



(12) 发明专利申请

(10) 申请公布号 CN 121166830 A

(43) 申请公布日 2025. 12. 19

(21) 申请号 202511228986.2

G06F 18/22 (2023.01)

(22) 申请日 2025.08.29

G06F 18/241 (2023.01)

(71) 申请人 广东省城乡规划设计研究院科技集团股份有限公司

G06N 5/025 (2023.01)

地址 510290 广东省广州市海珠区南洲路483号

G06F 18/25 (2023.01)

G06V 30/148 (2022.01)

G06V 30/262 (2022.01)

G06Q 50/26 (2024.01)

(72) 发明人 潘俊钊 王晓路 吴晓生 陈得冠 陈俊松 李志恒 陈美妍 黄梓浩 魏冀明 周元

(74) 专利代理机构 广州粤高专利商标代理有限公司 44102 专利代理师 高棋 甘雅舒

(51) Int. Cl.

G06F 16/29 (2019.01)

G06F 16/22 (2019.01)

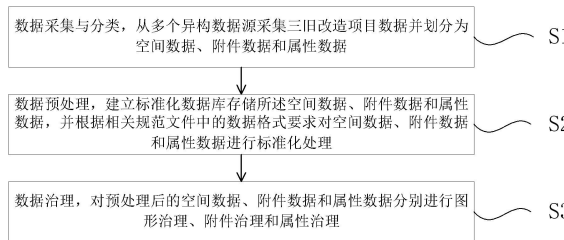
权利要求书4页 说明书14页 附图1页

(54) 发明名称

一种多源异构三旧改造项目数据治理方法及系统

(57) 摘要

本申请涉及一种多源异构三旧改造项目数据治理方法及系统,方法包括数据采集与分类,从多个异构数据源采集三旧改造项目数据并划分为空间数据、附件数据和属性数据;数据预处理,建立标准化数据库存储所述空间数据、附件数据和属性数据,并根据相关规范文件中的数据格式要求对空间数据、附件数据和属性数据进行标准化处理;数据治理,对预处理后的空间数据、附件数据和属性数据分别进行图形治理、附件治理和属性治理。本发明有效解决了三旧改造项目中多源数据整合难、标准化程度低、管理效率低下等问题,提升了数据质量与利用价值,为项目监管与决策提供了数据支撑。



1. 一种多源异构三旧改造项目数据治理方法,其特征在于,包括以下步骤:

数据采集与分类,从多个异构数据源采集三旧改造项目数据并划分为空间数据、附件数据和属性数据;

数据预处理,建立标准化数据库存储所述空间数据、附件数据和属性数据,并根据预定义的数据格式要求对空间数据、附件数据和属性数据进行标准化处理;

数据治理,对预处理后的空间数据、附件数据和属性数据分别进行图形治理、附件治理和属性治理。

2. 根据权利要求1所述的方法,其特征在于,所述数据预处理包括如下步骤:

将数据格式统一为标准化数据库的存储格式,在GIS平台建立标准化数据库,作为三旧改造项目数据的统一数据库容器,根据相关规范文件的要求,设置四个业务阶段子库,分别为三旧改造年度实施计划阶段、三旧改造方案阶段、三旧改造完善用地阶段、三旧改造项目实施阶段,在每个子库下设立对应的要素类,并为每个要素类导入对应的矢量范围,对同一项目的四个业务阶段先赋值一个同一标识码作为记录主键,以此统一全项目流程的唯一标识码;

检查每条属性数据记录所需附件是否齐全,并将附件统一为PDF标准格式。

3. 根据权利要求1所述的方法,其特征在于,所述数据治理包括如下步骤:

进行图形治理,对空间数据中的空间图形数据进行图形结构标准化处理、图斑拓扑检查与修复、图形版本与阶段归类管理并生成图形数据表;

进行附件治理,建立附件清单表,明确必备附件类型,对附件文件进行统一命名并存储至标准化数据库中;建立属性与附件的路径引用关系以构建附件映射表;进行附件完整性检查与缺漏预警;

进行属性治理,建立字段标准库,基于OCR技术从附件中识别文本信息,利用自然语言处理NLP技术从文本信息中提取语义字段并与字段标准库进行语义匹配,采用上下文结构判断字段归属,构建三旧改造领域知识图谱以对提取的字段进行容错修正与自动补全,获得标准化的属性字段数据。

4. 根据权利要求3所述的方法,其特征在于,所述图形治理具体包括:

图形结构标准化处理,统一空间图形数据的坐标系统与精度,对所有空间图形数据进行坐标转换并设置矢量数据精度,统一采用2000国家大地坐标系CGCS2000,投影方式为高斯-克吕格投影,并将空间图形数据中的各图形对象转化为面状要素,剔除冗余结构,包括多余节点、空几何对象和多余环;

