

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20222582

基于 DataG 的水库调度规程知识图谱构建方法

顾 丽, 陈瑜彬

(长江水利委员会水文局, 湖北 武汉 430010)

摘要: 随着近年来以水库为主的大量水利工程投入使用,水利部门积累了大量的水库调度规程信息。通过分析现有水库调度规程,引入知识图谱,提出了一种基于 DataG 的水库调度规程知识图谱构建方法,即在本体模型设计中,水库调度方式采用自顶向下的二次拆解法。在知识抽取时,对结构化和非结构化数据进行差异化知识抽取。在知识推理中,对前置条件、调度措施、调度目标的推理流程进行推理路径分析和定义。最后,以三峡水库为例介绍了水库调度规程知识图谱可视化展示。结果表明,水库调度规程知识图谱满足调度方式时空维度化的需求,解决了水库调度领域中多源异构数据的抽取问题,实现了水库调度方式的智能检索与推理。

关键词: 知识图谱; DataG; 水库; 调度规程; 推理路径

中图分类号: TV697

文献标志码: A

文章编号: 1000-7709(2023)07-0080-04

1 引言

水库调度规程是水库调度运用的依据性文件,水利部门在运行调度中需严格执行调度规程,保证水利枢纽工程安全,以便充分发挥防洪、发电、航运、水资源利用等综合效益^[1-2]。目前这些信息资源大部分为非结构化的文本信息,以文件形式存储于磁盘中,其存储分散、利用效率低;搜索技术为基于关键词索引的传统搜索,其内容不精准,缺乏关联性;表达方式传统的文本和表格展示,样式单一,调度规程中各对象间和对象内部的关联性弱。如何表达和利用水库调度规程,为决策者高效准确地提供关联性强的信息,成为亟需解决的问题。知识图谱最早起源于 Google,本质上是一种揭示知识之间关系的语义网络。相较于传统的存储、搜索技术和表达方式,知识图谱以三元组的形式描述客观世界中实体和实体间的关系,可有效地表达事物间的关联关系,支持更高效的数据搜索,且能模拟人类思考过程进行知识推理^[3]。在水利领域,尤其是水库调度规程方面相关的研究和应用工作还较少^[4-5]。其难点主要体现在水库调度规程中的调度方式有时间和空间条件,在本体模型和推理规则设计时扩展时空维度

较难;其次,水库调度规程包含结构化和非结构化等多源异构数据,难以用单一的抽取算法进行知识抽取。为此,本文提出了一种基于 DataG 的水库调度规程知识图谱构建方法,解决了水库调度规程在知识图谱构建中的难点问题,同时利用知识图谱的特性,有效地表达和利用水库调度规程,为决策者提供关联性强的水库调度信息。

2 DataG 平台简介

国内互联网公司百度、阿里等开始构建通用知识图谱平台^[6],其中 DataG 为阿里知识图谱开放平台。DataG 平台提供知识图谱开发、实体关系抽取、语义问答、可视化分析、开发服务等工具。其中知识图谱开发提供知识图谱的本体构建、结构化数据映射抽取、图谱数据查询及知识推理规则设计功能,实体关系抽取提供面向实体、关系及属性抽取的算法训练及打标工具,语义问答提供基于自然语言理解的 KBQA 查询检索,可视化分析提供图谱可视化分析工具,开放服务面向知识图谱开发者提供开发 API(Application Programming Interface)接口服务。本文采用 DataG 进行水库调度规程的知识图谱构建及应用,主要涉及到知识图谱开发中的本体构建、数据映射、数据导

收稿日期: 2022-12-13, **修回日期:** 2023-02-01

基金项目: 国家自然科学基金青年基金项目(52009012);国家重点研发计划(2021YFC3200301)

作者简介: 顾丽(1990-),女,工程师,研究方向为水文信息化,E-mail:1393116725@qq.com

通讯作者: 陈瑜彬(1981-),男,教授级高级工程师,研究方向为预报调度一体化,E-mail:chenyb@cjh.com.cn

人,实体关系抽取中的内容标注及开放服务中的图谱接口、节点接口、边接口和路径接口查询。

3 水库调度规程知识图谱构建

基于 DataG 的水库调度规程知识图谱构建方法自上而下包括本体模型设计、知识建模、知识抽取和知识推理四部分。本体模型设计采用自顶向下方法进行水库调度规程本体模型分析和定义,包括本体领域的确定及本体术语、实体关系和属性的定义,其中本体术语和属性的定义涉及到二次拆解。知识建模利用 DataG 中知识图谱本体建模工具构建水库调度规程知识图谱本体模型,包括实体、关系及属性的构建。知识建模对应本体层的构建,存储本体模型。知识抽取是指对结构化数据、非结构化文档中的实体、关系和属性进行提取,保存到图数据库中的过程。知识抽取对应实例层的构建,存储具体知识内容。知识推理利用 DataG 平台中的基于自然语言语法规则的图推理规则编辑引擎,通过制定一系列规则进行调度规程语义推理。基于 DataG 的水库调度规程知识图谱构建流程见图 1。

