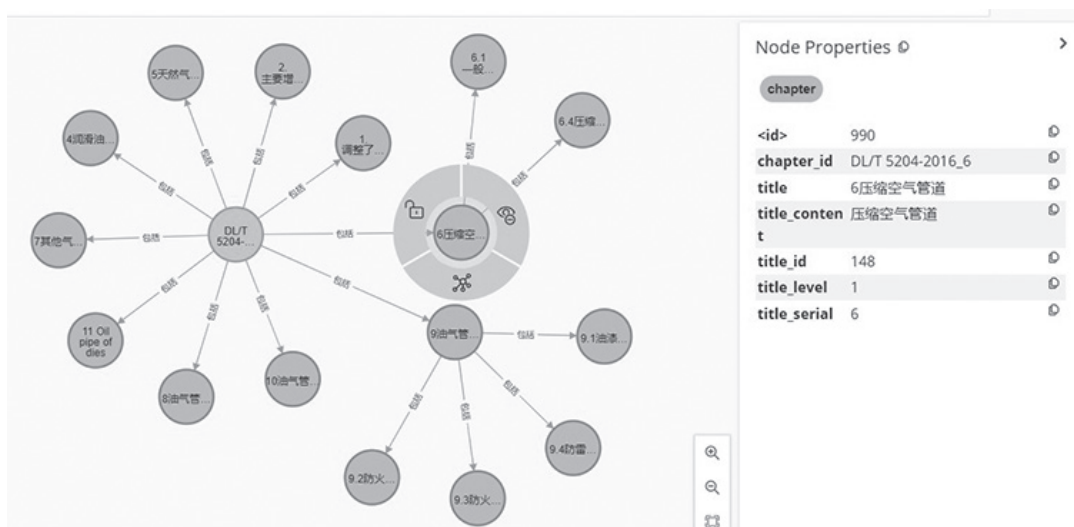


图2 标准主要内容展示图

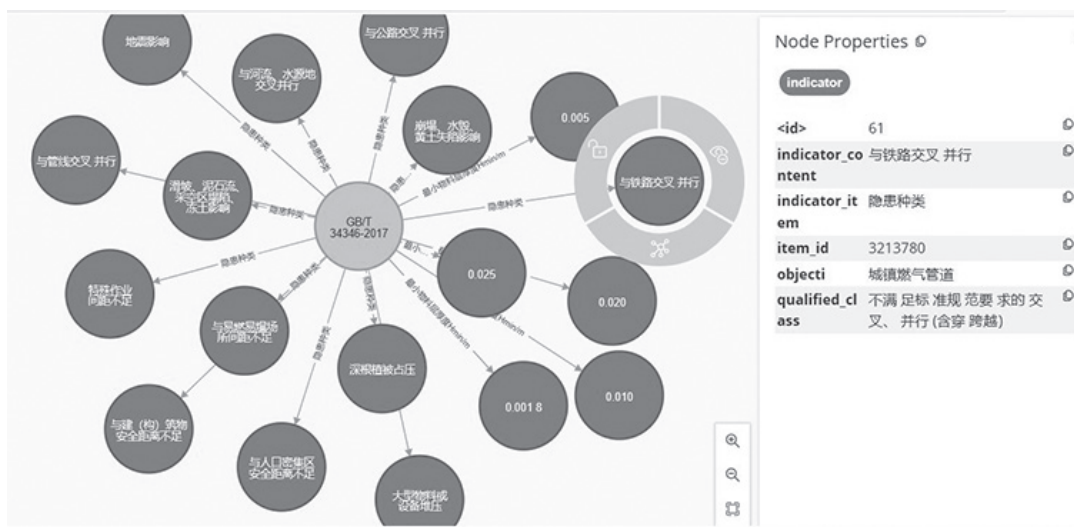


图3 标准技术内容展示示意图

的嵌套引用、交叉引用、采用以及标准新旧版本等多种工作中的难题。图4为标准知识关联关系展示示意图,可以直观发现各种关系。

1.1.3 标准机构及合作关系的可视化浏览和查询

标准中涉及机构相关的关系包括标准的发布机构、提出单位、归口单位、起草单位等,合作关系包括起草单位的合作关系,起草人的合作关系等。利用油气管道标准知识图谱服务,用户可以直观了解和分析标准发布机构、提出单位和归口单位所发布和制修订标准的总体情况,挖掘和分析制修订标准的起草单位、起草人之间的合作情况、合作紧密程度、合作领域以及起草单位和起草人牵头和参与标准制修订情况等,为标准化的规划发展和标准知识的溯源跟踪提供支撑。

