

# 《环境文本数据加工处理技术规范》

（征求意见稿）

编制说明

《环境文本数据加工处理技术规范》编制组

二〇二四年十月

# 目 录

1 任务来源.....	3
2 标准制定必要性.....	3
3 主要工作过程.....	3
4 国内外相关标准研究.....	6
5 同类工程现状调研.....	8
6 主要技术内容及说明.....	9
7 标准实施的环境效益与经济技术分析.....	9
8 标准实施建议.....	10
9 应用案例.....	11

# 《环境文本数据加工处理技术规范》编制说明

## 1 任务来源

2024年，生态环境部环境规划院承担了国家重点研发计划“**基于人工智能与区块链技术的生态环境新型治理体系研究与示范应用（编号：2021YFC1809000）项目**”，其中提出了制定《环境文本数据加工处理技术规范》团体标准的任务。生态环境部环境规划院承担该标准的编制工作。参编单位有北京市生态环境监测中心。

## 2 标准制定必要性、编制依据、编制原则

本团体标准《环境文本数据加工处理技术规范》的制定具有重要的现实意义和迫切需求。随着环保工作的深入发展，环境文本数据的规模日益庞大，涵盖政策法规、研究报告、监测报告等多种类型，这些数据对于环境治理、政策制定、科学研究等具有重要价值。然而，由于缺乏统一的数据加工处理规范和技术指导，导致数据质量参差不齐、信息提取效率低下等问题。因此，制定本标准旨在规范环境文本数据的加工处理流程，提高数据质量与可用性，促进环保领域的信息化建设。

本标准的编制依据主要包括但不限于 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》等相关国家标准，以及国内外先进的数据处理技术和经验。同时参考了国际标准化组织 ISO/IEC JTC 1/SC 36 和 ISO/IEC JTC 1/SC 32 的相关标准，确保本标准与国际接轨。

编制本标准遵循以下基本原则：一是科学性原则，确保技术内容符合科学技术成果和发展趋势；二是实用性原则，紧密结合环境保护的实际需求，保证标准的可操作性和实用性；三是兼容性原则，考虑到与其他相关标准的协调一致；四是安全性原则，确保数据加工处理过程中的信息安全和个人隐私保护；五是可持续性原则，注重环境保护和可持续发展的长远目标。通过上述原则的指导，本标准旨在为环境文本数据的加工处理提供一套系统化、规范化的方法论框架。

## 3 主要工作过程

### 1) 成立编制组时间。

本团体标准《环境文本数据加工处理技术规范》的编制工作得到了广泛的关注和支持。为了确保标准的科学性、实用性和先进性，我们于2022年9月正式成立了标准编制组。编制组成员由来自环保智能化系统平台建设、大数据分析、人工智能技术应用以及政策文本分析等多个领域的专家组成，涵盖了政策制定者、科研人员、企业代表和第三方服务机构的专业人士，共同致力于本标准的编制工作。

### 2) 经过哪些工作，什么时间完成了开题报告。

本团体标准《环境文本数据加工处理技术规范》的编制工作始于2022年9月正式成立了标准编制组。编制组成员由来自环保智能化系统平台建设、大数据分析、人工智能技术应用以及政策文本分析等多个领域的专家组成。编制组成立后，立即展开了紧锣密鼓的工作安排，明确了标准的编制目标、工作内容与程序，并制定了详细的时间表和路线图。

编制组成员积极参与各项讨论会议，进行了多次的技术交流和意见征求，确保标准内容既符合国际发展趋势又适应我国环保工作的实际情况。在此基础上，编制组于2023年8月初完成了开题报告的编写工作，报告详细阐述了标准的背景意义、编制目的、主要内容框架、预期成果以及具体的编制计划。开题报告的完成标志着本团体标准编制工作进入了实质性的

实施阶段。

开题报告的编制过程中，编制组充分考虑了国内外相关标准的发展情况，特别是在环境文本数据处理方面的最新研究成果和技术进展，并结合我国环保工作的实际需求，确保了标准内容的科学性、实用性和先进性。开题报告的完成为后续的标准编制工作奠定了坚实的基础。

3) 召开开题会时间，会上提出了哪些建议。

本团体标准《环境文本数据加工处理技术规范》的编制工作于2024年4月正式启动，并于同日召开了开题会。会议上，编制组成员及相关专家围绕标准的编制目标、技术路线、工作内容与程序等核心议题展开了深入讨论，提出了以下建议：