图斑拓扑检查与修复,采用GIS平台内置的拓扑检查工具或自定义拓扑规则,对空间图形数据中的图斑执行无缝检查,排查并自动修复拓扑错误,包括图斑未闭合、悬挂线、图斑重叠、缝隙、多边关系错误;

图形版本与阶段归类管理,依据空间图形数据所属业务阶段进行分类入库,并按项目编号或数据采集批次设置图层版本号,生成图形数据表,实现空间图形数据的阶段追溯和更新管理。

5. 根据权利要求3所述的方法,其特征在于,所述附件治理具体包括:

进行附件清单梳理与统一命名,对采集的附件数据进行集中清点与整理,建立附件清单表,明确每个项目所需的必备附件类型与文件数量要求,所有附件根据相关规范文件的

要求进行统一命名,并存放至对应的标准化数据库的附件路径或外部文档库中,便于识别与调用;

进行主键挂接与目录结构生成,以项目编号作为唯一主键,对每条属性数据建立对应的附件路径引用关系,在标准化数据库中自动生成项目附件映射表,记录字段,包括项目ID、附件类型和文件路径,实现数据层与文件层的精准挂接,并支持一对多附件关联;

进行附件完整性检查与缺漏预警,根据附件清单表逐项目核查每类附件是否存在,若发现缺漏,自动标记为缺失,并生成附件缺漏报告推送至责任单位或源数据来源方,生成附件完整性统计表,并按业务阶段和项目状态分类汇总。

6. 根据权利要求3所述的方法,其特征在在于,所述属性治理具体包括:

依据相关规范文件建立字段标准库,明确字段名称、数据类型、编码规则和同义词典,用于后续自动映射和一致性管理;

对含有文本信息的附件采用OCR模型识别文本信息,并消除噪声与排版异常,转化为纯文本格式;

利用中文自然语言处理NLP模型对识别出的文本信息进行分词词性标注与命名实体识别,提取语义字段,并与字段标准库进行语义匹配,生成映射结果,并采用上下文结构判断字段归属,提高映射精度;

构建面向三旧改造领域的知识图谱,包括以下步骤:

进行领域实体抽取,依据相关规范文件和历史治理数据,自动识别并归类高频字段与典型业务属性,形成基础实体集;

通过正则表达式、句法分析与人工标注,自动标注实体属性及其语义关系,采用RDF或OWL标准格式建立三元组并进行存储;

进行同义词典生成与规范映射,结合分词、Word2Vec技术、BERT嵌入及历史人工复核库,构建同义词库,并采用哈希索引或倒排索引提升归一检索效率;

定期爬取和同步三旧改造项目数据来源的标准字段词表,实现多源知识融合,完成知识图谱的构建;

利用三旧改造领域的知识图谱进行属性字段的容错修正与自动补全,包括:

先通过知识图谱对OCR提取字段进行直接实体匹配校验,若校验通过即视为正确;

将直接实体匹配校验未通过字段利用BERT嵌入模型进行语义向量化,计算未通过字段与知识图谱实体的余弦相似度,同时利用Levenshtein距离计算文本编辑距离,将余弦相似度与文本编辑距离相结合形成综合相似度评分,综合相似度评分高于预设阈值则自动修正属性字段,低于预设阈值则进行异常字段标记与人工复核;

利用知识图谱内存储的语义关系对相似字段进行上下文交叉校核,动态验证字段正确性;

当利用OCR字段提取模型识别文本信息遗漏或未识别字段时,根据知识图谱的关联推理路径自动推断并补全字段;

根据知识图谱集成的标准单位转换规则,对属性字段进行统一换算与字段值规范

在属性治理过程中进行实时反馈与动态优化;

将人工复核结果、容错修正数据与自动补全数据实时反馈至所述知识图谱和字段提取模型;

利用反馈数据对知识图谱进行增量更新,同时,将三旧改造领域最新标准术语和业务规范定期同步至知识图谱,并对OCR模型与NLP模型进行增量学习与微调,以形成闭环的动态优化机制。

7. 根据权利要求6所述的方法,其特征在于,所述异常字段标记与人工复核具体包括:

通过对文本信息的识别,返回每个字段的识别置信度,如果OCR模型识别的文本在字符级别的置信度低于预设阈值,则将该字段标记为异常;

进行NLP语义匹配置信度评估,利用NLP模型对OCR文本进行语义解析,识别项目字段的正确性和语义一致性,根据识别出的字段类别和上下文语义判断是否匹配标准字段库,若匹配度低于设定阈值,则将该字段标记为异常;