图5所示为起草单位和起草人的合作情况。在图5中,以标准SY 6652-2013《成品油管道输送安全规程》为例,可以直观地看到标准的提出单位、起草单位是哪些以及标准的起草人是谁,进而可以了解到不同单位、起草人之间的合作关系。例如:标准由石油工业安全专业标准化技术委员会提出,中国石油天然气股份有限公司管道分公司、中国石油天然气股份有限公司西南管道分公司、中国石化胜利油田分公司地质科学研究院等多家单位

共同起草完成。而起草人则包括闫啸、王琪、张彦敏、董晓燕等人。



图5 标准机构及合作关系展示图

1.1.4 标准技术内容的关联知识可视化浏览与查询

标准的技术内容是标准的核心,依据油气管道的知识体系,建立标准化对象、技术要求、技术指标、术语等知识图谱,提供油气管道标准之间

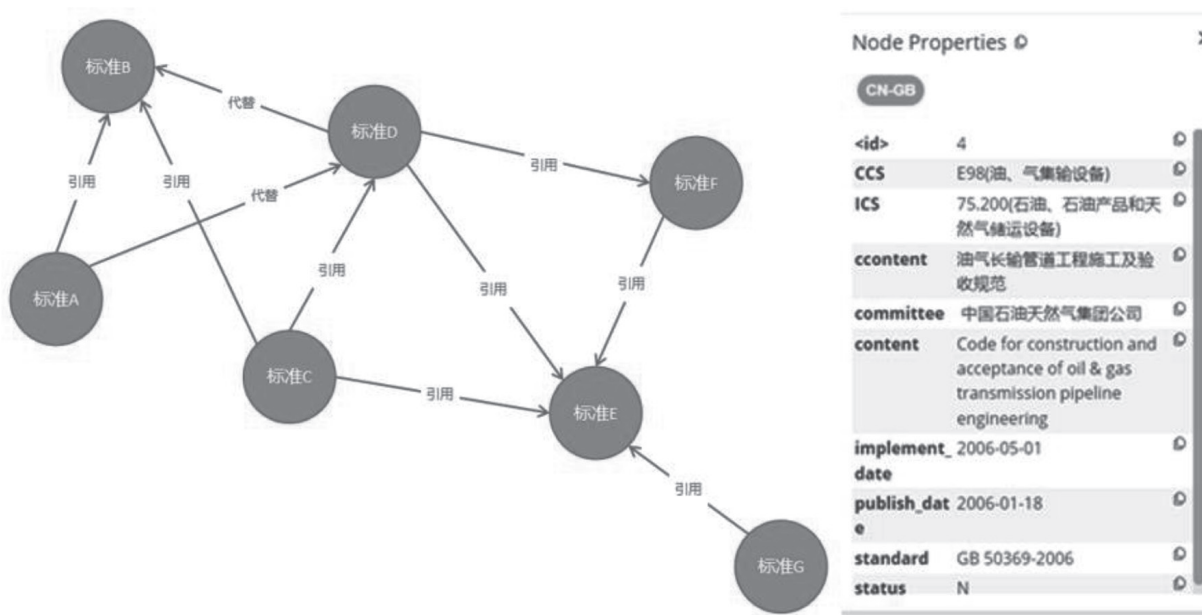


图4 标准知识关联关系展示示意图

技术内容关联知识的可视化浏览与查询,以具体示例说明如下。

图6所示为两个标准之间核心技术要素的对比与技术内容的关联示意图,可使技术人员快速掌握标准之间引用的技术内容。

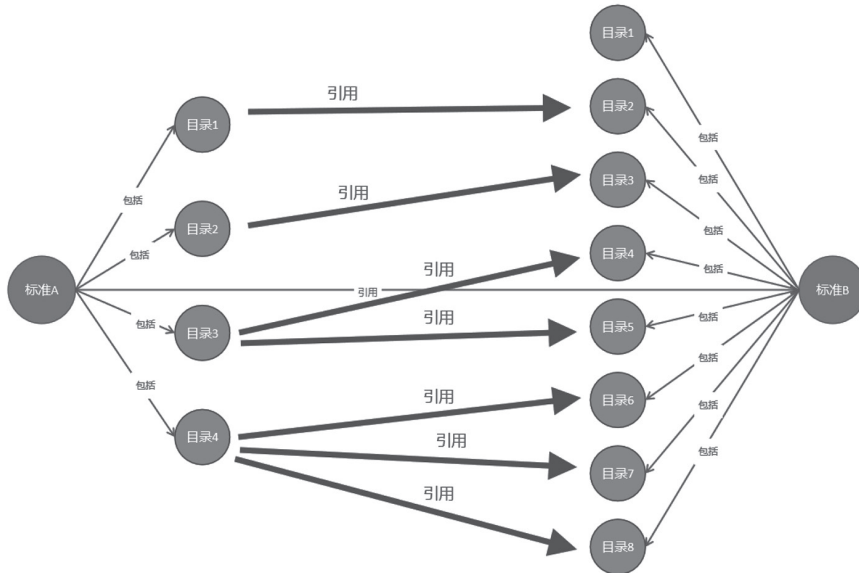


图6 标准间核心技术要素对比与关联知识图谱展示示意图

图7中,以标准SY/T 6652-2013《成品油管道输送安全规程》为例,可直接获得油气管道标准涉

