

# 基于油田领域标准知识图谱的数字化服务平台研究

黄珊<sup>1</sup> 牟建荣<sup>2</sup> 王凯月<sup>1</sup> 延伟<sup>1</sup> 杨欣欣<sup>1</sup> 张俊华<sup>3</sup>

(1.中国石油化工股份有限公司胜利油田分公司技术检测中心;

2.中石化胜利海上石油工程技术检验有限公司; 3.中国石油化工股份有限公司胜利油田分公司胜利采油厂)

**摘要:** 本文介绍了我国油田企业标准化发展的现状与问题,明晰了本体和知识图谱的概念内涵,创新性地提出了适用于油田领域的“标准化对象(产品)–体例(段落结构)–指标(属性)”三元组核心数据模型,据此研究了油田领域标准知识图谱的数字化服务平台的建设思路、主要功能与未来方向。

**关键词:** 油田领域, 知识图谱, 数字化, 服务平台

**DOI编码:** 10.3969/j.issn.1674-5698.2023.05.012

## Research on Digital Service Platform Based on Oilfield Domain Standards Knowledge Graph

HUANG Shan<sup>1</sup> MOU Jian-rong<sup>2</sup> WANG Kai-yue<sup>1</sup> YAN Wei<sup>1</sup>  
YANG Xin-xin<sup>1</sup> ZHANG Jun-hua<sup>3</sup>

(1. Shengli Oilfield Branch Technology Testing Center, China Petrochemical Co., Ltd.; 2. Sinopec Shengli Offshore Petroleum Engineering Technical Inspection Co., Ltd.; 3. Data Laboratory, Shengli Oil Production Plant Management Zone)

**Abstract:** This paper introduces the current situation and problems of standardization development in Chinese oilfield enterprises, clarifies the conceptual connotations of ontology and knowledge graph, and innovatively proposes a “standardized object (product)—style (paragraph structure)—index (attribute)” triple core data model suitable for the oilfield field. Based on this, the construction ideas, main functions, and future directions of a digital service platform for oilfield field standard knowledge graph are proposed.

**Keywords:** oilfield field, knowledge graph, digitization, service platform

---

**作者简介:** 黄珊, 高级工程师, 学士, 主要研究方向为企业标准化信息化。

牟建荣, 工程师, 学士, 主要研究方向为石油领域标准化。

王凯月, 助理工程师, 学士, 主要研究方向为标准化研究。

延伟, 高级工程师, 本科, 主要研究方向为标准化研究。

杨欣欣, 工程师, 学士, 主要研究方向为企业标准化信息化。

张俊华, 工程师, 中专, 主要研究方向为石油领域标准化。

## 1 标准制定的背景和目的

石油天然气工业发展至今,已形成涵盖石油勘探、开发、油气集输、储运等十几个专业的工业体系。随着我国石油天然气工业的迅速发展,企业规模不断增长,油气勘探开发难度也不断加大,油田企业的安全形势比较严重,面临的挑战和竞争前所未有,同时暴露出的安全、健康、环境问题也愈来愈多,在石油天然气勘探开发面临的对象、要求的技术条件、新工艺、新技术应用等方面的安全与环保问题日益突出<sup>[1]</sup>。当前,世界各国公众的安全和环保意识不断提高,对安全事故和环境问题的关注也达到了史无前例的高度,作为美国油气工业界标准制定机构的美国石油学会(API)组织,其下属三大部门之一的安全健康环保部,专门负责制定及修订健康、安全与环保标准,而勘探生产部和炼油部也都分别制定其相关领域的安全与环保标准。我国的安全生产技术标准化工作,已经形成设计类标准、安全生产设备/工具标准、生产工艺类安全卫生标准、防护用品类标准、管理类标准、安全技术规程等几大安全环保标准体系。

相关研究表明,不按标准规范操作的事故时有发生,由人为失误造成的约占事故总起数的70%左右。石油石化企业从事的是高危行业,一旦发生大型或特大型事故将给社会带来严重的灾难,给企业带来巨大的财产损失,且给企业形象造成恶劣的影响。

油田安全及环保问题是油田生产当中的两个重要的问题,社会也非常关注,油田的相关部门正加强对两个方面问题的认识程度,利用正确的策略,并且采取有效的方法,让标准数字化工作在油田环保安全工作中起到决定性和规范性作用是本文研究的基本出发点。

## 2 油田领域标准化现状与问题分析

### 2.1 油田现行标准法规数量庞大,但数字化程度不高

目前,我国油田企业在用标准系统包括标准动态管理系统、标准制修订系统、油田标准查询系统、技术监督管理平台等平台,各平台的历史标准数据独立存储,标准规范和格式也不统一,对标准

