

油气行业标准内容技术指标比对方法研究

徐 婷¹ 张自成² 李 茹¹ 张 华¹ 李炎华¹ 吕 华¹

(1. 中国石油集团工程材料研究院有限公司; 2. 大庆油田有限责任公司测试技术服务分公司)

摘 要: 油气行业标准内容技术指标比对对于明确油气行业发展目标和提升油气行业质量效率具有重大理论和现实意义。本文明晰了油气行业标准指标比对的概念内涵, 分别从人工比对与机器自动比对等两方面对现有研究进行了总结和归纳。围绕油气行业标准技术指标比对的业务流程, 提出了油气行业标准指标自动比对的关键技术的解决方案。在此基础上, 研判了油气行业标准指标比对技术或方法面临的问题, 并提出了针对性的对策建议。

关键词: 油气行业, 标准指标, 比对, 人工比对, 机器辅助比对, 对策建议

DOI编码: 10.3969/j.issn.1674-5698.2024.05.015

Research on Comparison Method of Technical Index in Standard Content in Oil and Gas Industry

XU Ting¹ ZHANG Zi-cheng² LI Ru¹ ZHANG Hua¹ LI Yan-hua¹ LYU Hua¹

(1. CNPC Tubular Goods Research Institute; 2. Logging & Testing Services Company of Daqing Oilfield Co., Ltd.)

Abstract: The comparison of technical index of oil and gas sectoral standard content has great theoretical and practical significance for clarifying the development goal of the oil and gas industry and improving the quality and efficiency of the industry. In the paper, the concept and connotation of standard index comparison in oil and gas industry are clarified, and the existing researches are summarized from two aspects: manual comparison and machine automatic comparison. Centering on the business process of standard technical index comparison in oil and gas industry, this paper puts forward the solution of key technology of automatic standard index comparison in oil and gas industry. On this basis, the problems about the technology or method of standard index comparison in oil and gas industry are evaluated, and the corresponding countermeasures are put forward.

Keywords: oil and gas industry, standard index, comparison, labor comparison, machine aided comparison, countermeasures and suggestions

基金项目: 本文受国家重点研发计划项目“油气能源产业链关键国际标准研制-油气能源勘探开发领域国际标准化研究”(项目编号: 2022YFF0609700-2022YFF0609701) 和中国石油科学研究与技术开发项目“石油标准资源数字化建设与应用研究”(项目编号: 2022DQ0108-17(JT)) 资助。

作者简介: 徐婷, 正高级工程师, 研究方向为石油装备与管材标准化。

张自成, 高级工程师, 研究方向为油气勘探标准化。

李茹, 工程师, 研究方向为石油标准数字化。

张华, 正高级工程师, 研究方向为石油标准数字化。

李炎华, 正高级工程师, 研究方向为石油装备与管材标准化。

吕华, 高级工程师, 研究方向为石油管材标准国际化。

0 引言

目前,“标准比对”尚没有明确的定义。百度百科对“比对”一词的解释是比较对照、核对;全国科学技术名词审定委员会审定的化学术语“比对”指通过对比试验确定和评价分析方法的可靠性、实验室的水平以及分析者技能的操作;与标准比对近似的电力术语“比对标准”指用于同准确度等级的标准器之间相互比对的标准器。

本研究尝试将“标准比对”定义为:针对同一标准化对象,对不同(同类或有代替关系)的标准文件按照相同或相近的规则、指南或特性相互比对,以确定和评价这些规则、指南或特性的差别。标准比对起源于20世纪70年代美国的对标管理,最初是人们利用对标寻找与别的公司的差距,把它作为一种调查比较的基准的方法。后来,对标管理逐渐演变成为寻找最佳案例和标准,加强企业内部管理的一种方法。我国对标准比对有一个逐步认识与拓展的过程,近年来,我国进出口企业和一些研究机构基于自身需求做过大量的标准比对,我国的标准比对研究开始于畜牧业,主要是国内外产品质量标准的不同要求或具体指标差异进行对比,包括国内外羊肉质量对比^[1]、梨质量国内外标准对比^[2]等。