及的标准对象以及相应标准对象应满足的技术指标,同时还可以通过查看同一技术指标对象的关系,直接获得不同标准文件内容上的内在知识关联。通过展开其包含“标准对象”的列表,可检索到该标准文件涉及的标准对象,以及相应标准对象应满足的技术指标。例如:

标准对象列表包括“设备、工艺”的标准对象节点。

图8中,为标准GB/T 9387.1-1998《信息技术 开放系统互连 基本参考模型 第1部分:基本模型》的关联情况,其中标准对象列表包含“标准对象”节点,节点满足多个指标对象。

1.2 语义搜索

语义搜索根据技术人员的自然语言查询,对查询进行分析

和理解,将其转换为知识图谱中的实体和关系的组合,然后利用图数据库或其他检索方式,从知识图谱中检索出与查询相关的知识,并以子图形式返回给用户,为用户提供智能化服务。



图7 油气管道标准技术指标的知识图谱

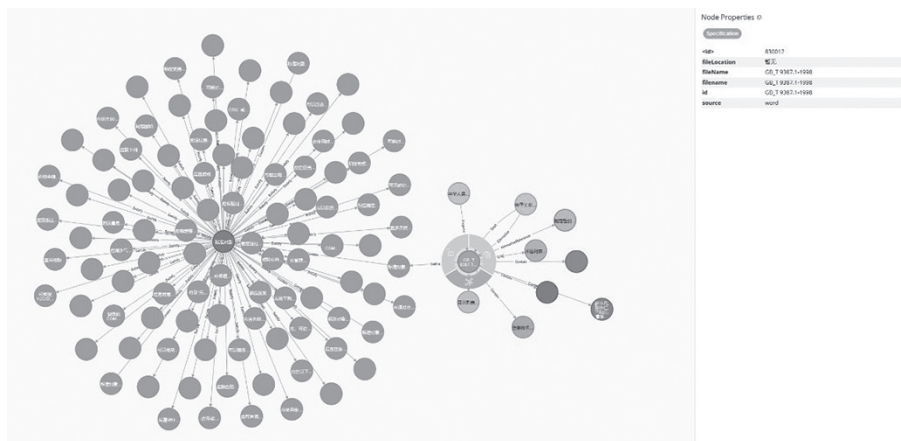


图8 油气管道标准之间的技术内容关联知识图谱

语义搜索服务的实现思路是:对于技术人员输入的自然语言句子,使用HanLP中文自然语言处理包进行依存句法分析,获得查询问句的特征词并获得关键词和词性,用对应关键词的词性等特点,利用HanLP中文自然语言处理包将关键词进行语义相似度转化为图数据库中的实体标签和关系类型,进行查询问句的抽象化并表示成知识图谱三元组,然后在图数据库中进行查询并返回子图。如图9所示为语义搜索服务实现流程图。