的统一使用和集中管理带来不便。此外,在用企业、国家、行业、地方等标准仍以非结构化电子文档形式存在,各标准内容、条款、指标等标准核心要素数字化程度不高,在指导油田环保安全生产活动方面存在较大短板<sup>[2]</sup>。

### 2.2 标准获取方式较为传统落后,人为主观活动导致标准使用的不确定性风险较大

油田安全风险点多、分布广、作业现场分散,油田企业环保安全管理是一项典型的知识密集型工作,但这些知识分散存储在各种资料中,如:标准规范、施工组织设计、安全技术方案、事故调查报告以及其他各种技术及管理资料等。在油田环保安全管理过程中需要大量的专业领域知识的支持,这些资料由于数据量大、种类繁多、来源广泛,很难快速准确地从中获取有价值的知识。

### 2.3 标准指导安全检查工作能力不足

当前油田现场操作实施及监督人员在日常工作中常常需要随身携带大量纸质标准文档和规章制度文件,发现问题后需要耗费大量精力去查阅,但现有的搜索引擎仍以文件名称或编号进行检索,纸质或电子版的标准文档、规章制度、法律法规中涉及的知识需要花费较多的人力成本和时间成本进行解读,存在检索标准、规章制度等资源不全,检索手段落后、查全率和查准率低、检索质量不高等问题,大大影响了标准、规章制度和相关法律法规的推广普及应用。

鉴于此,本文在数据库与文件存储中对油田领域的国家、行业、地方、团体等标准数据进行数据保存,通过多源异构数据集成的方式在数据管理系统中建成标准题录数据库,之后对标准数据进行内容识别,构建标准内容知识图谱,标准知识图谱中具有标准化对象、指标、限定类、体例等内容,最终在pc端上提供油田领域诸如:题录检索、全文检索、指标展示等知识图谱的数字研究服务。

## 3 基于油田领域标准知识图谱的数字化服务平台创建的思路

### 3.1 本体

本体的概念最初起源于哲学领域,用于描述事

物的本质。本体一词产生之后,学者们便对其概念及内涵作了不懈的研究和探讨。1998年Student等人对于本体的特点给出了一个更为明确和全面的解释:

“知识本体是对概念体系的明确的、形式化、可共享的规范说明”,并且进一步指出,知识本体是“领域知识规范的抽象和描述,表达、共享、重用知识的方法”。

如果把每一个知识领域抽象成一套概念体系,再具体化为一个词表来表示,包括每一个词的明确定义、词与词之间的关系(例如:代、属、分、参关系)以及该领域的一些公理性知识的陈述等,并且能在这个领域的专家之间达成某种共识,即能够共享这套词表,上述所有就构成了该知识领域的一个“知识本体”<sup>[3]</sup>。

最后,为了便于计算机理解 and 处理,需要用一定的编码语言(例如:RDF/OWL)明确表示上述体系(词表、词表关系、关系约束、公理、推理规则等)。在这个意义上,知识本体已经成为一种提取、理解 and 处理领域知识的工具,可以被应用于任何具体的科学和专业领域。总之,本体是构建知识图谱的基础。

### 3.2 知识图谱

知识图谱是对真实世界某些特定领域中的元素以及它们之间的相互关系的抽象表示,一种计算机可计算可理解的结构化知识表示模型<sup>[4]</sup>。知识图谱采用面向对象的方式对真实世界进行抽象,将领域中的“元素”抽象为“实体”,在图谱中以节点来表示,将两个元素之间的内在联系抽象为“二元关系”,在图谱中以边(连接两个节点)来表示,将实体和关系的名称用标签标注,据此形成两个实体与关系的三元组结构(如图1所示)。

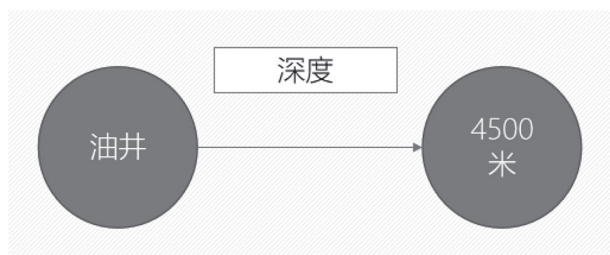


图1 知识图谱的“三元组”结构

其中,实体指客观世界中存在的人(例如:油

田工人)、事(例如:大气污染物排放控制要求)、物(例如:油井)、属性值(例如:4,500米)或者概念(例如:油田十四五规划)。关系指实体间的关系(例如:因果关系)或者属性(例如:深度)。