**明确标准定位：**强调本标准应立足于解决环境文本数据加工处理中的关键技术问题，为环保领域提供统一的技术指导和规范。

**技术选型：**鉴于当前环保工作中大量使用的文本数据种类繁多，建议采用先进的文本处理技术，特别是基于大语言模型的文本嵌入方法，以提高数据加工处理的效率和质量。

**适用范围界定：**明确本标准适用于各类环境文本数据的加工处理，包括政策文本、科研报告、监测数据、新闻报道等。

**跨学科合作：**鉴于环保工作涉及多个学科领域，建议加强与信息技术、环境科学、法律政策等领域的交流合作，确保标准内容的全面性和科学性。

**持续更新机制：**考虑到技术的快速发展，建议建立标准的定期审查和更新机制，确保标准内容能够及时反映行业发展的最新动态。

4) 根据开题会建议并做了哪些工作，什么时间形成征求意见稿。

根据开题会上提出的建议，本团体标准《环境文本数据加工处理技术规范》的编制组积极开展了一系列工作。编制组对开题会上提出的各项建议进行了细致梳理，并将其逐一纳入后续的工作计划之中，确保每一项建议都能得到有效落实和完善。首先，编制组明确了标准的技术定位和适用范围，确保本标准能够有效指导环境文本数据的加工处理过程。在确定技术定位的过程中，编制组深入分析了当前环境文本数据处理的实际需求和的发展趋势，明确了标准的核心内容应涵盖数据采集、预处理、分析、存储以及最终的应用等多个环节。为了使标准更具实用性和针对性，编制组还特别考虑了不同应用场景下的特殊要求，如环境监测数据的实时处理、环境事件的舆情分析等，确保本标准能够适应多样化的数据处理需求。此外，编制组还对标准的适用范围进行了详细界定，明确指出标准适用于各类环境监测机构、科研单位以及相关企业的数据处理活动，旨在通过统一的技术规范，提高整个行业的数据处理水平和效率。

其次，编制组对文本处理技术进行了深入研究，特别是基于大语言模型的文本嵌入方法，以提高数据加工处理的效率和质量。在这个过程中，编制组密切关注了最新的技术动态，包括自然语言处理领域的前沿研究成果，并将其合理融入标准之中。通过对大语言模型的研究，编制组发现这类模型在文本语义理解和上下文感知方面具有明显优势，能够显著提升文本数据的处理质量和准确性。因此，在标准中明确提出推荐使用此类模型来进行文本的预处理工作，如关键词提取、文本分类、情感分析等。同时，编制组还注意到，大语言模型的应用也需要考虑计算资源的消耗和模型训练所需的数据量等因素，因此在标准中给出了相应的指导原则，帮助用户根据自身条件选择合适的模型配置。此外，编制组还探讨了如何结合其他先进技术（如云计算、区块链等），进一步优化文本数据处理流程，提高数据处理的整体效能。

此外，编制组特别关注了数据安全和个人隐私保护问题，制定了严格的数据访问权限和

使用流程，确保数据的安全性。在标准编制过程中，编制组认识到随着数据量的增长和个人信息的数字化，数据安全和个人隐私保护已成为不容忽视的重要议题。为此，编制组专门设立了数据安全小组，负责研究制定符合国家标准和行业惯例的数据保护措施。编制组在标准中明确规定了数据分类分级的原则，确保敏感数据得到妥善处理；同时，制定了详细的用户权限管理体系，限制数据访问范围，防止未经授权的数据访问行为发生。此外，为了保障个人隐私权益，编制组还在标准中加入了关于数据匿名化处理的要求，确保在不影响数据使用价值的前提下，最大限度地减少个人信息泄露的风险。编制组还强调了数据生命周期管理的重要性，要求在数据的收集、存储、传输、使用以及销毁等各个环节都要遵循严格的规范，确保整个数据处理过程的安全可控。

基于上述工作，编制组于2024年8月中旬完成了征求意见稿的编制，并开始广泛征求业内专家和社会各界的意见。征求意见稿的形成标志着本团体标准进入了一个新的阶段，为进一步完善标准内容提供了宝贵的机会。编制组希望通过此次公开征求意见，能够汇集更多专业意见和实践经验，使标准内容更加全面、科学和实用。在未来的工作中，编制组将继续保持开放的态度，积极吸收各方反馈，不断完善标准内容，争取早日发布正式版本，为环境文本数据加工处理领域提供有力的技术支撑。