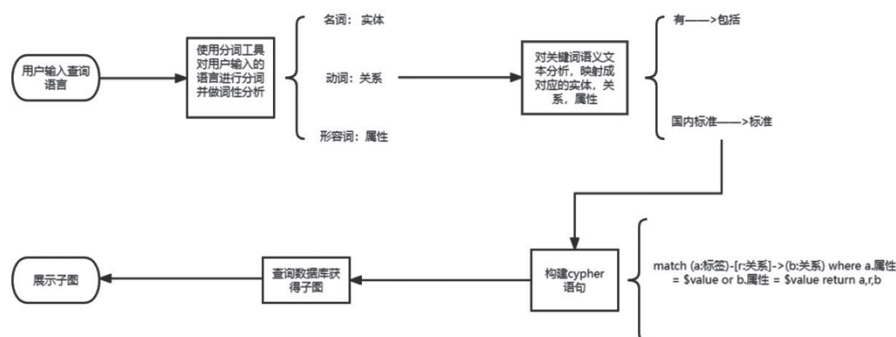


图9 语义搜索流程图

2 油气管道标准知识图谱服务系统需求分析与架构设计

油气管道知识图谱服务模式的理论研究完成后,进行相应的油气管道知识图谱服务系统研发,使用Neo4j图数据库存储知识图谱系统的图数据,系统需求分析如下。

2.1 功能点分析

油气管道标准知识图谱服务系统的功能点见表1。

表1。

2.2 系统用例分析

系统用例具体包括:

- 1) 查看数据用例: 用户查看全部抽取出的实体/关系的数据;
- 2) 实体/关系查询用例: 系统根据用户输入的实体/关系名称查询知识图谱相关实体/关系,并通过图的节点和关系形

表1 油气管道标准知识图谱服务系统功能点

功能	子功能	功能点	说明
标准知识图谱语义搜索		语义搜索并展示	将用户输入的简单查询进行关键词解析,基于关键词查询知识图谱中实体和关系,最后通过图的节点和关系形式展示搜索结果
知识图谱可视化浏览和查询	实体/关系查询		根据用户输入的实体/关系名称查询知识图谱相关实体/关系,并通过图的节点和关系形式展示搜索结果
	实体扩展		在图中,进一步展示与用户选择的实体相关的所有实体以及他们之间的关系
	路径查询		根据用户输入的两个实体名称查询这两个实体之间的6阶以内所有关联路径,并以图的形式展示

式展示搜索结果;3)实体扩展用例:用户在图中操作,进一步展示与用户选择的实体相关的所有实体以及他们之间的关系;4)路径查询用例:根据用户输入的两个实体名称查询这两个实体之间的6阶以内所有关联路径,并以图的形式展示;5)分类搜索用例中,用户可以输入类别关键词进行搜索。

数据查询相关需求比较类似,表2以比较复杂的路径查询用例为例,展示相关细节。

表2 路径查询用例描述表

用例名称	路径查询
主要参与者	所有用户
涉众及其关注点	所有用户:查询标准之间的关联
前置条件	用户登录
后置条件	展示标准之间的关联路径
主成功场景	1.用户输入关键词搜索标准1 2.系统展示搜索结果 3.用户选择某个标准作为标准1 4.用户输入关键词搜索标准2 5.系统展示搜索结果 6.用户选择某个标准作为标准2 7.系统返回标准1和标准2之间的所有关联路径,并以图的方式展示
扩展(可选)	*a.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 *b.用户未登录直接进入标准化对象属性检索结果页 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 2a5a.用户搜索的标准不存在 1.显示错误信息

2.3 系统架构设计

系统采用前后端分离的架构,技术路线为前端Vue3+js,后端采用springboot+java。应用服务器采用东方通TongWeb,关系数据库采用南大通用数据库Gbase8s,图数据库采用Neo4j。本项目以软件的内聚和耦合是软件设计的重要评价指标为落脚点,从架构设计层面尽可能降低模块间的耦合,设计出如图10所示的系统架构图,系统架构图包括了油气管道标准知识图谱构建与存储部分。

