

ICS 75.180.10

CCS E92

团体标准

T/CNPC10021--2024

天然气与管道领域 知识图谱技术规范

Natural gas and Pipelines

Specification for knowledge graph construction

2024-10-01 发布

2024-12-01 实施

中国石油和化工自动化应用协会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 天然气与管道领域本体技术要求	2
5.1 天然气与管道领域知识图谱本体顶层框架定义	3
5.2 天然气与管道领域本体框架扩展要求	3
6 天然气与管道领域知识图谱建设技术要求	4
6.1 天然气与管道领域知识图谱命名	4
6.2 天然气与管道领域知识获取	5
6.3 天然气与管道领域知识图谱融合	5
6.4 天然气与管道领域知识图谱质量评估	5
7 天然气与管道领域知识图谱应用	6
7.1 天然气与管道领域知识展示	6
7.2 天然气与管道领域知识搜索	6
7.3 天然气与管道领域知识推理	6
7.4 天然气与管道领域文本生成	7
7.5 天然气与管道领域智能问答	7
8 天然气与管道知识图谱的共享安全与隐私	7
8.1 数据、算法和模型安全	7
8.2 知识图谱技术和管理安全	7
附录 A（资料性）天然气与管道知识图谱本体设计	8
附录 B（资料性）天然气与管道关系设计	9
附录 C（资料性）通用关系设计	11
参考文献	12

前 言

本文件按照 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》的规定起草。

本文件由中国石油和化工自动化应用协会标准化工作委员会提出并归口。

本文件起草单位：中国石油天然气股份有限公司勘探开发研究院、天然气销售分公司（昆仑能源有限公司）、华为技术有限公司、山东胜软科技股份有限公司、天然气销售山东分公司、中国石油大学（北京）。

本文件主要起草人：龚仁彬、林霞、李欣、米兰、窦宏恩、惠思源、刘文岭、徐超、向冲、刘宗尚、江鹰、李帅、张思博、彭朱炜、李昱池、苏涛、王珍珍、于潇、王吉庆、来守龙、张俊涛、薄叶会、林伯韬、宋先知、廖广智。

天然气与管道领域知识图谱技术规范

1 范围

本文件规定了天然气与管道知识图谱建设技术要求。

本文件适用于天然气与管道知识图谱的设计、构建、融合与应用。

2 规范性引用文件

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。

GB/T 18391.1—2009 信息技术 元数据注册系统（MDR） 第1部分：框架

GB/T 5271.17—2010 信息技术 词汇 第17部分：数据库

GB/T 23703.2—2010 知识管理 第2部分：术语

GB/T 35295—2017 信息技术 大数据 术语

GB/T 42131—2022 人工智能 知识图谱技术框架

SY/T 5922—2012 天然气管道运行规范

T/CESA 1034—2019 信息技术 人工智能 小样本机器学习样本量和算法要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

知识图谱 knowledge graph

以结构化形式描述的知识元素及其联系的集合。

[来源：GB/T 42131—2022，3.6]

3.2

本体 ontology

在大数据语境下，本体是一些约束后续各种不同层次逻辑模型的语义模型。

[来源：GB/T 35295—2017，2.1.54]

3.3

实体 entity

存在或者可能存在的任何具体或抽象的事物，包括这些事物间的关联。

[来源：GB/T 5271.17—2010，17.02.05]

3.4

属性 attribute

实体的一种命名的性质。

[来源：GB/T 5271.17—2010，17.02.12]

3.5

关系 relation

具有相同属性的各实体值的集合以及这些属性。

[来源：GB/T 5271.17—2010，17.04.01]

3.6

实体抽取 entity extraction

实体抽取是在文本中定位实体的边界并分类到预定义类型集合的过程。

3.7

关系抽取 relation extraction

关系抽取是从文本中提取实体之间的关联关系的过程。

3.8

属性抽取 attribute extraction

属性抽取是从文本中提取实体的属性信息的过程。

3.9

知识表示 knowledge representation

用于对知识进行表达的符号体系或形式化。

[来源:GB/T 23703.2—2010, 4.4]