将字段值与预设格式规则进行比对,若字段值不符合规则,将其标记为异常;

为了全面评估异常字段的质量,采用综合评分机制,结合OCR置信度、NLP匹配置信度、格式校验结果等多个评分指标;每个字段的评分按照以下公式进行加权计算

$$\text{Field Confidence Score} = w_1 \times \text{OCR Confidence} + w_2 \times \text{NLP Confidence} + w_3 \times \text{Format Consistency}$$

其中, $w_1$ 、 $w_2$ 、 $w_3$ 分别为OCR识别置信度、NLP匹配置信度评估和格式校验的权重,权重值根据实际需求进行调整,若最终分数低于预设定的阈值,则该字段被认为异常并进行人工复核与修复;

在人工复核过程中,复核人员对字段进行标记,分类存储修正结果,标记包括:

修复完成,表示复核人员确认该字段已正确,更新字段数据库;

待补充,该字段值不完整或无法确认,需进一步补充或等待其他数据源支持;

无法修复,若复核人员认为该字段的错误无法通过当前数据进行修复,标记为无法修复并提交进一步处理;

修复后的字段将通过API更新到标准化数据库,并生成复核记录表,包括复核时间、复核人员、修正内容及原因。

8. 根据权利要求3所述的方法,其特征在于,在所述属性治理后,还包括数据关联与入库步骤:

通过标准化的属性字段数据中的项目标识码、标准化数据库的外键约束与图形数据表及附件索引表,将治理后的属性数据与对应的治理后的空间数据、附件数据建立逻辑关联,实现三类数据之间的一致绑定与联动更新;

通过标准化数据库的批量写入接口将进行逻辑关联后的标准化的属性字段数据持久化存入标准化数据库,并在写入前进行属性字段数据与目标数据库的结构映射,对属性字段数据进行Schema校验合法性、正则规则检查、数值范围检测和时间格式解析,并对异常写入进行记录。

9. 一种多源异构三旧改造项目数据治理系统,用于权利要求1-8任一项所述的一种多源异构三旧改造项目数据治理方法,其特征在于,所述系统包括:

数据采集分类与预处理模块,用于执行所述数据采集分类步骤和数据预处理步骤;

数据治理引擎模块,用于执行所述数据治理步骤,包括图形治理单元、附件治理单元和属性治理单元;

知识图谱服务模块,用于存储三旧改造领域知识并提供语义匹配、容错修正和知识推

理服务；

数据入库与管理模块,用于执行数据关联与持久化入库操作。

10. 根据权利要求9所述的系统,其特征在于,所述属性治理单元包括:

OCR识别子单元,用于从附件中识别文本;

NLP解析子单元,用于从识别文本中提取语义字段;

知识图谱应用子单元,用于调用所述知识图谱服务模块进行字段修正与补充;

人工复核交互子单元,用于提供图形化界面供用户对系统标记的异常字段进行核查与修正。

## 一种多源异构三旧改造项目数据治理方法及系统

### 技术领域

[0001] 本申请涉及地理信息数据处理技术领域,特别是涉及一种多源异构三旧改造项目数据治理方法及系统。

### 背景技术

[0002] 三旧改造是指广东省特有的改造模式,分别是“旧城镇、旧厂房、旧村庄”改造。随着城市化进程的加速,如何高效利用现有土地资源,提升土地利用效率,成为城市发展的重要课题。“三旧”改造作为城市更新的重要手段,涉及大量复杂的数据管理与质量控制问题。传统的数据治理方法在处理“三旧”改造项目数据时,存在数据来源分散、格式不统一、质量参差不齐等问题,难以满足项目实施与监管的需求。

[0003] 目前存在的几项问题如下:

- 1) 数据源异构性强:空间数据格式不一,属性命名不统一,缺乏标准字段结构;
- 2) 非结构化资料难以提取:PDF批复件、规划文本等大量关键信息通过人工录入,效率低下且易出错;
- 3) 图属档关联弱:空间图形、属性、附件之间缺乏逻辑关联机制,信息割裂严重;
- 4) 数据质量不可控:缺少自动化质检工具,逻辑错误、数据冗余、附件缺失等问题频繁;
- 5) 成果输出不规范:图层混乱、字段不全。

[0004] 此外,现有的三旧数据治理技术在数据更新和维护方面也存在不足。数据更新不及时或不准确,可能导致土地数据与实际情况不符,进而影响土地资源的合理利用和管理。因此,需要一种能够提取融合三旧改造项目的多源异构数据、提高数据准确性和完整性的三旧改造项目数据治理方法及系统。