#### 5) 公开征求意见的时间。

本标准拟于2024年10月进行公开征求意见。

#### 6) 标准验证情况。

本标准于2024年4月进行标准验证。标准验证是确保《环境文本数据加工处理技术规范》科学性、可行性以及有效性的关键步骤。在2024年4月进行的标准验证工作中，编制组采取了多种方法来检验标准的各项规定是否能够达到预期的目标。首先，编制组选择了几个具有代表性的环境监测机构作为试点单位，通过这些单位的实际运行情况来验证标准的有效性。在试点单位的选择上，编制组充分考虑了地域分布、规模大小以及业务类型的多样性，以确保验证结果具有广泛的代表性。试点单位按照标准的要求对环境文本数据进行了加工处理，并记录下了整个过程中遇到的问题以及标准的实际效果。此外，编制组还设置了详细的评估指标体系，包括数据处理效率、数据准确性、系统稳定性等多个维度，以此来量化标准实施前后的变化情况。通过这种方式，编制组不仅能够了解到标准在实际应用中的表现，还能及时发现存在的不足之处，为后续的修订提供依据。

其次，编制组组织了多次专家评审会议，邀请了来自信息技术、环境科学、法学等多个领域的专家学者，对标准的各项技术指标进行了全面的评审。专家评审团成员不仅具备深厚的理论知识，还拥有丰富的实践经验，他们的意见对于标准的完善具有重要价值。在评审过程中，专家们围绕标准的技术细节、实用性以及与其他相关标准的兼容性等方面展开了深入讨论。此外，编制组还请专家们就标准实施过程中可能出现的问题提出了预防措施和解决方案，确保标准在推广过程中能够顺利执行。评审会议还讨论了标准的国际化问题，考虑到环境保护是一个全球性的议题，标准的制定需要兼顾国际通行做法与本国实际情况，以促进国际间的交流与合作。专家们一致认为，标准应该具有一定的前瞻性，能够适应未来一段时间内技术发展和市场需求的变化。

最后，编制组还通过问卷调查、线上研讨会等形式，广泛征求了社会各界的意见。调查对象不仅包括了环境监测机构的专业技术人员，还包括了普通民众、非政府组织代表等利益相关方。这种多元化的意见征集方式有助于编制组从不同角度了解标准的影响，确保标准既符合专业需求，又能获得社会公众的认可和支 持。在线上研讨会期间，参与者有机会直接与

编制组成员交流，提出自己的看法和建议。编制组认真听取了各方意见，并在后续的标准修订中给予了充分考虑。此外，编制组还利用现代信息技术手段，如云计算和大数据分析，对收集到的反馈信息进行了整理和分析，以便更准确地把握社会大众的需求和期望。通过这一系列验证活动，编制组不仅验证了标准的技术合理性，还增强了标准的社会接受度，为标准的最终定稿奠定了坚实的基础。

#### 7) 根据专家意见形成报批稿的过程。

根据专家意见形成报批稿的过程是一项系统而细致的工作，旨在确保《环境文本数据加工处理技术规范》的标准内容既科学合理又具有可操作性。首先，在收到专家们的评审意见后，编制组立即组织召开了专题工作会议，对每一条意见进行了认真的讨论和分析。专家们的意见涵盖了技术细节、术语定义、应用案例等多个方面，既有宏观层面的指导建议，也有微观层面的具体修改意见。编制组成员根据各自的专业背景，对这些意见进行了分类汇总，并初步拟定了修改方案。在这一过程中，编制组特别注意保留原始标准的核心理念和技术框架，同时吸收专家们提出的有益建议，以期在保持标准连贯性的前提下，进一步提升其完整性和适用性。此外，编制组还设立了一个专门的审核小组，负责对照专家意见逐条核对标准文本中的修改内容，确保每一个修改点都能够准确无误地体现在新版标准中。

接下来，编制组对标准文本进行了详尽的修订工作。根据专家的意见，编制组对标准中的一些技术术语进行了重新定义，使之更加准确和易于理解。例如，针对“文本嵌入”这一概念，编制组不仅明确了其在环境数据处理中的具体含义，还提供了详细的实现方法和技术参数，便于读者根据实际情况灵活选择。此外，编制组还补充了若干典型应用场景下的示例说明，以帮助使用者更好地理解标准的应用范围和实践方法。在数据安全和个人隐私保护章节，编制组采纳了专家提出的增加数据加密传输和访问控制策略的建议，强化了标准中有关数据保护的规定。同时，编制组还特别注意到了专家关于标准实施过程中可能遇到的问题预警，增加了相应的风险提示和应对措施，以增强标准的实用性和指导意义。

最后，编制组对修订后的标准文本进行了反复校对和格式调整，形成了正式的报批稿。在这一阶段，编制组再次邀请部分专家对报批稿进行了审阅，确保所有修改意见都得到了妥善处理，并且标准文本的语言表达清晰准确、逻辑严密。此外，编制组还组织了内部的模拟评审，模拟标准审批机关的审查流程，对报批稿进行了全面检查，以确保其符合相关法律法规的要求和技术规范的标准。在完成上述准备工作后，编制组正式向标准审批机关提交了报批稿及相关支持材料。同时，编制组还准备了一份详细的修改说明，详细列出了每一次修改的具体内容及其依据，以便审批机关在审核过程中能够快速了解标准修订的全貌。通过这一系列严谨的工作流程，编制组最终形成了高质量的报批稿，为标准的最终批准奠定了坚实的基础。

