



数字中文建设背景下高质量民族语言文字语料库建设路径探析

● 淑琴

2025年,教育部、国家语言文字工作委员会、中央网信办联合印发《关于加强数字中文建设 推进语言文字信息化发展的意见》,标志着我国语言文字事业正式进入数字赋能、战略赋能、科技赋能的全新发展阶段。语言数据作为大语言模型研发的基础,其规模、质量和治理水平直接决定了模型的智能程度、应用范围和发展潜力。而高质量的民族语言文字语料资源,不仅是实现民族语言智能交互、精准翻译、智慧服务的前提,更是推动语言文字事业高质量发展的关键支撑。

内蒙古地处祖国北疆,作为我国向北开放的重要桥头堡及中蒙俄经济走廊的核心节点,不仅区位优势显著,更拥有独特的民族语言资源禀赋,为推进民族语言文字数字化建设奠定了坚实的基础,而建设高质量的民族语言文字语料库是紧扣铸牢中华民族共同体意识工作主线,助力“数字内蒙古”及现代化建设,推动中华优秀传统文化传承发展的关键举措。

一、高质量民族语言文字语料库建设的时代价值

数字中文建设将高质量民族语言文字语料库建设纳入国家语言资源战略布局,系统部署语言文字数字化、智能化、规范化建设工作,深度对接数字中国、教育强国、文化强国战略部署,统筹推进国家通用语言与民族语言数字化转型。这既是传承中华优秀传统文化、赓续民族文脉的基础工程,亦是赋能数字经济发展、驱动人工智能创新、服务国家对外开放大局的时代工程。

(一)服务数字中文建设大局,完善国家语言资源布局

数字中文建设旨在构建规范统一、互联互通、安全可控的国家数字语言生态,推动国家通用语言与民族语言共建共享、协同发展。民族语言文字语料库是国家关键语料库和国家语言文字大数据中心的重要组成部分,是国家语言信息化体系的基础支撑。高质量推进民族语言文字语料库建设,能够有效补齐民族语言数据短板,构建“通用语言为主、民族语言为辅、多语种协同”的国家语言资源新格局。

(二)赋能民族团结进步,有形有感铸牢中华民族共同体意识

语言文字是民族交往交流交融的重要纽

带,是凝聚民族情感、传承民族文化、汇聚民族共识的核心载体。高质量民族语言文字语料库建设,将系统归集民族团结史料、民族交融案例、惠民宣传资源,搭建数字化、规范化的传播教育平台。通过常态化普及国家通用语言文字,传播共有共享的中华文化符号,推动各民族互学互鉴、交融共生,持续增强各族群众中华文化认同感和民族向心力,以语言文字数字化建设筑牢祖国北疆民族团结的坚固屏障。

(三)赋能自贸区建设,优化沿边开放营商环境

中国(内蒙古)自由贸易试验区肩负着沿边制度创新、产业集聚、开放示范的重要职能,是我国向北开放的重要窗口。自贸区商事办理、跨境电商、涉外法务、外资服务等场景,对专业化、标准化、智能化多语种服务需求迫切。建设高质量民族语言文字语料库,能够为自贸区开展多语种政务服务、涉外法律咨询、跨境商事办理、国际化宣传推广提供坚实的数据支撑,推动政务服务智能化、涉外服务精细化、对外开放便捷化,从而促进沿边营商环境持续优化、开放能级持续提升。

二、民族语言文字语料库建设面临的现实问题

在民族语言文字语料库的建设过程中,常常会面临多维度的现实挑战,但大多体现在资源、标准和技术三个方面。从资源角度看,总量不足、零散分布、专业领域语料匮乏是现实问题;从标准角度看,采集标准不统一、技术标准缺失、管理标准空白是亟待解决的问题;而基础技术薄弱、智能化水平低、多模态融合不足则是技术层面的短板。

(一)标准体系滞后,规范化建设水平不足 标准化是语言信息化建设的基础和前提。目前,我区民族语言文字语料库建设尚未形成统一完备的地方标准体系,在语料采集、分类标注、元数据规范、质量评价、入库审核等方面缺乏统一准则。各高校、科研院所、文旅机构分头建设,存在编码不统一、标注不规范、格式不兼容等问题,造成资源难以互通共享、重复建设突出、资源内耗严重等问题。同时,适配大模型训练的专用语料标准、数据安全规范、版权使用规则、隐私脱敏机制等尚不健全,难以适配数字中文建设法治化、规范化发展的要

求。

(二)资源归集零散,优质专业语料供给短缺

我区民族语言文字语料资源整体呈现碎片化、分散化格局,尚未形成统筹归集、分类管控、动态更新的资源体系。现有语料多集中在日常通用文本、教材读物等基础领域,政务法治、跨境经贸、口岸物流、基层治理、应急服务等领域专业语料储备不足。资源形态以静态书面文本为主,语音、影像、交互式多模态语料储备滞后,难以适配大模型多维训练需求。此外,大量民间特色用语、边境交流语料、非遗口述史料分散在基层、民间和各类机构,挖掘转化不足,整体语料体量偏小、精品占比偏低,难以支撑民族语言大模型高质量迭代。