油气行业在国民经济中具有重要地位,其发展状况对经济、政治和军事等方面都有着深远的影响。石油和天然气是重要的能源和化工原料,广泛应用于交通、工业、农业等领域,满足人们日常生活需求。当前,在油气行业,标准化意识相对薄弱,主要表现在油气管网建设和运营的参与主体较多,涉及指标范围广泛,缺乏统一规范的组织与引导。随着油气管网建设的快速推进,油气行业标准指标比对工作也相对滞后,同时由于新技术的快速发展和市场需求的变化,导致已有的标准指标不能完全适应当前的发展需求。尽管我国已经发布了一些与油气勘探开发相关的标准,但在油气勘探开发、生产及储运等环节仍存在许多不规范的行为。因此,需要开展油气行业标准指标比对工作,进一步加强行业规范。鉴于此,针对目前国内

外重点发展领域标准指标的比对需求,本研究将开展标准指标比对的通用方法和技术研究,在标准文献结构分析基础上提出标准内容指标比对的通用技术和方法,重点解决国内外重点发展领域标准比对技术问题,实现国内外重点领域标准内容获取、标注和比对,为产业发展和技术创新提供技术和数据支撑。

1 研究现状

1.1 专家主导的人工标准比对

标准作为专业技术文件,包含有众多技术指标,随着生产和管理需求不断深入和扩大,同一对象涉及的标准不断增加,相关技术指标也越来越多。因此在做人工对比时,大部分是根据具体比对目标和需求,专家提取核心指标进行对标对比,比对标准范围包括新旧标准比对和国内外标准比对。

在国内外标准比对研究方面,目前研究内容集中在具体产品或领域的核心指标统一认定和提取上,采用方式和流程大都相似,比对目的是从技术标准角度说明产品的差异,研究给出直观对比分析结论,目的是通过对比给产品生产和研究提供技术参考。如:陈晓穗等对国内外LED台灯标准进行了比对,对安全、电气性能、光度学性能、色度学性能等要求和测试方法进行了对比分析,对比全球各地标准对LED台灯的要求^[3]。朱晓春等进行了中蒙《铜精矿·二氧化硅含量的测定方法》标准指标比对研究,对蒙古国国家标准与中国国家标准中的方法类标准指标进行了对比^[4]。

在国内新旧标准比方面,目前研究成果集中在新增或差异指标的选择和指标值差异研究上,服务于新旧不同标准的宣贯和进一步的新技术宣传,如:丁莉等对我国轻型车国六与国五排放标准进行对比分析,方便使用者理解标准差异^[5]。马冬妮等通过研究三峡工程水泥新旧标准比对方法,得出在我国新旧标准水泥强度检验方法使用上有着很大的差异,主要表现在试验设备、使用标准砂、胶砂组成、人员操作、养护条件等指标要求上,以及在计算方法上的差异等,为相关企业和技

术人员使用新标准提供相应的解释材料,节省产业熟悉新标准时间,进一步提高效率^[6]。

在比对流程方面,蔺菲等在实验室间电能计量标准比对的方法与数据中,介绍了电能计量标准实验室间标准比对的过程、方法,通过实例给出比对结果,分析试验数据并进行数据处理,得出比对试验的结论,通过这一比对流程,真实地反映参加比对工作的实验室电能计量的综合技术水平^[7]。付卉青和刘霞在开展消费品安全标准比对方法研究中,把消费品安全标准对比的步骤分为4步:确定消费品比较的范围,收集和整理资料,建立消费品安全标准比对指标体系,比较分析并得出结论^[8]。

1.2 基于技术手段的自动比对研究

由于人工开展标准比对的效率问题和开展比对人员的专业局限性,标准比对的范围和比对结果时效性受到极大制约。近年来国内技术人员也对标准自动比对进行过研究,技术核心是解决对海量标准文本进行标准内容指标提取处理,从而建立基础的指标数据库,再通过语义识别、归一化处理等技术,加上机器自动比对与专家经验比对的综合分析,对相关标准化对象的不同标准进行标准指标比对分析,最终输出结果。目前整体研究还处于起步阶段,当前研究成果集中在前期的不同类型标准数据库建设和语言处理上。