## 4 国内外相关标准研究

在编制本团体标准《环境文本数据加工处理技术规范》的过程中，编制组对国内外相关标准进行了深入研究。国际上，ISO/IEC JTC 1/SC 36 和 ISO/IEC JTC 1/SC 32 等技术委员会制定了一系列与文本数据处理相关的标准，如 ISO/IEC 2382-36《信息技术——词汇——第 36 部分：学习、教育和培训》等，这些标准为文本数据的处理提供了重要的参考。同时，国际标准化组织 ISO 发布的 ISO 3166-1《国家代码》等标准也为文本数据中的地理信息处理提供了标准化的框架。在国内标准方面，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》等国家标准为团体标准的编制提供了基本的指导原则。此外，GB/T 38660—2020《信息技术 大数据 数据质量》等相关标准也对数据质量控制提供了

具体的要求和指导。编制组在制定本团体标准时，充分借鉴了上述国内外相关标准的核心理念和技术内容，并结合环保领域的需求和特点进行了创新性开发。本标准不仅吸收了国际先进的数据处理技术，还融入了环保智能化系统建设的实践经验，旨在为环境文本数据的加工处理提供一套系统化、规范化的方法论框架。

在国内方面，随着环保意识的增强和技术的发展，环境监测与治理工作越来越依赖于大数据的支持。在环境文本数据加工处理上，国内学者和机构已展开了广泛的研究。一方面，随着物联网、云计算等技术的应用，环境监测数据的获取更加实时和全面，这些数据经过自然语言处理（NLP）技术，能够快速转化为有价值的决策依据。例如，通过分析社交媒体上的公众意见，可以及时发现环境污染事件并采取措施。另一方面，针对环境文本数据的特点，如信息碎片化、语义复杂等问题，国内科研人员提出了多种数据清洗与整合的方法，以提高数据的质量和可用性。此外，考虑到数据安全与隐私保护的重要性，相关机构还探索了基于区块链技术的数据共享机制，确保在数据流通中保护个人隐私与企业商业秘密。

与此同时，国内对于环境文本数据加工处理的标准建设也日益重视。政府部门和行业协会正在积极推动相关标准的制定和完善，以填补现有标准体系中的空白。近年来，一系列涉及环境保护领域的信息化建设指南相继出台，涵盖了数据采集、存储、传输等多个环节的技术要求。这些努力不仅有助于提升环境管理效率，也为实现可持续发展目标提供了坚实的技术基础。然而，在实际操作中，如何平衡数据开放与信息安全之间的关系仍然是一个挑战，需要跨学科的合作来共同解决。

在国际视野下，环境文本数据加工处理同样是全球关注的重点领域之一。发达国家在这一领域的研究与实践相对成熟，尤其是在利用人工智能算法进行复杂环境问题分析方面取得了显著进展。例如，通过机器学习模型对大量历史环境数据进行训练，可以预测未来可能出现的污染趋势，并为政策制定提供科学依据。此外，跨国合作项目也在不断推进，如欧盟支持的多个环保项目就强调了数据共享与标准化的重要性，力求通过统一的数据处理流程促进不同国家和地区间的信息交流与合作。此类国际合作不仅有助于提升各国环保工作的协同效应，也为发展中国家提供了宝贵的经验借鉴。

此外，国际社会在推动环境文本数据加工处理技术规范的全球化进程中也发挥了重要作用。诸如 ISO 等国际组织已经发布了一系列与环境保护相关的标准文档，其中包括对数据质量控制的具体要求。这些标准不仅为企业和个人提供了统一的操作指南，也为评估和比较不同地区环境状况提供了可靠基准。值得注意的是，随着全球化进程加快，跨国公司越来越多地参与到全球供应链管理中，因此建立一套被广泛认可的数据处理规范变得尤为重要。这不仅有助于降低因数据不兼容而导致的成本，也有利于构建更加透明、公平的国际贸易环境。