### 发明内容

[0005] 基于此,本发明的发明目的是解决上述技术问题,实现提取融合三旧改造项目的多源异构数据并对数据进行治理以提高数据准确性和完整性。

[0006] 为实现上述发明目的,本申请第一方面提供一种多源异构三旧改造项目数据治理方法,包括:

数据采集与分类,从多个异构数据源采集三旧改造项目数据并划分为空间数据、附件数据和属性数据;

数据预处理,建立标准化数据库存储所述空间数据、附件数据和属性数据,并根据相关规范文件中的数据格式要求对空间数据、附件数据和属性数据进行标准化处理;

数据治理,对预处理后的空间数据、附件数据和属性数据分别进行图形治理、附件治理和属性治理。

[0007] 优选地,所述数据预处理包括如下步骤:

将数据格式统一为标准化数据库的存储格式,在GIS平台建立标准化数据库,作为

三旧改造项目数据的统一数据库容器,根据相关规范文件的要求,设置四个业务阶段子库,分别为三旧改造年度实施计划阶段、三旧改造方案阶段、三旧改造完善用地阶段、三旧改造项目实施阶段,在每个子库下设立对应的要素类,并为每个要素类导入对应的矢量范围,对同一项目的四个业务阶段先赋值一个同一标识码作为记录主键,以此统一全项目流程的唯一标识码;

检查每条属性数据记录所需附件是否齐全,并将附件统一为PDF标准格式。

[0008] 优选地,所述数据治理包括如下步骤:

进行图形治理,对空间数据中的空间图形数据进行图形结构标准化处理、图斑拓扑检查与修复、图形版本与阶段归类管理并生成图形数据表;

进行附件治理,建立附件清单表,明确必备附件类型,对附件文件进行统一命名并存储至标准化数据库中;建立属性与附件的路径引用关系以构建附件映射表;进行附件完整性检查与缺漏预警;

进行属性治理,建立字段标准库,基于OCR技术从附件中识别文本信息,利用自然语言处理NLP技术从文本信息中提取语义字段并与字段标准库进行语义匹配,采用上下文结构判断字段归属,构建三旧改造领域知识图谱以对提取的字段进行容错修正与自动补全,获得标准化的属性字段数据。

[0009] 优选地,所述图形治理具体包括:

图形结构标准化处理,统一空间图形数据的坐标系统与精度,对所有空间图形数据进行坐标转换并设置矢量数据精度,统一采用2000国家大地坐标系CGCS2000,投影方式为高斯-克吕格投影,并将空间图形数据中的各图形对象转化为面状要素,剔除冗余结构,包括多余节点、空几何对象和多余环;

图斑拓扑检查与修复,采用GIS平台内置的拓扑检查工具或自定义拓扑规则,对空间图形数据中的图斑执行无缝检查,排查并自动修复拓扑错误,包括图斑未闭合、悬挂线、图斑重叠、缝隙、多边关系错误;

图形版本与阶段归类管理,依据空间图形数据所属业务阶段进行分类入库,并按项目编号或数据采集批次设置图层版本号,生成图形数据表,实现空间图形数据的阶段追溯和更新管理。

[0010] 优选地,所述附件治理具体包括:

附件清单梳理与统一命名,对采集的附件数据进行集中清点与整理,建立附件清单表,明确每个项目所需的必备附件类型与文件数量要求,所有附件根据相关规范文件的要求进行统一命名,并存放至对应的标准化数据库的附件路径或外部文档库中,便于识别与调用;

主键挂接与目录结构生成,以项目编号作为唯一主键,对每条属性数据建立对应的附件路径引用关系,在标准化数据库中自动生成项目附件映射表,记录字段,包括项目ID、附件类型和文件路径,实现数据层与文件层的精准挂接,并支持一对多附件关联;

附件完整性检查与缺漏预警,根据附件清单表逐项目核查每类附件是否存在,若发现缺漏,自动标记为缺失,并生成附件缺漏报告推送至责任单位或源数据来源方,生成附件完整性统计表,并按业务阶段和项目状态分类汇总。

[0011] 优选地,所述属性治理具体包括:

依据相关规范文件建立字段标准库,明确字段名称、数据类型、编码规则和同义词典,用于后续自动映射和一致性管理;

对含有文本信息的附件采用OCR模型识别文本信息,并消除噪声与排版异常,转化为纯文本格式;

利用中文自然语言处理NLP模型对识别出的文本信息进行分词词性标注与命名实体识别,提取语义字段,并与字段标准库进行语义匹配,生成映射结果,并采用上下文结构判断字段归属,提高映射精度;

构建面向三旧改造领域的知识图谱,包括以下步骤:

进行领域实体抽取,依据相关规范文件和历史治理数据,自动识别并归类高频字段与典型业务属性,形成基础实体集;