3.10

知识获取 knowledge acquisition

知识获取是组织从某种知识源中总结和抽取有价值的知识的活动。

[来源:GB/T 23703.2—2010, 2.13]

3.11

结构化数据 structured data

结构化数据是一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的，并且可以使用关系模型予以有效描述。

[来源:GB/T 35295—2017, 2.2.13]

3.12

非结构化数据 unstructured data

非结构化数据是不具有预定义模型或未以预定义方式组织的数据。

[来源:GB/T 35295—2017, 2.1.25]

3.13

半结构化数据 semi-structured data

介于结构化数据和非结构化数据之间的数据。

3.14

天然气与管道领域语料库 natural gas and pipeline corpus

指经科学取样和加工的天然气与管道领域大规模电子文本库。

3.15

元数据 metadata

定义和描述其他数据的数据

[来源:GB/T 18391.1—2009, 3.2.16]

4 缩略语

下列缩略语适用于本文件。

URI: 统一资源标识符 (Uniform Resource Identifier)

5 天然气与管道领域本体技术要求

5.1 天然气与管道领域知识图谱本体顶层框架定义

5.1.1 天然气与管道领域顶层本体定义

天然气与管道领域是指天然气勘探、开发、储存、运输和销售的统称。

天然气与管道领域顶层本体是知识图谱中顶层概念，它提供了一个框架来组织和描述天然气与管道领域内的知识，并定义了一些基本的概念、属性和关系，天然气管道与场站本体设计见附录 A。

在天然气与管道领域顶层本体中：

- a) 概念是对天然气与管道领域内事物的抽象定义。
- b) 属性是对概念的描述。
- c) 关系用于描述概念之间的客观联系，天然气与管道关系设计见附录 B，通用关系设计见附录 C。

5.1.2 天然气与管道领域顶层本体业务范围

天然气与管道业务通常可以分为上游、中游和下游三个环节。

- a) 上游主要涉及天然气的勘探和开发。
- b) 中游主要涉及天然气的输送。
- c) 下游主要涉及天然气的销售。

5.1.3 天然气与管道领域顶层本体构建流程

在天然气与管道领域的顶层本体构建中，宜采用以下流程：

- a) 确定本体领域：明确知识图谱的应用场景和需求，确定天然气与管道领域本体的专业领域和范畴；
- b) 识别重用本体：考虑对齐、复用现有的本体；
- c) 列举重要术语：列出本体涉及的天然气与管道领域中的重要术语；
- d) 确定本体概念：梳理天然气与管道领域的知识，抽取出核心概念和概念的层次关系；
- e) 定义概念属性：定义每个概念数据属性和对象属性；
- f) 定义约束：定义概念的数据类型、关系的指向、关系的特征；
- g) 概念实例化：将天然气与管道领域的数据和知识填充到本体中，进行概念、关系和属性的实例化。

5.2 天然气与管道领域本体框架扩展要求

5.2.1 本体扩展原则

天然气与管道领域本体扩展应遵循以下原则：

- a) 一致性：新的本体扩展应与现有本体保持一致。确保新概念、关系及属性与已有的定义相协调，避免出现冲突或重叠。
- b) 可重用性：应尽量重用现有的概念和关系。避免重复定义相同或相似的内容，利用已有的本体资源和知识库来减少冗余并提高效率。
- c) 模块化：应根据不同的领域、知识主题，将本体划分为不同的模块或分层，以便更好地管理和维护。
- d) 扩展性：应保持本体的可扩展性，以便容纳未来的新知识和需求。设计灵活的本体结构，允许新概念和关系的添加，同时确保已有本体的稳定性和可用性。
- e) 明确性：在扩展本体时，应使用清晰明确的定义和术语。确保其他使用本体的人或系统可以理解 and 解释新概念的含义，避免歧义和模糊性。
- f) 可维护性：应考虑本体的维护成本和复杂性。在设计本体扩展时，应便于更新、调整和修正，