在非结构化文本数据分析方面，国内外的研究均展现了显著的进展与应用。在国内，随着互联网技术的发展，社交媒体、新闻报道等渠道产生的海量文本数据成为了重要的研究对象。研究人员利用先进的自然语言处理技术，如中文分词、情感分析和主题模型，从这些非结构化数据中挖掘出公众对于环境问题的态度和情绪反应，为政府决策提供了有力支持。同时，借助大数据处理框架如 Hadoop、Spark，以及深度学习模型如 CNN、RNN 等，实现了对大规模文本数据集的有效处理，提高了文本分析的准确性和效率。然而，数据标注成本高、模型泛化能力不足等问题依然存在，需要进一步研究解决。而在国际上，预训练语言模型（如 BERT、GPT 系列）的应用极大提升了文本数据的理解能力，特别是在环境科学中的自动摘要和关键信息提取等方面表现出色。此外，国际研究还特别关注文本数据分析的伦理和社会影响，强调算法的透明度和公平性，避免在情感分析或文本生成过程中产生偏见，并确保在处理包含敏感信息的文本数据时保护用户隐私。国际组织和行业联盟提出的指导原则和最佳实践，为全球范围内的非结构化文本数据分析提供了伦理和技术标准。结合国内外的研究进展，可以看出非结构化文本数据分析在技术革新与应用拓展方面均取得了实质性突破。国内研究

侧重于利用先进技术挖掘文本数据的社会价值，特别是在环境保护领域的应用，而国际研究则更加强调技术的伦理性和跨模态学习能力，以确保技术发展的可持续性和社会包容性。未来的研究将致力于优化现有的分析工具，并探索更具创新性的解决方案，以应对非结构化文本数据处理中面临的各种挑战，促进其在全球范围内更广泛的应用。

## 5 同类工程现状调研

在编制本团体标准《环境文本数据加工处理技术规范》的过程中，编制组对同类工程现状进行了深入调研。当前，随着环保工作的日益精细化，环境文本数据的加工处理已成为环保领域信息化建设的重要组成部分。国内外已有不少相关的实践案例和技术成果，为本标准的制定提供了丰富的参考素材。例如，我国部分地方政府和环保机构已经开展了环境文本数据的收集、加工和分析工作。以深圳市为例，该市启动了减污降碳协同综合模型与智慧决策应用项目，该项目通过整合环境政策文本、科研报告、监测数据等多源异构数据，利用大数据分析和人工智能技术，实现了对环境数据的有效整合与利用。该项目不仅提升了环境监测数据的处理速度，还通过智能分析技术，为环保决策提供了科学依据。此外，还有多个省市在环保信息化建设中引入了类似的技术手段，通过构建环保大数据平台，实现了对环境信息的实时监控与预警，大大提高了环保工作的效率和响应速度。这些成功的实践案例表明，环境文本数据的加工处理技术在实际应用中具有广泛的前景和潜力。

与此同时，国际上也涌现出一批先进的环保数据处理技术与应用。各国都在积极探索环保领域的数据处理技术，通过构建环保智能化系统来提高环保工作的效率和精准度。例如，欧盟的环境信息系统（EIONET）就是一个典型的例子。EIONET 不仅实现了环境数据的共享与分析，还通过建立统一的数据交换平台，促进了成员国之间的信息互通与协作。此外，美国环保署（EPA）也推出了多项环保数据管理系统，如 AirToxics Portal 等，这些系统通过先进的数据处理技术，实现了空气质量数据的实时监测与分析，为公众健康提供了有力保障。日本则在环境监测网络建设和数据可视化方面积累了丰富的经验，通过引入物联网技术，构建了覆盖全国的环境监测网络，能够实时收集并分析各地的环境数据，为环保政策制定提供了强有力的数据支持。通过对这些国际案例的学习与借鉴，我国在环保数据处理技术方面也取得了显著进步。

通过对这些工程的现状调研，编制组发现虽然目前已有不少成功的实践案例，但在数据标准化、数据清洗、数据加工处理等方面仍存在一定的挑战。首先，在数据标准化方面，由于环境数据来源多样，格式不一，缺乏统一的数据标准导致了数据整合难度大、互操作性差等问题。其次，在数据清洗环节，由于环境数据本身具有复杂性，存在大量缺失值、异常值等情况，如何高效准确地进行数据清洗，是当前面临的一大难题。再者，在数据加工处理过程中，如何利用先进的技术手段，如自然语言处理、机器学习等，实现对环境文本数据的深度挖掘与分析，也是一个需要深入研究的方向。这些问题的存在，制约了环境数据的有效利用，影响了环保工作的整体效能。

本团体标准旨在解决这些问题，提供一套统一的技术规范，以促进环保领域数据处理的标准化和规范化，从而更好地服务于环保工作的实际需求。标准将从数据标准化入手，制定统一的数据格式和编码规则，确保数据的一致性和互操作性；在数据清洗方面，提出科学合理的数据清洗策略和技术方案，提高数据质量；在数据加工处理方面，结合最新的自然语言处理技术和机器学习算法，为环境文本数据的深度分析提供技术指南。通过本标准的制定与实施，希望能够推动环保数据处理技术的标准化进程，为环保工作的科学化、信息化提供坚实的技术支撑，最终实现环境数据的最大化利用，促进环保事业的健康发展。