计雄飞等提出实现标准文献内容挖掘与比对需要经过标准文献文本识别、特征提取、内容加工、内容检索4个步骤^[9]。王昕等基于语义网的理论,研究一种标准指标比对的方法,以“产品—体例—指标”三元组的方式将标准文献碎片化,对细粒度的指标碎片进行知识化组织,最终形成“标准指标比对”的应用系统并应用于电力行业^[10]。周生龙等提出了在古文献版本差异比对中的图像文字自动比对方法,主要内容包括信息采集与分类、文献图像预处理、文字切分与存储、文字比对与文献差异标注等^[11]。吴建港等以标准制修订、标准内容研究、指标比对分析、产品质量提升等为最终目的,在传统标准指标比对分析研究的经验基础上,通过构建以“标准体系—标准—产品—指标项—指标值”为数据模型,引入图像识别技术,

构建指标库^[12]。

1.3 研究述评

目前开展的标准比对工作大都选定具体产品(对象)采用人工方式进行,需要专业技术人员从大量标准中找到相关的、有差异的信息,虽然有能对Word、PDF等格式的标准文献全文文件进行检索的软件工具,但是很难快速获取所需要的信息,缺乏油气行业标准内容技术指标比对方法研究,主要由于:(1)油气行业标准文献数量巨大,题录检索和手工查找难于满足需求;(2)油气行业纸质的标准文献,虽然能够扫描制作成PDF等格式的电子文件,建立了全文数据库,实现了电子化馆藏,但由于本身未进行文字识别,实现不了全文检索功能;(3)现有油气行业Word、PDF标准文献全文文件为非结构化数据,很难实现如:结构化数据那样的检索、提取、挖掘和比对分析等功能。对于需要开展比对的油气行业技术人员和企业来说,如何从海量的基础标准、技术标准、管理标准、工作标准、产品标准中快速获取所需相关信息显得尤为重要。

2 研究方法

(1)文献研究法。查阅相关资料,学习研究相关理论。着眼于标准比对的内涵和标准自身特征,进行综合分析,力求方法的创新。

(2)调查研究法。调查当前开展油气行业标准内容技术指标比对工作的现状,分析判断当前油气行业标准比对工作开展的影响和制约因素,了解掌握丰富的第一手资料,为任务研究提供强有力的支撑。

(3)知识库方法。使用专家+计算机结合的方式,进行油气行业比对标准集成、内容指标提取,构建包含比对标准文本、内容、指标、比对结论、比对专家等相互关联的知识库。

3 油气行业标准技术指标比对的实现路径

3.1 油气行业标准指标比对的流程

要开展油气行业标准内容指标比对,实现比对任务,均需要在油气行业不同的标准中找到相关内容,并根据具体判定标准进行对比,给出结论。按照解决问题的操作流程,需要解决3个关键问题:需要比对什么标准?需要比对的内容、指标是什么?结论是什么?

要解决这3个问题需要研究油气行业标准的集成组织方法和工具,标准的结构化分析方法和工具,内容指标提取方法和工具以及标准内容指标组织、关联、比对方法和工具。本研究提出的业务流程如下。

(1) 梳理油气行业中外技术标准体系,确定油气行业比对基准标准及相关标准集合;

(2) 建立油气行业标准专业技术树,确定标准比对指标点;

(3) 结合油气行业专业技术树选择待比较标准;

(4) 识别油气行业中外技术标准中的关联条款;

(5) 对油气行业关联条款进行差异分析比较;

(6) 形成油气行业标准指标比对结果报告。

3.2 油气行业标准指标自动比对的关键技术

3.2.1 油气行业标准内容指标比对层级

以油气行业实际需求来说标准比对分3个层次:了解油气行业宏观层次的差异、了解油气行业标准内容及工作思路差异、了解油气行业标准技术细节差异。针对这些需求开展比对的层次也有所区别。在大多数情况下体系对比可以满足宏观层面管理及发展布局差异的需求,内容比对能了解标准内容结构、文本结构差异,了解不同操作方式的工作思路;指标比对能进一步确定技术细节差异,明确不同技术的实现目标,了解指标取舍的判定过程。因此,在设计油气行业内容指标比对流程和给出比对结果时,也需要满足3个层次的需求。给比对人员足够的自由度,对结果的判定也需要油气行业基础专业背景并能提供可验证的证据。