通过正则表达式、句法分析与人工标注,自动标注实体属性及其语义关系,采用RDF或OWL标准格式建立三元组并进行存储;

进行同义词典生成与规范映射,结合分词、Word2Vec技术、BERT嵌入及历史人工复核库,构建同义词库,并采用哈希索引或倒排索引提升归一检索效率;

定期爬取和同步三旧改造项目数据来源的标准字段词表,实现多源知识融合,完成知识图谱的构建;

利用三旧改造领域的知识图谱进行属性字段的容错修正与自动补全,包括:

先通过知识图谱对OCR提取字段进行直接实体匹配校验,若校验通过即视为正确;

将直接实体匹配校验未通过字段利用BERT嵌入模型进行语义向量化,计算未通过字段与知识图谱实体的余弦相似度,同时利用Levenshtein距离计算文本编辑距离,将余弦相似度与文本编辑距离相结合形成综合相似度评分,综合相似度评分高于预设阈值则自动修正属性字段,低于预设阈值则进行异常字段标记与人工复核;

利用知识图谱内存储的语义关系对相似字段进行上下文交叉校核,动态验证字段正确性;

当利用OCR字段提取模型识别文本信息遗漏或未识别字段时,根据知识图谱的关联推理路径自动推断并补全字段;

根据知识图谱集成的标准单位转换规则,对属性字段进行统一换算与字段值规范。

[0012] 在属性治理过程中进行实时反馈与动态优化:

将人工复核结果、容错修正数据与自动补全数据实时反馈至所述知识图谱和字段提取模型;

利用反馈数据对知识图谱进行增量更新,同时,将三旧改造领域最新标准术语和业务规范定期同步至知识图谱,并对OCR模型与NLP模型进行增量学习与微调,以形成闭环的动态优化机制。

[0013] 优选地,所述异常字段标记与人工复核具体包括:

通过对文本信息的识别,返回每个字段的识别置信度,如果OCR模型识别的文本在字符级别的置信度低于预设阈值,则将该字段标记为异常;

进行NLP语义匹配置信度评估,利用NLP模型对OCR文本进行语义解析,识别项目字段的正确性和语义一致性,根据识别出的字段类别和上下文语义判断是否匹配标准字段

库,若匹配度低于设定阈值,则将该字段标记为异常;

将字段值与预设格式规则进行比对,若字段值不符合规则,将其标记为异常;

为了全面评估异常字段的质量,采用综合评分机制,结合OCR置信度、NLP匹配置信度、格式校验结果等多个评分指标。每个字段的评分按照以下公式进行加权计算:

$$\text{Field Confidence Score} = w_1 \times \text{OCR Confidence} + w_2 \times \text{NLP Confidence} + w_3 \times \text{Format Consistency}$$

[0014] 其中, $w_1$ 、 $w_2$ 、 $w_3$ 分别为OCR识别置信度、NLP匹配置信度评估和格式校验的权重,权重值根据实际需求进行调整,若最终分数低于预设定的阈值,则该字段被认为异常并进行人工复核与修复;

在人工复核过程中,复核人员对字段进行标记,分类存储修正结果,标记包括:

修复完成,表示复核人员确认该字段已正确,更新字段数据库;

待补充,该字段值不完整或无法确认,需进一步补充或等待其他数据源支持;

无法修复,若复核人员认为该字段的错误无法通过当前数据进行修复,标记为无法修复并提交进一步处理;

修复后的字段将通过API更新到标准化数据库,并生成复核记录表,包括复核时间、复核人员、修正内容及原因。

[0015] 优选地,在所述属性治理后,还包括数据关联与入库步骤:

通过标准化的属性字段数据中的项目标识码、标准化数据库的外键约束与图形数据表及附件索引表,将治理后的属性数据与对应的治理后的空间数据、附件数据建立逻辑关联,实现三类数据之间的一致绑定与联动更新;

通过标准化数据库的批量写入接口将进行逻辑关联后的标准化的属性字段数据持久化存入标准化数据库,并在写入前进行属性字段数据与目标数据库的结构映射,对属性字段数据进行Schema校验合法性、正则规则检查、数值范围检测和时间格式解析,并对异常写入进行记录。

[0016] 为实现发明目的,本申请第二方面提供一种多源异构三旧改造项目数据治理系统,用于上述技术方案所述的一种多源异构三旧改造项目数据治理方法,所述系统包括:

数据采集分类与预处理模块,用于执行所述数据采集分类步骤和数据预处理步骤;

数据治理引擎模块,用于执行所述数据治理步骤,包括图形治理单元、附件治理单元和属性治理单元;