以便及时处理错误、修复 bug 以及适应领域变化。

- g) 文档化：应为扩展的本体提供清晰的文档和说明，以便理解和使用。

5.2.2 属性扩展原则

天然气与管道领域属性扩展应遵循以下原则：

- a) 一致性：新的属性定义应与已有属性保持一致性。确保新属性使用的命名规范、数据类型、值域等与已有属性相协调，避免冲突和混淆。
- b) 可重用性：在扩展属性时，应尽量重用已有的属性定义。避免重复定义相同或相似的属性，以减少冗余并提高效率。
- c) 明确性：属性的定义和含义应明确清晰，使用户或系统能够理解其含义和用途。
- d) 关联性：应考虑属性之间的关联和依赖关系。当扩展一个属性时，需要考虑与其他已有属性的关系，以确保属性之间的一致和相互补充。
- e) 准确性：添加属性时，应考虑属性的信源和可信度，以增强属性的准确性。
- f) 扩展性：属性的扩展宜通过增量的方式进行，逐步添加新的属性，方便验证和维护。
- g) 文档化：应为扩展的属性提供清晰的文档和说明，以便理解和使用。

5.2.3 关系扩展原则

天然气与管道领域关系扩展应遵循以下原则：

- a) 一致性：新的关系定义应与已有关系保持一致，避免产生冲突或混淆。
- b) 准确性：新的关系应能够准确描述本体之间的连接和依赖关系，使图谱更加详细和精确。
- c) 扩展性：新的关系应考虑未来的扩展和演化。
- d) 文档化：应为扩展的关系提供清晰的文档和说明，以便理解和使用。

5.2.4 实体扩展原则

天然气与管道领域实体扩展应遵循以下原则：

- a) 唯一性：实体扩展应保持唯一性，避免重复或混淆。
- b) 丰富性：实体扩展应注重信息的丰富性，即新加入的实体应该能够填补知识图谱中的空白，丰富知识图谱的内容。
- c) 关联性：实体扩展应考虑实体之间的关联性，即新加入的实体应该与已有实体存在一定的关联，这有助于构建更加完整和连贯的知识图谱。
- d) 准确性：实体扩展应考虑实体信息的可信度，即新加入的实体应该来源可靠，信息准确。
- e) 文档化：应为扩展的实体提供清晰的文档和说明，以便理解和使用。

6 天然气与管道领域知识图谱建设技术要求

6.1 天然气与管道领域知识图谱命名

应遵循如下规则，天然气与管道_xx 单位_知识图谱_V 主版本号.子版本号.修正版本号：

- a) 专业：指天然气与管道专业；
- b) 单位：指构建的单位；
- c) 版本：由主版本号、子版本号、修正版本号三项组成。版本命名规则：应按照 V 主版本号.子版本号.修正版本号的格式命名，各版本之间应采用“.”连接。主版本号和子版本号是必选的，主版本号可从 0 开始；
- d) 各项之间应采用下划线“_”连接。

6.2 天然气与管道领域知识获取

6.2.1 天然气与管道领域知识获取要求

知识获取应遵循以下要求：

- a) 应支持人工标注，标注时一次标注实体、关系及属性，避免一段文本多次标注；
- b) 人工标注应按照标注规范进行；
- c) 应支持主流的实体抽取及关系抽取模型；
- d) 应支持基于模板的知识获取方法；
- e) 对模型自动获取的实体及关系宜进行领域专家审核；
- f) 知识获取的准确率和召回率宜大于 80%。

6.2.2 天然气与管道领域结构化数据知识获取

结构化数据获取可参考如下规则并根据具体数据进行相应调整：

- a) 表映射到本体中的概念；
- b) 列映射到表对应概念的属性；
- c) 行映射到表对应概念的实体，主键映射到 URI、其他字段映射到实体对应属性值；
- d) 外键映射到概念的关系。

6.2.3 天然气与管道领域非结构化数据知识获取

非结构化数据宜遵循以下要求：

- a) 文本数据的知识获取应利用规则和字典法、统计法、机器学习法对实体、关系进行抽取；
- b) 图像数据的知识获取应从图像的元数据中获取，或采用人工智能图像算法对图像或视频内容进行知识抽取；
- c) 音频数据的知识获取应从音频的元数据中获取，或采用语音识别技术转换音频为文本，并从文本中抽取知识；
- d) 其他非结构化数据应从智能分析或从该数据的元数据中进行知识获取。