(1) 运行环境

本系统部署在Ubuntu18.04LTS版本的服务器

上,运行环境为JDK8,使用TongWeb应用服务器进行请求分发和负载均衡等。

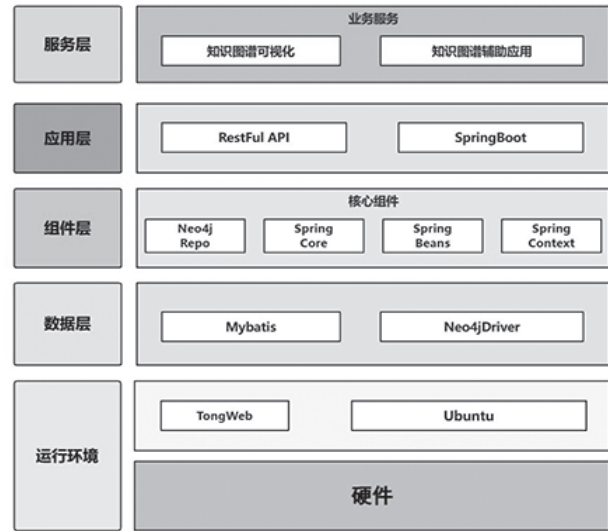


图10 油气管道标准知识图谱系统体系结构图

(2) 数据层

本项目主要采用了GBase8s关系数据库、Neo4j图数据库两种存储介质存储系统数据。GBase8s用于存储一些日志信息,以及一些需要持久化的关系型内容,Neo4j用于存储标准图结构数据,是本项目的核心数据库。Mybatis 作为ORM框架,封装了GBase8s JDBC连接的框架,从而对GBase8s进行数据访问等操作。Neo4j Driver是一个用于Java中连接Neo4j并执行Cypher语句的框架,结合Neo4j Driver,根据需求构建响应的Neo4j Cypher并执行,从而实现对Neo4j的数据访问等操作。

(3) 组件层

后端系统使用SpringBoot框架进行开发,因此系统的核心组件包括Spring框架的3个核心组件, Spring Core、Springboot Beans及Springboot Context。Bean组件用于Springboot Bean的定义创建以及解析。Context组件用于给Spring提供一个运行时的环境,用以保存各个对象的状态。Core组件包含了很多Spring的关键类,其中一个重要组成部分就是定义了资源的访问方式。Neo4j Repo封装了执行有关知识图谱操作的功能,屏蔽底层Cypher并进行ACID业务层面的事务控制,为控制器类提供所必须的服务,是Neo4j的应用类。

(4) 应用层

应用层包含业务层和控制器层两个层面。业务层实现了系统的基本业务功能,是系统功能需求的核心部分,业务层通过Neo4j Repo调用Neo4jDrive去执行Cypher语句形成业务流程,并形成结果。控制器层起着承上启下的作用,接收前端传来或者网关转发的API请求,并调用业务层相应业务,形成结果返回给前端。同时控制器层还起着参数校验等功能。Neo4j Repo有别于OOAD领域设计中访问仓库(Repository),是与数据访问层DAO以及业务层Service的一个混合,并不是对持久化代码完全隔离开来的访问仓库。

(5) 服务层

直接供用户使用的层面架构,即浏览器层,也即本系统的前端。用户通过浏览器端的交互向后端发送请求,处理请求的结果并渲染成可视化交互界面,实现系统的交互功能。

3 油气管道标准知识图谱服务系统实现

油气管道标准知识图谱服务系统提供油气管道标准知识图谱语义搜索、图谱可视化浏览查询等智能化服务,支持油气管道领域相关人员进行油气管道标准知识浏览和学习,主要功能包括:AB路径、节点展开、语义搜索。

3.1 AB路径

AB路径主要用于展示两个实体(节点)之间的关系路径,包括同类型节点和不同类型节点间关

系路径,如:两个标准之间的关系路径、某标准和某机构的关系路径等等。通过输入两个实体信息和路径步长,查询符合条件的所有路径。

在系统中,可选择第一个进行路径搜索;可在起始节点和终止节点中输入关键词进行路径搜索,并选中所显示的列表中相关项;可输入最短步长检索。通过选择左侧提供的路径选项,查看不同路径具体内容并在图谱高亮该路径。例如:起始节点选择ISO 6976《天然气 发热量、密度、相对密度和沃泊指数的计算方法》,终止节点选择GB/T 22723《天然气能量的测定》,最短步长输入3,点击确认按钮,左侧展示起始节点与终止节点间最短步长为3的检索路径结果,如图11所示,展示检索节点的关系图,高亮展示的为起始节点与终止节点通过“规范性引用文件列表442”节点连接的长度为3的路径。