三元组则是由两个实体及其之间关系构成的组合。例如:实体1油井,属性为深度,实体2(它是一个属性值)为4,500米。

### 3.3 基于油田领域标准知识图谱的数字化服务平台的创建思路

针对油田领域标准文件间碎片化、标准间数据难以关联、数据结构复杂导致机器难以理解等难点,将石油行业油田企业特点、环保安全领域属性与标准文本数据特点结合,以环保安全标准规范等技术资料为底层数据,设计油田环保安全标准知识组织方法论,明确知识的展示形态;研究基于自然语言处理和机器学习的半自动化标准知识组织技术;梳理建立知识体系各要素之间的关联关系,形成面向油田环保安全领域的标准规范知识图谱,并据此构建基于油田领域标准知识图谱的数字化服务平台,为在其他领域的延展应用奠定了夯实的技术基础。基于油田领域标准知识图谱的数字化服务平台的创建思路主要有以下4个方面。

#### 3.3.1 研发基于知识本体理论的油田环保安全标准知识的组织方法论

标准知识组织方法论,是实现知识图谱中标准内容的实体识别和关系抽取的基础和依据:1)结合油田环保安全管理体系及油田标准体系,设计形成适合于油田环保安全领域的标准知识分类体系,其中,涵盖工作场景、业务流程、应用设备等多种组织维度;2)在此基础上,结合油田业务维度(如:勘探开发、地面工程、公用工程及海上工程等),基于知识本体理论,采用叙词表等组织方式,研究适用于油田企业的三元组核心数据模型,即标准化对象(产品)一体例(段落结构)一指标(属性),如:

“轻质石油产品—酸度—中和100ml轻质石油产品所需氢氧化钾的毫克数”(GB/T 258-2016《轻质石油产品酸度测定法》),其中,产品和体例均需要建立同义词和上下位的关系,指标包括:指标项、指标值、计量单位、限定类等,实现文献碎片化,形成基于知识本体理论,采用叙词表等组织方式的油田环



保安全标准知识组织方法论。

### 3.3.2 研发基于自然语言处理和机器学习的半自动化标准知识组织技术

与通用知识图谱不同,适用于油田企业环保安全领域的知识图谱面向特定领域,深度和完备性、针对性较强,实体属性较多且具有行业意义。通常,行业领域知识图谱的数据模式构建由本领域专家来承担,但专家对计算机中知识图谱的理解有限,一般建设周期长,人力成本高,知识图谱构建启动难。为大幅度降低行业知识图谱启动时的构建成本,通过对半结构化数据及非结构化数据做半自动化处理:1)以人工处理的结构化数据为训练集,应用机器学习框架,针对半结构化数据,实现自动的实体与关系标注,并构建标准知识图谱;2)以人工构建的词表和语法规则范式为基础,针对非结构化数据,实现实体识别与消歧、关系标注,并构建标准知识图谱。再由专家对关键信息进行总结,通过迭代的方式优化标注结果,供专家筛选判断,以此加快知识图谱的构建过程。

### 3.3.3 建立基于关联网络和知识图谱的标准知识基础框架

在定义及梳理研究对象,梳理对象的类型、明确对象模板的基础上,研究基于领域预训练模型的知识抽取技术,形成树形结构为主,属性分类为辅的知识架构,以及三元组为主的知识关联网络。通过研究应用主题模型、标签关系挖掘、语义表示、多层次知识融合等方法,构建集成短语级、句子级、篇章级等不同粒度层次的知识对象识别及标注的统一体系,满足通用细粒度知识对象和专业领域知识的标注和抽取需求。

### 3.3.4 创建面向油田领域知识图谱的数字化服务平台

针对石油行业油田企业在实际工作中的标准文献管理、标准文献研究、标准知识利用等应用场景,形成不同层级的标准知识图谱数据库,最终在pc端上提供包括标准知识检索、分类导航、知识图谱、信息推荐等知识服务,建成面向油田领域知识图谱的数字化服务平台,切实解决用户在标准文献搜索精度、知识发现的智能化、文献关联推荐的准确度方面存在的问题。

## 4 基于油田领域标准知识图谱的数字化服务平台的主要功能

知识图谱可以在标准数字化转型过程中得到应用和发展,以帮助用户实现更智能、高效的标准信息处理和管理。具体功能包括以下几项。

### 4.1 搜索引擎优化

通过构建油田领域标准知识图谱的数字化服务平台,搜索引擎可以更好地理解用户搜索结果,并提供更准确、有用的信息。知识图谱中包含了丰富的标准结构化数据和实体关系,可以让搜索引擎更好地理解用户查询意图,提供更准确的搜索结果。