6.3 天然气与管道领域知识图谱融合

6.3.1 融合内容

- a) 概念层面应利用人工审核或技术手段将抽取的知识与本体中的概念对齐；
- b) 实体层面应利用人工审核或技术手段识别同一实体的不同名称，进行实体对齐。

6.3.2 融合方法

概念、实体对齐宜采用如下方法：

- a) 编辑距离：从一个字符串转成另外一个字符串所需的编辑操作数目，编辑距离越短，相似度越高；
- b) 向量距离：首先将知识表征为向量，向量为信息检索中基于词频的统计方法生成的向量，或为深度学习算法生成的向量，然后通过余弦相似度等方法计算得到向量间的相似度。

6.4 天然气与管道领域知识图谱质量评估

6.4.1 天然气与管道领域知识图谱质量评估方法

应遵循以下要求：

- a) 应允许专家对获取的知识进行人工审核。

- b) 应对融合的知识进行质量控制、舍弃专家界定的可信度低的知识。
- c) 应利用公认的衡量指标对知识获取、知识融合步骤中使用的算法进行质量评估，遵守 T/CESA 1034 中的算法要求量化方法。当衡量指标达到用户指定的阈值，即认为知识图谱获取的知识满足最低质量要求。
- d) 对知识图谱中潜在偏见和错误从数据来源、知识更新、用户反馈等方面建立监控、检测与纠正机制，保障知识图谱的准确性和可靠性。

6.4.2 天然气与管道领域知识图谱补全

应遵循以下要求：

- a) 针对知识图谱不完整的应进行知识图谱补全；
- b) 知识图谱补全分为概念层面补全和实体层面补全；
- c) 宜采用基于规则、基于统计学习等方法对知识图谱的内容进行补全。

7 天然气与管道领域知识图谱应用

7.1 天然气与管道领域知识展示

对天然气与管道领域包含合理数目实体、关系的知识图谱或子图进行可视化展示，内容应包括实体、关系、实体的属性等。知识展示的技术要求如下：

- a) 应解决计算效率问题，如分布式表示，能够高效地实现语义相似度计算等操作，显著提升计算效率；
- b) 应缓解数据稀疏问题，使每个对象均对应一个稠密向量；
- c) 应实现异质信息融合，不同来源的异质信息需要融合为整体，才能得到有效应用；
- d) 应支持复杂推理模式的知识表示，考虑关系路径的知识表示学习；
- e) 应支持展示节点数量配置，满足天然气与管道领域业务范围的展示要求。

7.2 天然气与管道领域知识搜索

针对天然气与管道领域实体、关系、实体的属性进行搜索的技术要求如下：

- a) 应支持扩展成为构建在语义 Web 上的新一代搜索引擎；
- b) 应支持天然气与管道领域本体知识库的构建、维护与进化；
- c) 宜满足推理机制，提高基于描述逻辑的本体推理技术的推理效率；
- d) 应支持结果排序，通过基于语义相似度的排序算法提高返回结果的相关性。
- e) 应支持原型系统的实现，具体方法为：
 - 1) 构建本体：使用本体构建工具创建一个天然气与管道领域的本体，包括实体、属性和关系；
 - 2) 本体推理：使用推理器对本体进行推理；
 - 3) 语义相似度计算：使用相似度计算算法，计算天然气与管道领域内术语的语义相似度，以优化搜索结果的相关性；
 - 4) 结果排序：使用统计算法，对搜索结果进行排序，确保最相关的信息排在最前位置；
 - 5) 原型系统实现：开发 Web 原型系统，集成上述工具和算法，实现知识搜索。

7.3 天然气与管道领域知识推理

提供基于规则引擎的推理能力，允许用户配置规则推理获得新知识、更新知识图谱，并提供基于图向量知识表征的推理能力，知识推理的技术要求如下：

- a) 应支持关联分析能力，基于天然气与管道领域知识图谱实现知识的关联分析；

- b) 应支持规则推理能力，通过路径和节点属性计算等方法，基于专家经验，在图上进行基于路径和节点属性计算的迭代推理能力；
- c) 应支持表示推理能力，包括但不限于通过联动图谱存储、图训练框架实现表示推理能力，通过对知识的嵌入表示，结合深度学习实现推理能力，同时可支持基于图神经网络的表示推理算法等，结合业务下游模型进行关系预测、属性预测、推荐排序等。