3.2.2 油气行业标准比对结果的判定

专家比对采用的主要方式为两两比对,选定一个油气行业标准作为基准比对标准,选定待比对的内容和指标,检索其他标准中的相同或类似指标,将不同标准中的主要技术指标分别与基准标准中的指标进行对比。结果判定方法如下。

(1) 基准标准的要求低于其他比对标准时,该项指标的单项评价为“低于”;

(2) 基准标准的要求等同于其他比对标准时,该项指标的单项评价为“等同于”;

(3) 基准标准的要求严于其他比对标准时,该项指标的单项评价为“高于”;

(4) 若基准标准的该项指标在其他比对标准中未提及,该指标的单项判定结果为“自定义新指标”;

(5) 若基准标准缺少其他比对标准中的指标时,该指标的单项判定结果为“xx指标缺失”

结论判断的一般描述为“***标准的此项要求等同于***标准”“***标准的此项要求高于/低于***标准,具体为***”“***标准的此项要求与***标准存在差异为指标缺失/自定义新指标,具体为***”。

3.2.3 油气行业标准计算机辅助比对模型

(1) 提取关键字

计算机通过分词组件提取比对内容涉及的关键字。在油气行业标准文本结构中,字和标点符号共同构成一个段落,标点符号只是标准内容的分隔符,并没有特殊的意义。因此,标点符号不能够成为文件内容的关键字,同时停顿字在文件内容中也没有特殊的意义,因此停顿字也不能够成为文档内容的关键字。

通过分词组件完成以下功能:将比对内容分成单独的字;去掉标点符号;分词处理。

油气行业标准比对内容经过以上的过程处理后,就成为算法处理的词元。系统通过分词处理的方法将词元变成比对内容的关键字。油气行业标准内容指标比对数据模型的研究单元为指标化数据,因此标准内容中的指标化数据可以作为关键字提取的最小单元,有效降低词元的复杂性,提高了关键字的提取准确率。根据对汉语语法和统计学

规律的研究,制定了特有的关键字提取机制,即段落中关键字的字数应尽可能多、单个分词的根数尽可能少和总词数尽可能少。

(2) 权重计算

权重是比对模型中一个相对的概念,同时权重是一个可调的值,权重一般表示该元素对系统输出结果的影响力,影响力越大表示该元素权重越高,也表示该元素和系统处理结果越相关。标准内容指标比对的对象为标准内容中的指标数据,通过对标准内容的研究和分析,确定影响关键字在指标数据中权重的因素。

通过两个层次对标准文档进行研究,第一个层次为针对单个标准文档的研究,第二层次为针对整个标准文档库的研究,研究发现,每篇标准文档都包含不同的主题,而工作人员为了描述该主题都采用了大量的专业词汇,如果一篇文档中某些关键字的词频比较高,这些关键字可能是用于描述文档主题,那么这些关键字应该有较高的权重。同时文档中词频较高的关键字并不都是用于描述文档主题,有些关键字为生活中的常用词汇,文档在编写过程中会用到较多的常用词汇,通过统计学方法的分析,得出了相应的处理方法,即如果在一个数量比较大的文档集群中,包含某些关键字文档的数目越多,这些关键字越不重要,那么这些关键字的权值较低。

(3) 相关性判断

通过对指标数据之间相关性的判断,就能够实现指标数据之间的比对功能。将每段内容指标数据看作是由N个关键字(term)构成,每个关键字有一个权重,不同的关键字根据在指标数据中的权重来影响比对结果的相关性。通过处理将所有关键字的权重看作一个向量,将比对对象看作是由N个关键字(term)构成,也用向量表示。通过计算两个向量之间的夹角来判断搜索内容和文档之间的相关性,两个向量之间的夹角越小,相关性越大。可以采用余弦公式作为向量相关性的打分标准,余弦值越大,分数越高,相关性越大。通过上述过程的处理,系统就能够根据条件,自动实现指标数据的比对功能。