3.2 节点展开

用户输入关键词查询实体(节点)一步以内的子图,即直接和该实体关联的实体以及关系子图。

如图12所示,在页面上方输入GB/T 22723,在下方提示框选择相应选项,即可在下方展示区展示与输入内容直接相关的实体及关系,例如:有“包含”关系的“规范性引用文件442”节点等。

展开工具,选择第一个为“筛选”,分为节点和关系两种筛选,会展示图谱中的节点类型和关系类型,节点类型包括“目次”“前言”“范围”等,选择相应类型后,点击确认按钮,可进行节点和关系的筛选。在两种选择中,勾选某个节点类型(或

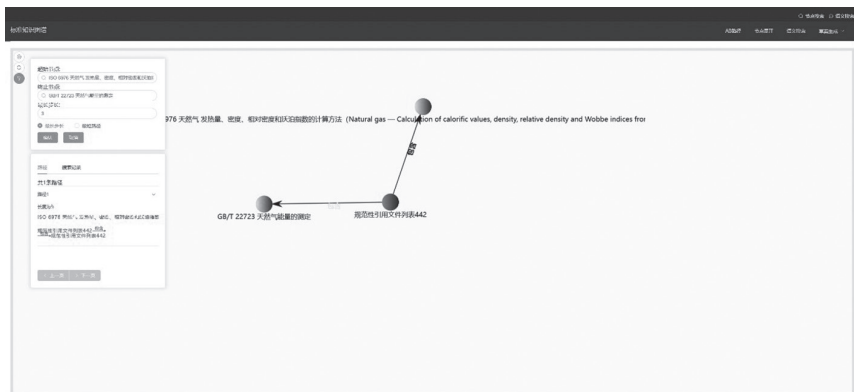


图11 路径检索结果示例

关系类型)的选中框并点击确认,即显示这种节点类型(或关系类型);取消选中某个节点类型(或关系类型)的选中框并点击确认,则不显示该类型这种节点类型(或关系类型),如图13所示,筛选结果如图14所示。

3.3 语义搜索

用户输入简单的句子,如:“**标准包含什

么”,系统显示该标准及其包含的内容。在输入框输入语句,格式为“主体-关系-客体”,按下回车键即可显示内容。

如:在页面上方搜索框中按照“主图-关系-客体”格式输入“GB/T 17286.4-2006规范性引用什么”,检索结果在下方图谱中展示,可以看到GB/T 17286.4-2006规范性引用节点“规范性引用文件