### 4.2 内容推荐

基于用户行为和偏好,利用油田领域标准知识图谱的数字化服务平台来推荐相关内容,提升用户体验和满意度。其中丰富的语义信息可以拓展检索结果,具有更好的检索体验。

### 4.3 智能客服

油田领域标准知识图谱的数字化服务平台将本体、知识库和知识图谱结合起来,让智能客服系统能够更快速、准确地回答用户问题,提高标准信息和服务效率。

### 4.4 数据分析

油田领域标准知识图谱的数字化服务平台借助知识图谱的关系网络,进行数据挖掘和分析,发现潜在的规律和趋势,支持决策制定。

### 4.5 机器学习

知识图谱可以作为机器学习模型的输入数据,用于训练模型并提高预测准确率。

## 5 未来技术展望

目前基于知识图谱的信息化系统发展缓慢,随着深度学习、神经网络等技术快速涌现,结合油田领域的标准知识图谱的数字化服务平台也得到了推动发展,且有着广阔的研究空间。油田领域标准知识图谱的数字化服务平台未来发展方向如下。

(1) 基于神经网络的深度学习,知识图谱的质量决定标准数字化平台的体验,因此对于油田领域知识图谱的链路预测问题是不可忽视的。例如:油

田知识图谱中的实体关系若是存在缺失则会导致油气生产过程中无法完成闭环,并且因为缺少事实关系,推荐的质量可能会下降。因此将知识图谱推理和推荐任务联合部署在同一个模型中是增强推荐质量的有效途径。另外,多任务油田领域信息化系统也是智能的、有发展前景的,可以释放和减轻生产资源紧缺的压力,使生产任务更加便捷高效。

(2) 基于知识图谱的智能化油田管理系统可以通过对各种数据、设备和工艺的建模和分析,实现全面监测、预警和优化操作。这将显著提高油田生

产效率 and 安全性。建立油田领域的知识图谱,可以将各种数据、文献和专家经验进行汇聚和整合,从而支持人工智能算法开展更为精准的油田决策和风险评估。随着能源互联网的建设和发展,知识图谱可以扮演连接各个环节的重要角色,例如:从天然气输送到加工和终端应用等多个环节的效率优化和智能调度。建立行业级的油田知识图谱也将有利于推动油田领域标准化和规范化,从而进一步提高行业效率和质量水平。

#### 参考文献

- [1] 施政. 石油天然气勘探开发数据标准体系研究[J]. 中国石油和化工标准与质量, 2018,38(21):5-6.
- [2] 陈心怡,方伟,徐婷,等. 标准数字化在石油工业的探索研究[J]. 中国石油和化工标准与质量, 2023,43(05):7-9.
- [3] 王宁. 语言使用与框架设置——框架分析的本体论基础与方法论意义[J]. 广东社会科学, 2023,220(02):192-202+287.
- [4] 宫法明,董文吉,袁向兵. 基于知识图谱的潜油电泵井故障诊断[J/OL]. 计算机系统应用: 1-10[2023-04-08].<https://doi.org/10.15888/j.cnki.csa.009102>.

(上接第64页)

方开展畜禽养殖、水产养殖和有机加工互认比较的技术工作的同时,应积极建立双方有机产品认证主管机构、认证机构、认可机构定期技术交流机制,推动双方标准、认证认可和检验检测方法与结果互认,探索开展证书联网核查系统。中国有机认证要积极到国外开展有机认证,有机产品贸易便利化互认关键技术,形成不同类型有机产品认证结果互认实施方案和技术准则,并在区域国家示范应用,满足区域合作贸易畅通需求,促进区域有机产品贸易

便利化。

### 3.3 推动中国与东非贸易便利化

以有机产品为抓手,充分利用中非技术性贸易措施研究评议基地、中非经贸博览会等平台资源,引导企业参与非洲农产品标准的制定、技术性贸易措施通报评议等工作,提升企业应对技术性贸易措施主体意识,营造企业深度参与中非贸易的良好氛围,不断促进中国与非洲地区食品及农产品的贸易与合作。

#### 参考文献

- [1] OJ Lim Tung: African Organic Product Standards for the African Continent? Prospects and Limitations [R].2018(8).
- [2] EAC Secretariat Arusha, Tanzania: EAST AFRICAN COMMUNITY VISION 2050.2016(2).
- [3] 国际有机联盟: world-organic-agriculture-2021.
- [4] EAS 456:2018 Organic production standard[s].
- [5] GB/T 19630-2019 有机产品 生产、加工、标识和管理体系要求[S]. 北京: 中国标准出版社, 2019.