7.4 天然气与管道领域文本生成

天然气与管道领域文本生成的技术要求如下：

- a) 应正确识别用户所输入文本内容，基于知识图谱进行查询、推理；
- b) 应支持大规模知识推理，通过知识推理手段获取数据；
- c) 应支持异构知识关联，综合多个知识库输出报告。

7.5 天然气与管道领域智能问答

天然气与管道领域智能问答的技术要求如下：

- a) 应正确理解用户所提问题的语义内容，基于知识图谱进行查询、推理；
- b) 应支持大规模知识推理，通过知识推理手段获取隐含答案；
- c) 应支持异构知识关联，综合多个知识库给出答案；
- d) 可结合大语言模型进行相关问答、推理等应用。

8 天然气与管道知识图谱的共享安全与隐私

8.1 数据、算法和模型安全

8.1.1 数据安全应包括天然气与管道领域图数据的生命周期，基于对称加密或非对称加密等算法对数据进行加密保护，保障数据标注过程安全，保证数据质量可靠。

8.1.2 隐私保护类标准应重点防范因隐私数据过度采集、逆向工程、隐私数据滥用等造成的隐私数据安全风险。

8.1.3 算法、模型安全应包括模型文件的加密保护、算法模型鲁棒性、安全防护、可解释性和算法偏见的安全需求。

8.1.4 应保障图数据库和基于知识图谱所建立的智能系统的本质安全。

8.2 知识图谱技术和管理安全

8.2.1 天然气与管道领域知识图谱应设置数据访问控制功能，对访问人员或应用要有明确的知识范围控制。

8.2.2 天然气与管道领域知识图谱应设置管理员，负责知识图谱的权限控制和日常运维。

8.2.3 天然气与管道领域知识图谱应用应有日志监控、更新记录、操作记录等。

8.2.4 天然气与管道领域知识图谱应定时备份、异地灾备。

附录 A

(资料性)

天然气与管道知识图谱本体设计

表A. 1给出了天然气管道与场站本体设计示例。

表 A. 1 天然气管道与场站本体设计示例表

一级业务	二级业务	三级业务	四级业务
气田集输业务管道	单井采气管道		
	集气管道		
	气田集液管道		
	海底集气管道		
	储气库采气管道		
	储气库注采合一管道		
	净化气管道		
	商品气外输管道		
	海底净化气输送管道		
气田集输业务厂站	集输	气田集气站	集气站
			集气脱水站
			集气脱硫站
			集气增压站
			独立增压站
			集输线路截断阀室
			清管站
		储气库集配站	
	处理	天然气处理（净化）厂	
		储气库集注站	
	储运	独立天然气外输首站	
		独立天然气外输末站	
商品气外输管道中间站			
储气库（地下空间）			

表 A.1 天然气管道与场站本体设计示例表（续）

一级业务	二级业务	三级业务	四级业务	
天然气运销业务长输管道	输气管道			
	复线管道			
	干线			
	支线			
天然气运销业务长输厂站	分输站			
	计量站			
	压气站			
	调压站			
	截断阀室			
天然气管输业务城镇燃气网	城镇燃气管网			
	预留管道			
	庭院管网			
	立管			
	引入管			
天然气管输业务城镇燃气厂站	门站			
	调压站			
	CNG 场站	CNG 母站		
		CNG 标准站		
		CNG 减压站		
		CNG 子站		
	LNG 场站	LNG 接收站		
		LNG 工厂		
		LNG 气化站		
		LNG 加气站		
LNG 储配站				

附录 B

(资料性)