知识图谱服务模块,用于存储三旧改造领域知识并提供语义匹配、容错修正和知识推理服务;

数据入库与管理模块,用于执行数据关联与持久化入库操作。

[0017] 优选地,

OCR识别子单元,用于从附件中识别文本;

NLP解析子单元,用于从识别文本中提取语义字段;

知识图谱应用子单元,用于调用所述知识图谱服务模块进行字段修正与补全;

人工复核交互子单元,用于提供图形化界面供用户对系统标记的异常字段进行核查与修正。

[0018] 本发明与现有技术相比,其有益效果是:

本发明通过多源异构数据采集与分类、标准化预处理及分类治理,实现了三旧改造项目数据的一体化整合与深度规范化。其通过图形治理保障了空间数据的拓扑正确性与版本可控性,通过附件治理实现了附件的标准化管理与智能检漏,通过属性治理结合OCR、NLP与知识图谱技术实现了非结构化数据的语义识别、智能匹配与自动补全,显著提升了数据质量与一致性。该方法有效解决了多源异构数据融合难、标准化程度低、附件管理混乱、属性提取效率低下等问题,为三旧改造项目的科学决策、动态监管与业务协同提供了完整、准确、可靠的数据基础。

## 附图说明

[0019] 图1为一个实施例中一种多源异构三旧改造项目数据治理方法的步骤流程示意图。

## 具体实施方式

[0020] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。以下实施例用于说明本发明,但不用来限制本发明的范围。

[0021] 实施例1

本申请实施例1提供一种多源异构三旧改造项目数据治理方法,如图1所示,包括如下步骤:

S1:数据采集与分类,从多个异构数据源采集三旧改造项目数据并划分为空间数据、附件数据和属性数据;

S2:数据预处理,建立标准化数据库存储所述空间数据、附件数据和属性数据,并根据相关规范文件中的数据格式要求对空间数据、附件数据和属性数据进行标准化处理;

S3:数据治理,对预处理后的空间数据、附件数据和属性数据分别进行图形治理、附件治理和属性治理;

其中,步骤S1包括:S11

S11:从省土地市场系统、自然资源部门、住建部门数据与其他相关部门采集三旧改造相关资料数据,其中数据类型至少包括pdf档案文件、shp文件、gdb文件、cad文件、excel文件、宗地图文件,txt文件等。

[0022] 所述步骤S2数据预处理具体包括:S21-S22

S21:将数据格式统一为标准化数据库的存储格式File Geodatabase (FGDB),在ArcGIS或等效GIS平台中创建标准的File Geodatabase (FGDB)数据库,作为三旧改造项目数据的统一数据库容器,根据相关规范文件的要求,设置四个业务阶段子库,分别为三旧改造年度实施计划阶段、三旧改造方案阶段、三旧改造完善用地阶段、三旧改造项目实施阶段,在每个子库下设立对应的要素类,并为每个要素类导入对应的矢量范围,对同一项目的四个业务阶段的先赋值一个同一标识码作为记录主键,以此统一全项目流程的唯一标识码;

S22:检查每条属性数据记录所需附件是否齐全,并将附件统一为PDF标准格式。

[0023] S3数据治理:本步骤针对多源异构“三旧”改造项目数据在属性和图形层面存在的标准不统一、语义不一致、逻辑冲突及拓扑错误等问题,分别开展属性治理与图形治理,提升数据规范化程度与质量可控性,包括步骤S31-S33:

S31:数据的图形治理

进行图形治理,对空间数据中的空间图形数据进行图形结构标准化处理、图斑拓扑检查与修复、图形版本与阶段归类管理并生成图形数据表。该步骤聚焦空间图形数据的合法性、完整性与逻辑正确性,确保图斑数据符合拓扑规则并支持分析应用。

[0024] 其中,步骤S31包括步骤S311-S313:

S311:图形结构标准化处理

统一所有空间图形数据的坐标系统(CGCS2000)和投影方式(高斯-克吕格投影),并将各图形对象转化为面状要素,剔除多余节点、空几何对象和多余环等冗余结构。

[0025] S312:图斑拓扑检查与修复

采用GIS平台内置的拓扑检查工具或自定义拓扑规则,对图形执行无缝检查,排查并自动修复图斑未闭合、悬挂线、图斑重叠、缝隙、多边关系错误等拓扑错误。

[0026] S313:图形版本与阶段归类管理

依据图形数据所属业务阶段(如:年度计划、方案、完善用地、项目实施等)进行分类入库,并按项目编号或数据采集批次设置图层版本号,实现图形数据的阶段可追溯、更新可管理。