图12 节点展开显示结果

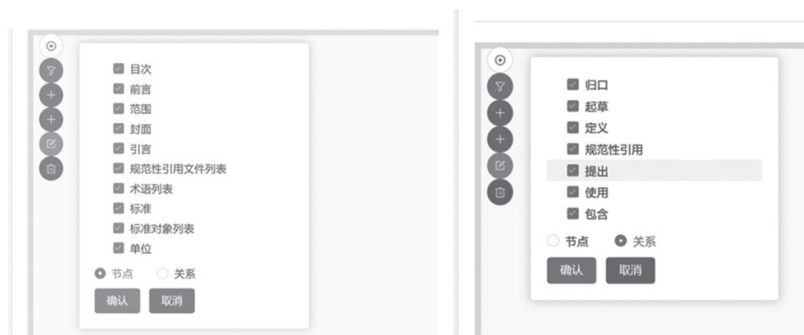


图13 节点和关系筛选选择

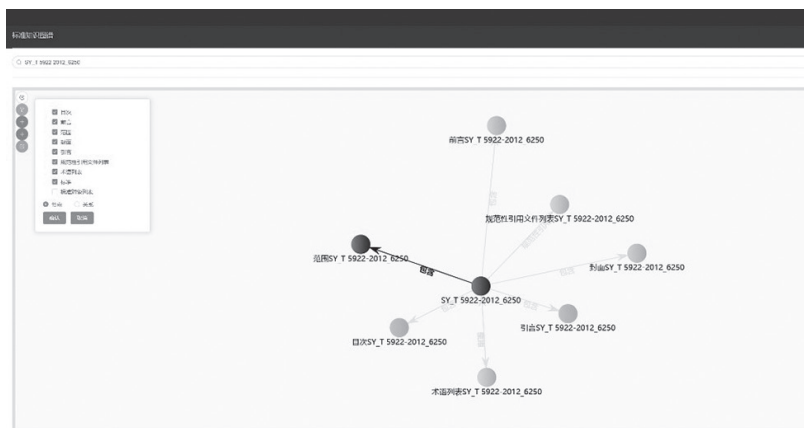


图14 筛选结果

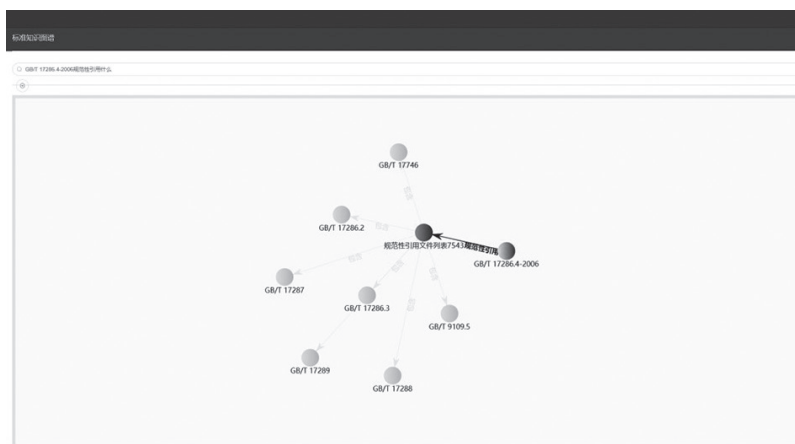


图15 语义搜索示例1

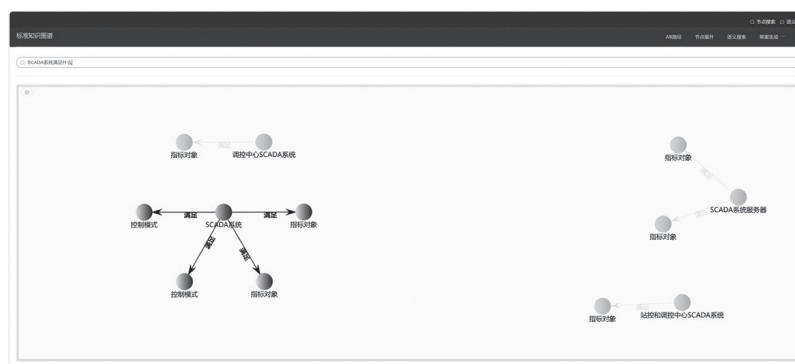


图16 语义搜索示例2

列表7543”，如图15所示。

同理，在检索框中输入“SCADA系统满足什么”，检索结果在下方图谱中展示，SCADA系统满足的节点包括“控制模式”“指标对象”等节点，如图16所示。

4 结语

本文针对油气管道标准知识图谱服务模式进

行了探索性理论研究和应用实践研究，提出了油气管道标准知识图谱的服务模式，研究开发了相应的服务系统，实现了油气管道标准知识图谱的可视化浏览和查询、语义搜索等标准智能化服务，为油气管道标准数字化发展奠定了理论和实践基础，也为我国其他行业的标准数字化和标准知识图谱研究提供了可供借鉴的理论和实践成果，将有力促进油气管道标准数字化和我国标准数字化工作的发展。

参考文献

- [1] 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):163-173.
- [2] 赵伟,张览,望俊成. 金融领域标准文献知识图谱的构建与实现[J]. 情报技术, 2022,8(6):103-113.
- [3] 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.
- [4] 刘鹏,曹新晨,耿念,等. 煤矿标准文件知识图谱构建与应用[J]. 情报工程, 2023(5): 73-83.
- [5] 郭德华. 标准文献知识链接服务模式研究[J]. 图书情报工作, 2011,55(9):76-79.