## 6 主要技术内容及说明

**适用范围：**本标准明确了其适用范围，即适用于各类环境文本数据的加工处理，包括但不限于政策文本、科研报告、监测数据、新闻报道等。这些数据类型涵盖了环境管理与科学研究的多个方面，从政府发布的政策文件到科研机构的研究报告，再到媒体的新闻报道，以及各类环境监测站点采集的数据信息。标准的广泛适用性意味着它可以作为一个统一的框架，指导不同来源和形式的环境文本数据进行标准化的加工处理，确保数据的一致性和可比性。无论是政府部门制定的环境法规解读，还是科研人员对环境变化趋势的分析，或是新闻媒体对环境污染事件的报道，都可以参照本标准进行数据的标准化处理，从而提高数据使用的效率和准确性。

**规范性引用文件：**本标准编制过程中引用了多项国家及行业标准，这些规范性文件为本标准的技术内容提供了坚实的科学基础和权威性支持。引用的文件包括但不限于《信息分类与编码的基本原则与方法》、《环境信息系统通用技术要求》等行业标准，以及相关国际标准。通过规范性引用文件的确立，本标准的技术要求与国内外现行标准体系保持了一致，确保了技术内容的科学性和权威性，同时也为使用者提供了明确的参考依据，便于在实际操作中遵循统一的技术规范。

**术语和定义：**标准中对关键术语进行了明确定义，以确保各方对术语的理解一致。例如，“环境文本数据”是指与环境相关的、以文本形式存在的信息集合；“文本嵌入”是指将文本信息转化为数字向量的过程，以便于计算机处理和分析；“数据清洗”则是指去除数据中的错误、不一致和冗余信息，提高数据质量。通过对术语的精确定义，本标准消除了因术语理解差异而可能导致的混淆，为后续的技术操作和应用提供了清晰的指引，确保各方在使用标准时能够达成共识。

**工作内容与程序：**本标准详细规定了环境文本数据加工处理的各个环节，包括数据收集、预处理、分析、加工与处理等步骤的操作流程。在数据收集阶段，明确了数据来源的合法性与可靠性要求；在预处理阶段，则对数据进行了初步的清洗和格式化处理；分析阶段采用了先进的自然语言处理技术，对数据进行了深入挖掘和分析；加工与处理阶段则根据分析结果对数据进行了进一步的加工，使其更加符合特定的应用需求。通过这一系列标准化的操作流程，本标准确保了数据处理的系统性和科学性，为环境文本数据的高效利用奠定了基础。

**数据分类：**本标准界定了环境文本数据的主要类别，并对其特性进行了描述，为后续的数据处理提供了分类依据。数据分类涵盖了政策文本、科研报告、监测数据、新闻报道等多个方面，并对每一类数据的特点进行了详细说明。例如，政策文本通常包含明确的指令性和指导性信息，科研报告则更加注重数据的准确性和结论的可靠性。通过对数据的分类描述，本标准为使用者提供了清晰的数据管理指南，便于根据不同类型的数​​据采取相应的处理策略，从而提高数据处理的针对性和有效性。

**数据清洗：**本标准制定了数据清洗的具体策略，包括清洗前评估、清洗策略、清洗方法及清洗后验证等，确保数据质量。在清洗前评估阶段，对数据进行了全面的检查，识别出需要清理的数据元素；清洗策略则根据数据的不同类型和问题制定了相应的清洗规则；清洗方法采用了自动化与手动相结合的方式，确保清洗过程的高效性和准确性；清洗后验证则是对清洗结果进行复查，确保数据清洗达到了预期效果。通过这一整套的数据清洗流程，本标准有效提高了环境文本数据的质量，为后续的数据分析和利用提供了可靠的数据基础。

## 7 标准实施的环境效益与经济技术分析

本团体标准《环境文本数据加工处理技术规范》的实施将带来显著的环境效益与经济技术效益。通过规范化的数据加工处理流程，能够快速准确地提取环境文本中的关键信息，为

环保政策制定和执行提供有力支持，提升决策效率。环境问题的复杂性和多变性要求决策者能够及时获取准确的信息，以便作出科学合理的决策。标准的实施意味着环保机构可以通过自动化工具高效地从海量文本数据中筛选出有价值的信息，例如，通过对政策文本、科研报告和监测数据的深度分析，可以识别出当前环境治理中的薄弱环节，并据此调整政策方向。此外，标准还为不同部门之间的信息共享提供了统一的平台，促进了信息的透明度和可访问性，使得决策过程更加民主化和科学化。例如，在应对突发环境事件时，相关部门可以迅速调用相关数据，分析事件影响范围及程度，制定紧急应对措施，提高应急响应速度和效果。通过这种方式，标准不仅优化了决策流程，还增强了政府与公众之间的信任，提升了政府形象和社会治理水平。