[0027] S32:附件治理

进行附件治理,建立附件清单表,明确必备附件类型,并按照相关规范文件的要求对附件文件进行统一命名并存储至标准化数据库中;建立属性与附件的路径引用关系以构建附件映射表;进行附件完整性检查与缺漏预警。本步骤针对“三旧”改造项目中广泛存在的附件材料(如供地合同、批复文件、图则图纸、会议纪要、项目备案函等),开展系统化、结构化的治理,确保档案资料的完整性、关联性与可检索性,为后续图属档一体化管理与汇交提供支撑。

[0028] 其中,步骤S32包括步骤S321-S323:

S321:附件清单梳理与统一命名

对从各部门采集的原始档案资料进行集中清点与整理,建立《附件清单表》,明确每个项目所需的必备附件类型(如:“批复文件”、“供地协议”、“改造方案”、“补偿协议”、“完成改造证明照片”等)与文件数量要求。所有附件统一命名规范为:项目编号\_附件类型.pdf,如HY202501001供地批复.pdf,并存放至对应的FGDB附件路径或外部文档库中,便于系统识别与调用。

[0029] S322:主键挂接与目录结构生成

以项目编号作为唯一主键,对每条属性记录建立对应的附件路径引用关系。系统在FGDB中自动生成项目附件映射表(AttachmentLinkTable),记录“项目ID-附件类型-文件路径”等字段,实现数据层与文件层的精准挂接,并支持一对多附件关联。

[0030] S323:附件完整性检查与缺漏预警

系统按清单项逐项目核查每类附件是否存在,若发现缺漏(如某项目缺失“供地协议”),自动标记为“缺失”,并加入《附件缺漏报告》,推送至责任单位或原数据来源方。支持

生成附件完整性统计表,并按阶段、项目状态分类汇总。

**[0031] S33:数据的属性治理**

进行属性治理,建立字段标准库,基于OCR技术从附件中识别文本信息,利用自然语言处理NLP技术从文本信息中提取语义字段并与字段标准库进行语义匹配,采用上下文结构判断字段归属,构建三旧改造领域知识图谱以对提取的字段进行容错修正与自动补全,获得标准化的属性字段数据。该步骤融合OCR识别、NLP语义分析、知识图谱及容错机制等技术,针对“三旧”改造项目中结构化与非结构化字段提取难、字段错误率高、数据一致性的问题,设计了一套具备自学习和容错修复能力的属性治理流程。治理对象涵盖各类PDF批复材料、纸质扫描件及历史项目档案,确保数据字段标准化、可追溯与一致性闭环。

**[0032]** 其中,步骤S33包括步骤S331-S337:

**S331:字段标准库构建**

依据相关规范文件的要求建立属性字段标准库,明确字段名称、数据类型、编码规则和同义词典,用于后续自动映射和一致性管理。a“三旧”改造年度实施计划阶段,含批准日期、项目面积、项目位置等字段;b“三旧”改造方案阶段,含地块面积、方案批准机关、方案批准时间、改造模式、改造类型、改造类型细分、供应方式.等字段;c“三旧”改造方案阶段,含地块面积、完善用地手续方式、是否已完成完善用地手续、批准文号、批准时间、批准机关.等字段;d“三旧”改造项目实施阶段,含改造前土地用途、改造后土地用途、电子监管号、项目总投资等字段。

**[0033] S332:PDF档案OCR识别与文本预处理**

对来源于住建、自然资源等部门的批复文件、项目方案、供地合同等PDF格式档案,采用OCR技术(如Tesseract或PaddleOCR)进行识别,提取正文信息,消除噪声与排版异常,转化为纯文本格式。

**[0034] S333:NLP语义解析与字段提取**

利用中文自然语言处理工具(如LTP、BERT等)对识别文本进行分词词性标注与命名实体识别,抽取如“项目编号”、“批准文号”、“宗地面积”、“土地用途”等语义字段,并与字段标准库进行语义比对,生成映射结果,同时结合预定义的字段标准库进行语义匹配,采用上下文结构判断字段归属,提高映射精度。

**[0035] S334:基于知识图谱的字段容错修正机制**

本步骤通过构建面向“三旧”领域的专用知识图谱,实现对属性字段提取结果的自动纠错、同义归一、字段补全与异常修复。其核心包括领域知识建模、知识图谱实体关系抽取、规则与模型双驱动的修正算法、智能映射与容错推理。

**[0036]** 其中,步骤S334包括步骤S3341-S3343:

**S3341. 知识图谱构建:**

领域实体抽取:依据相关规范文件和历史治理数据,自动识别并归类如“项目类型”、“批准机关”、“土地用途”、“地名”、“批文号”等高频字段与典型业务属性,形成基础实体集。