(三)技术支撑薄弱,数字化建设效能有限 相较于通用汉语智能技术成熟发展的态势,蒙古语等民族语言自然语言处理技术存在明显短板。例如,智能分词、语义识别、文本对齐、智能纠错等核心技术精度不足,自动化、智能化处理工具匮乏;语料整理、精细标注、内容校对多依赖人工完成,成本高、周期长、效率低。同时,缺乏一体化民族语言智能语料管理平台,数据存储、检索、清洗、去重、共享等功能不完善,海量语料难以实现集约化管理和高效利用,制约了语料库数字化、智能化升级步伐。

三、建设高质量民族语言文字语料库的实践路径

众所周知,数字中文建设为民族语言文字数字化转型、智能化升级提供了重大机遇。高质量民族语言文字语料库建设,是夯实民族语言智能发展根基,赋能内蒙古高质量发展的基础性、战略性工程。破解民族语言文字语料库建设中的问题,需要政府、科研机构、高校和企业等各方共同努力,通过加强顶层设计、加大资源投入、完善标准体系、提升技术水平、加强人才培养等措施加以解决。

(一)健全标准规范体系,夯实法治化规范化建设根基

严格对标国家数字中文建设和语言信息化标准化部署,结合我区的实际情况及发展需求,需统筹教育、网信、民委、文旅、司法及科研院所等力量,构建适配蒙古语言文字、符合地方发展需求的建设体系。例如,需要统

一语料采集范围、分类规则、分词标注、语义规范、元数据格式、入库标准和质量评价体系,推动民族语言标准与国家通用语言标准、人工智能行业标准无缝衔接;还要同步完善数据安全、版权管理、隐私脱敏、内容审核等制度机制,规范语料采集、存储、流转、共享、应用全流程监管,严守数据安全、伦理规范和法治底线,实现语料库建设全程有据可依、有规可循、合法合规。

(二)统筹全域资源归集,构建分层分类特色语料体系

应坚持“政府主导、统筹规划、突出特色、共建共享”原则,打破部门壁垒与行业分割,建立全区统一的民族语言资源归集机制,盘活存量、挖掘增量、抢救濒危资源。打造战略专属语料,聚焦民族团结、向北开放、自贸区建设,构建交融叙事、口岸经贸、跨境交流、商事法治、基层治理等专项语料库。抢救特色文化语料,系统整理蒙古古籍、英雄史诗、民间故事、民俗谚语、非遗口述史料,实现民族文化资源数字化留存。

(三)强化智能技术赋能,提升数字化建设质效

依托国家语言文字大数据中心建设布局,对接国家语言资源服务平台,搭建内蒙古民族语言文字智能语料管理共享平台,集智能采集、批量去重、精准纠错、智能标注、分类检索、安全共享于一体。推动语料平台与智慧教育、政务服务、跨境服务、自贸区服务平台互联互通,打通数据流转壁垒,实现优质资源高效调用、全域共享、深度赋能,全面提升民族语言数字化、智能化水平。

(四)构建多元协同格局,激活共建共享发展活力

秉持数字中文共建共享理念,构建“政府统筹、科研主攻、市场赋能、社会参与”的立体化建设体系。政府强化顶层设计、政策保障、资金投入和统筹协调;科研院所聚焦标准研制、资源整理、技术攻关和理论创新;广泛征集民间特色语言资源,汇聚全民共建合力。主动融入全国数字中文建设联盟,深化跨区域交流协作,借鉴先进建设经验,推动民族语言资源互联互通、共建共享,持续释放数字化建设综合效能。

(作者单位:呼和浩特民族学院智能科学学院)

新时代国有施工企业党建与生产经营融合路径探析

● 李雅楠

新时代推动党建与生产经营深度融合是国有施工企业高质量发展的必然要求。本文以内蒙古地区企业为样本,剖析问题成因,从六个维度构建融合路径,为边疆民族地区同类企业提供实践参考。

对国有施工企业而言,如何推动党建深度融合融入治理体系、嵌入生产经营全过程,一直是其坚持党的领导、激活治理效能、推动高质量发展必须破解的重大命题。

以我区国有施工企业为样本,当前融合实践仍存在突出瓶颈。一是融合型党建思维缺位,部分管理者“重业务、轻党建”;二是量化考评体系不健全,考核指标精细化、差异化不足,刚性约束不强;三是融合深度不足,多停留在“物理叠加”层面,在“党建+科技创新、数智化转型”等领域探索有待深入,引领赋能作用未充分释放。究其根源,集中在四个层面。思想认识层面上,对“围绕发展抓党建、抓好党建促发展”辩证关系理解不深,从“事务型党务”向“价值型党建”转变不彻底;机制建设层面上,制度供给不足、闭环管控不精细,激励容错机制不完善;责任落实层面上,压力传导衰减,末梢督导不严;方法创新层面上,方式手段与行业特点、职工需求适配度不高,品牌化意识不强、新媒体运用滞后。