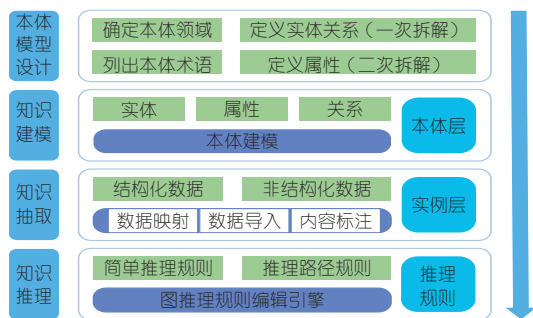


图 1 知识图谱构建流程图

Fig. 1 Flow chart of knowledge graph construction

3.1 本体模型设计

本体是对特定领域中某套概念及其相互之间关系的形式化表达。本体由实体、属性和关系组成,其中实体是指特定领域中的某个概念,关系表示两个实体间的联系,属性描述实体或关系所具有的特征。水库调度规程本体模型的设计采用自顶向下方法,具体包含以下步骤。

步骤 1 确定本体领域范围。通过分析现有长江流域大型水库调度规程,结合知识图谱构建的目的,确定本体的领域范围为水库调度规程中与防洪、发电、航运等各项调度相关的调度条件和调度方式。

步骤 2 列出本体重要术语。在此领域范围内的重要术语包括调度规程、规程对象、调度要

求、调度方式、前置条件、调度措施、调度目标和联合调度水库群及水利对象。其中调度规程与水库调度运用的依据性文件文本对应^[7];规程对象为调度规程规定的水库;调度要求是指防洪调度、发电调度、航运调度等各项水库调度项目;调度方式指明某个调度目标;联合调度水库群由多个水库组成,水库间协同合作以减轻流域或水系防洪压力;水利对象为调度方式中涉及的水库、水文站。前置条件、调度措施和调度目标构成某个调度方式的具体内容,前置条件是指采取具体调度措施的前提,调度目标为采用某个调度措施带来的结果。前置条件、调度措施和调度目标与传统的实体概念不同,有流程化的特点。

步骤 3 定义实体和关系。水库调度规程本体模型中实体和实体间的关系可用图 2 表示。其中规程对象为水利对象的子集,即规程对象继承水利对象。

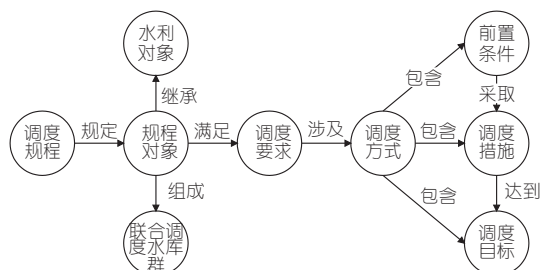


图 2 水库调度规程本体模型实体关系图

Fig. 2 Ontology model relationship of reservoir operation regulation

步骤 4 定义实体和关系属性。水库调度规程本体模型中各实体的属性见表 1。前置条件、调度措施和调度目标实体涉及到调度规程中具体的调度规则,在这三个实体的内部属性中定义对象、属性、条件(操作或范围)的目的在于进一步拆解调度方式,将非结构文本结构化,便于后续的知识推理规则设置。前置条件中的对象包括测站、城市和河段,其中城市和河段均有具体的代表测站;属性包括水情类和雨情类两类;条件包括比较类、模糊范围类和时间范围类三类。调度目标中的对象、属性和范围与前置条件一致。

3.2 知识建模

本体模型设计完成后,利用 DataG 中知识图谱本体建模工具,构建水库调度规程知识图谱本体模型,包括实体、关系及属性的构建。关系包括对称关系、非对称关系、传递关系,其中规程对象继承水利对象为非对称关系,即继承关系为单向传递。

3.3 知识抽取

在水库调度规程知识图谱中,水利对象等实

的调度措施有关。设计以下三类推理规则:

(前置条件对象,前置条件属性,满足,条件)
→(前置条件)

(前置条件对象,前置条件属性,满足,条件)
→(调度措施)

(调度措施对象,采取,操作集合)→(调度目标)

将设计好的水库调度推理规则利用 DataG 平台中的图推理规则编辑引擎编辑导入后,即可进行水库调度规程知识图谱推理,再利用 DataG 平台中的节点接口、边接口、路径接口等开放服务即可进行推理结果的查询和展示。