标准化的数据处理方法能够减少数据重复收集与处理的成本，避免资源浪费，实现资源的有效利用。在传统的数据处理模式下，由于缺乏统一的标准，各部门往往需要独立收集和處理相同的数据，这不仅耗费了大量的时间和财力，还容易导致数据不一致的问题。本标准通过建立统一的数据格式和处理流程，确保了数据的一致性和可比性，减少了不必要的重复劳动。例如，在环境监测领域，多个监测站采集的数据如果采用统一标准处理，可以轻松进行横向比较，发现不同区域之间的差异，为区域环境治理提供依据。此外，标准还促进了数据的共享与交换，使得不同机构之间可以更加便捷地共享数据资源，共同推进环保工作的开展。标准化处理流程还可以提高数据处理的自动化程度，减少人为干预，降低错误率，进一步提升数据处理的效率和准确性。

通过对环境监测数据的有效加工与处理，能够实时监控环境质量变化，及时发现环境问题，加强环境监管力度。环境监测数据是评估环境质量、制定环保政策的重要依据。标准的实施使得这些数据得以高效加工处理，从而为环境监管提供了强有力的支持。例如，通过对空气质量监测数据的实时分析，可以迅速发现污染物超标的情况，并及时通知相关部门采取措施，避免污染扩散。此外，标准还支持对历史数据的长期跟踪分析，帮助研究人员发现环境变化的趋势和规律，为制定长远的环保规划提供科学依据。通过对环境数据的深度挖掘，还可以揭示不同污染源之间的关联性，为源头治理提供线索。例如，在水污染治理中，通过对上游河流水质数据的分析，可以追踪污染源，指导地方政府加强水源地保护。总之，标准的实施有助于构建一个更加智能高效的环境监管体系，提高环境管理的精细化水平。

本标准的实施能够显著提高数据处理的自动化程度，减少人力成本，降低整体运营成本。随着信息技术的不断发展，自动化数据处理已成为大势所趋。标准通过引入先进的数据处理技术，如自然语言处理、机器学习等，实现了对环境文本数据的自动化处理。这样一来，不仅可以大幅降低人力投入，减轻工作人员负担，还能够提高数据处理的速度和精度。例如，在处理环境科研报告时，自动化工具可以迅速提取出研究报告中的核心结论，为决策者节省大量阅读时间。此外，标准还促进了数据处理流程的标准化，使得企业可以更加方便地引入自动化设备和技术，进一步降低成本。通过数据清洗与加工处理，提高数据质量，使得环境文本数据更具分析价值，为企业和研究机构创造更多商业机会。例如，高质量的数据可以为企业提供市场趋势分析，帮助其制定更合理的经营策略；对于研究机构而言，准确的数据是科学研究的基础，有助于推动环保技术的创新与发展。标准的推广与应用将促进环保领域内数据处理技术的进步与发展，推动相关技术创新，为环保产业带来新的增长点。

## 8 标准实施建议

**加强宣传培训：**组织面向环保机构、企业及科研单位的培训活动，详细介绍标准的内容、意义及实施方法，提高相关人员对标准的认知度和应用能力。通过举办培训班、研讨会和在线课程等多种形式，确保参与者能够全面掌握标准的核心要义，并能够在实际工作中灵活运用。

用，从而提升整体的数据处理水平。

**建立示范项目：**选取具有代表性的项目作为示范，通过实际应用案例展示标准的实施效果，为其他机构提供参考。示范项目应当覆盖不同类型的数据来源和应用场景，展示标准在提高数据质量、优化处理流程等方面的积极作用，以此激发更多机构主动采用标准的热情。

**技术支持服务：**设立专门的技术支持团队，为标准实施过程中遇到的技术难题提供解决方案，确保数据加工处理工作的顺利进行。技术支持团队应具备丰富的专业知识和实践经验，能够及时响应用户需求，帮助解决实施中的具体问题，保证标准在不同环境下的稳定运行。