此外,国有施工企业还具有项目分散、流动性强、野外作业为主、社会关联度高等特征,因此,必须立足行业规律与区域实际,以系统思维构建党建与生产经营深度融合新格局。

一、强化思想铸魂,夯实融合发展的政治根基

坚持把党的政治建设摆在首位,以思想高度统一凝聚改革发展合力。严格执行“第一议题”制度,深入开展正确政绩观专题教育,将正确政绩观要求纳入党委理论学习中心组学习、党支部“三会一课”以及党员日常教育的核心学习内容,推动党的创新理论学习与企业战略谋划、生产经营、党建研究深度融合。

常态化开展警示教育,对照“两张皮”、形式主义等典型问题深入反思剖析,破除思想认知误区,牢固树立党建业务融合思维。健全覆盖学习教育全过程的督导机制,将落实成效纳入年度党建工作考核,推动全体干部职工思想统一到“十五五”规划与企业改革发展上来,凝聚起干事创业的强大合力。

二、聚焦民族工作,推动党建与民族团结双融互促

立足边疆民族地区定位,把铸牢中华民族共同体意识作为主线贯穿融合全过程。将主题教育纳入党委理论学习中心组、“三会一课”、干部培训必修课,纳入党建考核和干部考评体系。对内广纳各民族优秀人才,促进交往交流交融;对外深化企地共建,尊重民族习俗、解决群众急难愁盼,实现企业发展与民族团结双向促进。坚持以宣传教育为重要抓手,高质量推进各项工作落地,让铸牢中华民族共同体意识真正融入干部群众思想深处、落到日常工作实处。

三、深化多维融合,打造“党建+”特色载体

坚持把生产经营主战场作为党建主阵地,推动全方位嵌入、全链条融合、全要素赋能。在政治引领上,将政策宣讲融入施工现场,与安全培训、技术交流相结合,有效凝聚参建各方的思想共识。在组织体系上,不折不扣落实“四同步、四对接”,实现党对项目的全覆盖,做到项目建到哪里、党建工作覆盖到哪里;同时深化跨单位联建模式,打造联动共建的党建格局。在人才培养上,实施“双培工程”,把骨干培养成党员、把党员培养成骨干;在载体创新层面,聚焦中心工作,打造特色党建品牌,重点推出“党建+安全生产”“党建+创新攻坚”“党建+绿色施工”“党建+数智化转型”四类融合载体,真正推动党建工作融入生产、嵌入经营、落到一线。

四、加强品牌塑造,构建全域融合工作矩阵

深入推进“一支部一特色、一支部一品牌”创建工作,指导各基层党支部结合业务特点提炼品牌内涵、设计特色载体,定期组织跨支部交流观摩,互学互鉴提升创建质量。遵循“培育一成熟一推广”的梯度建设思路,对创建成熟、成效突出的品牌进行重点培育,选树一批可复制、可推广的标杆样板,最终形成“党委有主导品牌、支部有特色品牌、全员共同参与”的全链条品牌建设格局,推动党建品牌持续赋能企业高质量发展。

五、夯实基层基础,健全融合常态长效机制

以制度化、规范化建设保障党建业务融

合常态长效。严格落实周提醒、月调度、季督导、年总结的全周期管控机制,细化形成任务、责任、时限“三张清单”,强化日常调度督导,精准破解融合过程中的堵点难点。健全党务干部专业化培养机制,推行跟班轮训制度,全方位提升党务干部综合能力,推动基层党建工作专业化水平稳步提升。

探索创新党建共建路径,推行大支部联建带动小支部、机关结对帮扶项目一线的共建模式,推动优质党建资源下沉共享、业务发展协同联动,夯实全域协同、上下联动的党建工作格局。

六、抓实宣传赋能,塑造企业良好形象

以宣传讲好融合故事、凝聚奋进力量。完善宣传管理制度,明确分工、流程、考核,构建党委领导、党群牵头、部门协同、全员参与格局。精心策划主题、专题宣传活动,打造“党员的一天”“项目攻坚现场”“匠心筑路人”等系列宣传专栏,讲好党员先进事迹、项目一线奋斗故事,凝聚奋进正能量。同步强化外宣统筹与舆情管控:对外对接主流媒体策划重点选题,扩大企业品牌的社会影响力;对内健全舆情预警处置机制,第一时间回应公众关切,主动防范化解舆论风险,牢牢守住企业品牌声誉底线。

推动党建与生产经营深度融合是国企高质量发展的必然选择。内蒙古国有施工企业需多措并举,实现党建与经营深度融合,以党建优势赋能企业及地区经济高质量发展。

(作者单位:内蒙古路桥集团有限责任公司)