3.5 图谱展示

利用构建好的知识图谱,结合当前水雨情实况及预报信息,即可进行知识图谱推理路径可视化展示,推理路径图谱展示是将符合当前水雨情或预报信息的推理路径以图谱的形式进行展示,同时将同一调度方式所属的其他不符合当前流域形势的推理路径置灰。

本文以三峡水利枢纽调度规程为例来说明图谱展示方式,即根据长江中下游流域实时形势进行调度规程的推理,查找到符合当前水雨情实况及预报信息的调度方式,并进行知识图谱可视化的相关展示。推理路径展示见图 5,图 5 中横条纹

圆形代表该条推理路径的前置条件,格子圆形代表调度措施,竖条纹圆形代表调度目标,实线有向箭头表示符合当前水雨情或预报信息的推理路径,虚线有向箭头表示同一调度方式中不符合当前流域形势的推理路径。通过该推理路径可视化展示可清晰地查看调度规程的推理过程,为水库调度提供一定的技术和参考依据。

4 结论

本文通过分析现有水库调度规程,引入知识图谱,提出了一种基于 DataG 的水库调度规程知识图谱构建方法。通过此方法构建的知识图谱可将查询内容精准化、关联化展示,满足调度方式时空维度化的需求,解决了水库调度领域中多源异构数据的抽取问题,实现了水库调度方式的智能检索与图谱推理,具有一定的实用和推广价值。

参考文献:

- [1] CHAI Y F, LI Y T, YANG Y P, et al. Influence of climate variability and reservoir operation on streamflow in the Yangtze River[J]. Scientific reports, 2019, 9(1): 1-10.
- [2] 马建华. 2020 年长江流域防洪减灾工作实践及思考[J]. 人民长江, 2020, 51(12): 1-7.
- [3] 张祎, 孟小峰. InterTris: 三元交互的领域知识图谱表示学习[J]. 计算机学报, 2021, 44(8): 1535-1548.
- [4] 吴娟, 林荷娟, 季海萍, 等. 城镇化背景下太湖流域湖西区汛期入湖水量计算[J]. 水科学进展, 2021, 32(4): 577-586.
- [5] 冯钧, 杭婷婷, 陈菊, 等. 领域知识图谱研究进展及其在水利领域的应用[J]. 河海大学学报(自然科学版), 2021, 49(1): 26-34.
- [6] 李芳国, 张贝克, 高东. HAZOP 知识图谱构建方法[J]. 化工进展, 2021, 40(8): 4666-4677.
- [7] 王俊, 郭生练. 三峡水库汛期控制水位及运用条件[J]. 水科学进展, 2020, 31(4): 473-480.
- [8] CAI S Y, SUN L, LIU Q T, et al. Research on the dispatching rules of inter-basin water transfer projects based on the two-dimensional scheduling diagram[J]. Frontiers in earth science, 2021, 9: 1-14.

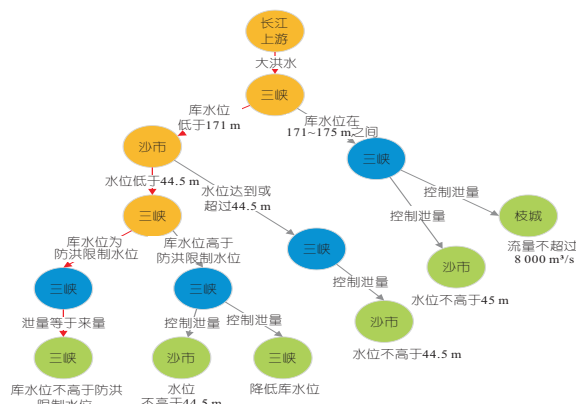


图 5 推理路径图谱展示

Fig. 5 Display of reasoning path map

Knowledge Graph Construction for Scheduling Regulation of Reservoirs Based on DataG

GU Li, CHEN Yu-bin

(Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 430010, China)

Abstract: As a large number of water conservancy projects have been put into use in recent years, water conservancy departments have accumulated a large amount of information about reservoir scheduling regulations. This paper analyzed the existing reservoir scheduling regulations, and proposed a construction method of reservoir scheduling regulations knowledge graph based on DataG. In the design of the ontology model, the top-down method of disassembling the ontology twice was adopted on the reservoir scheduling mode. The differentiated knowledge extraction was performed on the structured and unstructured data. In the knowledge reasoning, the reasoning path of preconditions, scheduling measures and scheduling objectives were analyzed and defined. Finally, the visualization of the reservoir scheduling regulation knowledge graph was introduced for the Three Gorges Reservoir. The knowledge graph of reservoir scheduling regulation meets the needs of spatiotemporal dimension, solves the problem of multi-source heterogeneous data extraction, and realizes intelligent retrieval and recommendation in the field of reservoir operation.

Key words: knowledge graph; DataG; reservoir; scheduling regulation; reasoning path