**持续监督评估：**建立定期评估机制，跟踪标准实施的效果，收集实施过程中存在的问题与建议，为后续修订提供依据。评估工作应涵盖标准的适用性、有效性等多个方面，确保标准始终符合行业发展需求，并能及时调整改进，以适应不断变化的技术环境。

**促进交流合作：**鼓励跨部门、跨行业的合作交流，共享成功经验和最佳实践，促进环保领域的数据共享与协作。通过搭建合作平台，组织交流活动，推动各方共享资源、优势互补，共同提升环保工作的整体效能。

**强化政策配套：**建议相关部门出台配套政策措施，为标准的实施提供必要的政策支持，确保标准得到有效贯彻。政策配套可以包括资金扶持、技术支持、人才培养等多个方面，形成合力推动标准在实际工作中的广泛应用。

## 9 应用案例

气候政策在减缓气候变化和适应气候变化方面展现出了明显的成效。然而，相较于控制全球气温升幅这一气候变化目标，当前气候政策的实际效果与政策需求之间仍存在较大的差距。IPCC 第六次评估报告（AR6）明确指出，尽管国际社会已经在应对气候变化方面取得了一定进展，但仍需付出更大的努力才能实现巴黎协定设定的目标。报告呼吁科学家、经济学家和决策者共同努力，拓展气候治理的范围，以更有效地应对气候变化带来的挑战。AR6 提出，当前一个亟待面对的挑战是缺乏一套基于统一基准的可靠方法和量化手段来评估各层级各领域气候政策强度，进而理解这些政策工具的实施情况与实际效果。传统的评估方法往往依赖于专家打分或者简单的统计分析，难以全面捕捉到政策执行的复杂性和多样性。此外，由于缺乏统一的评估框架，不同国家和地区之间的政策比较也面临着一定困难，这限制了全球气候治理的协调性与有效性。因此，开发一种能够客观、系统地评估气候政策强度的方法显得尤为迫切。

在我国，气候治理体系是一个交叉行政架构，多个部门参与其中，涉及气候相关的政策分散在各个机构之中。生态环境部、自然资源部、科技部等政府部门都承担着不同的职责，它们发布的政策文件涵盖了从工业减排到森林碳汇管理等多个方面。这种多头管理的局面虽然体现了国家对于气候问题的高度重视，但也带来了政策协调方面的难题。全面系统回顾、梳理和评估这些气候政策对于我国气候治理取向一致性、统筹融合各领域气候政策工具、协同政策目标与措施效果极为重要。只有通过系统的评估，才能确保各项政策措施相互配合，共同推进我国气候治理目标的实现。然而，由于政策文件数量庞大，且内容繁杂，单纯依靠人工阅读和总结难以保证评估的全面性和准确性。

生态环境部环境规划院环境规划与政策模拟重点实验室（以下简称：重点实验室）技术组构建了一套集成高维词典-大语言模型-文本分析-机器学习的综合量化评估方法。该方法利用中国国务院公报等官方出版物刊载的数据，从涉及气候变化的减缓、适应和政策执行三个主要维度，以及 149 个子维度量化了 1954 年至 2022 年间中国气候政策强度（PSCC）。这一创新性的工作不仅填补了国内气候政策评估领域的空白，也为其他国家提供了可供借鉴的经验。技术组首先建立了包含数千个关键词和短语的高维词典，这些词汇涵盖了气候治理的主要方面，如碳排放、森林覆盖率、清洁能源等。然后，利用预训练的大语言模型对这些关

关键词进行文本嵌入处理，从而将非结构化的文本数据转换成结构化的数值表示。接着，采用机器学习算法对这些数值表示进行分析，提取出政策强度的相关特征。这一过程中，技术组特别关注了算法的可解释性，确保每一个决策步骤都有据可依，避免了黑箱操作带来的不确定性。

为了验证所提方法的有效性，技术组利用机器学习可解释器 SHAP (SHapley Additive exPlanations) 分析识别了影响 PSCC 变化趋势的关键驱动因素。通过与其他数据集进行验证对比，发现 PSCC 能够较好地表征我国气候治理的变化趋势与关键节点，有效回应了 AR6 提出关于气候政策的重要挑战。研究结果显示，经济水平、科技发展程度、国际合作力度等因素对 PSCC 的变化具有显著影响。特别是在经济转型期，随着产业结构的调整和技术进步，气候政策强度呈现出明显的上升趋势。此外，国际合作项目的增多也为气候治理注入了新的动力。技术组还对 PSCC 的时间序列进行了分析，揭示了政策强度随时间演变的规律性特征。研究发现，我国气候政策经历了从单一减排向多元化适应转变的过程，政策重心逐渐由短期应急转向长期战略规划。

该数据产品包含一套气候政策高维辞典、一套长时间序列（1954-2022 年）中国气候政策量化评估数据集和一套基于 149 个维度的 PSCC 变化趋势分解与关键变量识别数据集。这些成果为政策制定者提供了宝贵的参考信息，有助于他们在未来的气候治理中做出更加科学合理的决策。相关成果已公开发布在科学数据银行。