4 研究结论

当前,油气行业标准内容指标比对模型提供的是计算机辅助功能,深层次的判定和比对工作还需要专业人员进行。对于油气行业外文标准文献,加工工具在人工比对操作时支持中英文比对,但机器辅助目前仅适用于中文标准(或中文译文)。目前油气行业国内外标准比对需求旺盛,在开展比对前需要对国外标准文献进行转化(翻译),因此接入多语种自动翻译平台,研制支撑多语言的标准比对模型和工具,实现国内外标准自动翻译、内容快速检索、指标比对是下一步工作的重点。目前的方法对油气行业标准关键技术指标的自动判定和提取仅限于表格和关键内容中,但技术指标的自动归一化处理和组织还需要进一步研究,自动比对和自动结果判定方法也需进一步研究。因此需要对油气行业标准文献智能比对方法或技术开展更为深入的研究,突破油气行业标准比对关键或共性技术。同时在油气行业组织管理与制度机制建设、人才队伍建设、资金投入与基础设施建设、应用推广服务等方面加强支持力度。

(1) 加强油气行业标准比对人才队伍建设

油气行业标准文献内容指标比对是一项以应用和服务为目标的基础性工作,是标准化科研的一个重要组成部分。但是,在比对方法、比对内容以及专业技能和知识构成上有别于传统的标准化科研工作,需要进一步加强油气行业数据资源建设、更新维护、软件开发与市场营销相关的人才队伍建设,逐步完善标准比对人才队伍。加强油气行业科研人员和业务人员的相关意识和能力培养,建立一支由标准化研究和标准化服务机构共同组成的标准内容指标比对、翻译、计算机与服务营销队伍。

(2) 建立油气行业标准比对资金持续投入机制

油气行业标准内容指标比对数据库建设是一项长期持续性工作,需要长期投入。采用科研投入和市场投入相结合的方式,多方调动资金,加强油气行业标准内容指标库更新维护和开发应用工作。同时,重视资源建设与应用的投入产出评估。

(3) 加强油气行业标准比对方法的推广应用与服务工作

油气行业标准内容指标库建设不同于标准化科研工作, 有其自身的特点, 需要按照自身的规律科学建设才能够可持续发展。该项工作涉及多学

科、多领域的知识, 是最基础的资源建设, 目标是应用服务, 评价内容指标库建设的重要指标是推广应用。因此, 在油气行业标准内容指标库建设立项的初期就要十分强调推广应用, 明确其市场化思路和步骤以及所服务的对象和市场前景。

参考文献

- [1] 刘春卉. 水泥窑余热发电验收标准法规比较手册[M]. 北京: 中国质检出版社, 2018.
- [2] 甘克勤. 标准大数据实践[M]. 北京: 中国质检出版社, 2016.
- [3] 陈晓穗, 江绍华, 徐晨. LED台灯国内外标准比对[J]. 机械工业标准化与质量, 2019(07):32-36.
- [4] 朱晓春, 唐文洁, 斯庆图娅. 中蒙《铜精矿·二氧化硅含量的测定方法》标准指标比对研究[J]. 标准科学, 2019(05): 22-25.
- [5] 丁莉, 邹雄辉, 戴春蓓, 等. 轻型车国六与国五排放标准比对分析[J]. 小型内燃机与车辆技术, 2019, 48(03):54-59.
- [6] 马冬妮, 张淑芝, 叶国强. 浅议三峡工程水泥新旧标准比对方法[J]. 云南水力发电, 2002(03):81-82.
- [7] 蔺菲, 庄磊, 吴蕴洁. 实验室间电能计量标准比对的方法与数据处理[J]. 安徽电力, 2009, 26(03):43-48+84.
- [8] 付卉青, 刘霞. 消费品安全标准比对方法研究[J]. 中国标准化, 2018(09):49-52.
- [9] 计雄飞, 张宝林, 李抵非, 等. 标准文献内容挖掘与比对[J]. 标准科学, 2012(08):16-19.
- [10] 王昕, 王宏, 周育忠, 等. 标准指标比对的方法与实践[J]. 中国科技资源导刊, 2017, 49(04):83-92.
- [11] 周生龙, 张忠林. 古文献版本考究中的图像文字自动比对方法设计与应用研究[J]. 河南图书馆学刊, 2018, 38(09):72-74.
- [12] 吴建港, 国玉宝, 崔绍辉. 基于标准指标库的塑料注塑机产品标准指标比对分析[J]. 工程塑料应用, 2019, 47(04):107-111.