天然气与管道关系设计

表B.1给出了天然气与管道关系设计示例。

表 B.1 天然气与管道关系设计示例表

关系名	说 明	示 例
包含	A 包含 B, A 与 B 是从属关系	省份“包含”城市
位于	B 是 A 所处的位置, 或是 A 的放置位置	德州末站“位于”德州
属于	A 和 B 都是事件, A 属于 B	天然气“属于”能源
产量	A 用于解释 B	LNG“产量”50000 立方米
来源	A 和 B 都是事件, A 发生的前提是 B 先发生	能源政策“来源”政府机构
类别	A 用于定义 B	工业用户“类别”用户类别
生产	A 是 B 的一种形式	企业“生产”LNG
储量	A 是 B 的一种形式	天然气“储量”12.4 万亿立方米
运行	A 是 B 的一种形式	场站“运行”时间
属性	用于描述本体之间的关系, A 的属性是 B	时间“属性”年
等价	A 和 B 相等	LNG“等价”液化天然气
互斥	用于表示两个类或属性之间是互斥的, 不能同时成立。	供应“互斥”需求

附录 C

(资料性)

通用关系设计

表C.1给出了通用关系设计示例。

表 C.1 通用关系设计示例表

关系名	说 明	示 例
用于	A 用于 B，或者说 A 的目的是 B。	长输管道“用于”输送天然气
位于	B 是 A 所处的位置，或者是 A 的放置位置。	滕州分输站“位于”枣庄
导致	A 和 B 都是事件，A 导致 B	温度降低“导致”天然气消耗量增加
子事件	A 和 B 都是事件，B 是 A 的一个子事件	天然气消耗“子事件”城市燃气消耗
前序事件	A 和 B 都是事件，A 发生的前提是 B 先发生。	管道输送“前序事件”修建管道
后序事件	A 和 B 都是事件，A 发生的前提是 B 后发生。	天然气供应增加“后序事件”天然气价格下降
前提	A 的前提是 B。B 先发生，A 才会发生，A 依赖于 B。	天然气产量增加“前提”天然气可采储量充足
阻碍原因	B 阻碍了 A。A 是目标，B 是实现 A 过程中的阻碍或者障碍。	天然气产量增加“阻碍原因”天然气可采储量不足
同义	A 和 B 有非常相近的意思。	天然气供应“同义”天然气供给
反义	A 和 B 是意义相反或相对的词。	销售“反义”购买
定义	利用 B 对于 A 的本质特征、或者内涵和外延进行表述	消费者“定义”客户
形式	A 是 B 的一种形式。“形式”关系是动词之间的关系。	输送“形式”管输
附近	A 和 B 可以一起被找到或者发现。	阀门“附近”管道
上下文	A 是一个词，通常出现在 B 的上下文里。	天然气价格“上下文”消费需求
接受	B 动作可以作用到 A 上。	阀门“接受”关闭

参 考 文 献

- [1] 中国石油天然气集团有限公司标准化委员会信息技术专业标准化技术委员会. 勘探开发知识图谱与人工智能平台技术规范: Q/SY 10552—2022[S]. 北京: 中国石油天然气集团有限公司, 2022: 9
- [2] 龚仁彬, 李欣, 李宁等. 油气人工智能[M]. 北京: 石油工业出版社, 2021: 38-45
- [3] 刘志刚, 王禹钦, 王学力. 油气管道运行维护安全技术研究新进展[J]. 天然气与石油, 2018, 36(4): 112-118
- [4] 吴冕, 余海涛, 史博会. 天然气市场知识图谱本体构建[J]. 油气与新能源, 2022, 34(6): 71-76
- [5] 王向前, 张宝隆, 李慧宗. 本体研究综述[J]. 情报杂志, 2016, 35(6): 163-167
- [6] 周树理, 严建文, 包红林. 石油勘探开发领域本体构建及应用[J]. 计算机系统应用, 2015, 24(5): 172-176
- [7] 岳丽欣, 刘文云. 国内外领域本体构建方法的比较研究[J]. 情报理论与实践, 2016, 39(8): 119-125
- [8] 姚渝琪, 杨毛佼. 基于知识图谱的天然气技术谱系管理平台研究[J]. 电子技术与软件工程, 2022(22): 182-186
- [9] 张富利, 张恩莉, 向永惠等. 知识图谱技术在石油天然气勘探开发知识管理中的应用探讨[J]. 信息化建设, 2020(01): 128-131