**[0037]** 属性与关系建模:通过正则表达式、句法分析与人工标注,自动标注实体属性(如地名、地址、单位、批文时间、项目编号)及其语义关系(如“隶属于”“归属于”“变更为”等),采用RDF/OWL标准格式建立三元组(subject-predicate-object)存储。

[0038] 同义词典生成与规范映射:结合分词、Word2Vec/BERT嵌入及历史人工复核库,构建同义词库(如“旧厂区”“旧工厂”“厂区用地”归一化为“工业用地”),并以哈希索引或倒排索引提升归一检索效率。知识融合:定期爬取和同步住建、自然资源等部门标准字段词表,实现多源知识融合。

[0039] S3342.容错修正与自动补全算法:

字段初步匹配校验:OCR提取字段先通过知识图谱进行直接实体匹配校验,若校验通过即视为正确。

[0040] 模糊匹配与相似度计算:未直接匹配字段通过BERT嵌入模型进行语义向量化,计算待校验字段与图谱实体的余弦相似度,同时利用Levenshtein距离计算文本编辑距离。两种算法相结合形成综合相似度评分,高于阈值自动修正字段值;低于阈值标记为异常待复核。

[0041] 上下文语义动态校核与推理:对于中度相似字段,利用知识图谱内存储的语义关系(如“项目编号-项目位置-批准机关”等多元关系)进行上下文交叉校核,动态验证字段正确性。例如,当OCR识别结果与图谱记录的项目位置不一致时,系统自动调用上下文关系链对字段进行动态纠错。

[0042] 自动补全缺失字段:OCR遗漏或未识别字段时,系统根据知识图谱的关联推理路径自动推断补全字段。例如,若OCR缺失“项目编号”,但存在“项目位置”和“批准机关”,系统自动推理匹配该项目的编号信息,并进行自动补全。

[0043] 字段单位标准化与换算:知识图谱集成标准单位转换规则,如面积单位(亩、公顷、平方米)、时间单位(年、月、日)的识别与标准化,自动进行统一换算与字段值规范存储。

[0044] S3343.实时反馈与动态优化:

实时反馈更新机制:每次系统完成字段容错修正与补全后,即刻将修正后的映射规则与结果实时反馈入知识图谱作为增量知识,以提升后续OCR识别容错修正的自动化精度。

[0045] 定期动态更新与模型联动:知识图谱定期自动同步领域最新标准术语和业务规范,并通过API通知OCR与NLP模型进行参数微调与再训练。动态更新机制确保系统始终具备最新领域知识,从而持续提高识别准确性。

[0046] 识别-反馈-训练-优化闭环:构建一个闭环机制,系统在实际应用中自动记录并分析纠错成功与失败案例,动态微调模型与知识图谱规则,建立长期持续优化和自适应进化机制,确保系统在处理复杂文书、历史档案、模糊扫描件时识别与纠错性能不断提升。

[0047] S335:主键绑定与数据关联结构设计

在结构化属性字段写入前,系统首先根据每条记录中识别出的项目编号SJGZ\_XMBH建立主键索引1。该主键作为属性数据的唯一标识符,并通过外键约束与图形数据表(如T\_PROJECT\_GEOM)及附件索引表(如T\_PROJECT\_ATTACHMENT)进行逻辑关联,实现图、属、档三类数据之间的一致绑定与联动更新,确保数据链路的完整性与可追溯性。

[0048] S336:字段写入与数据库持久化机制

在完成主键绑定后,系统调用数据库写入接口(ArcPy.da 模块),将标准化后的属性字段内容写入指定的成果数据库表(如T\_PROJECT\_ATTR)。在写入过程中,字段名称、字段类型与字段长度必须严格匹配数据库 Schema 定义的表结构,防止因结构不一致引发字

段错写、类型冲突或截断异常,确保数据入库的准确性与结构兼容性。

[0049] 其中,步骤S336包括步骤S3361-S3363:

S3361字段结构映射与Schema校验机制

在字段标准化完成并完成主键绑定后,系统首先执行属性字段与目标数据库Schema的结构映射过程。利用字段映射配置文件(Field Mapping JSON 或映射表),系统将提取字段名称与数据库目标字段进行一一对应匹配。映射过程中调用元数据校验逻辑,自动检查字段类型(如 Text、Float、Date)、长度限制、是否为必填字段等参数,并验证与目标表(如 T\_PROJECT\_ATTR)的Schema一致性,避免因字段错配或数据溢出导致写入失败。

[0050] S3362批量写入流程与接口调用技术路径

属性字段写入通过 Python ArcPy.da 模块中的 UpdateCursor 或 InsertCursor 实现高性能批量入库。在调用 ArcPy 的数据访问模块(Data Access Module)过程中,系统按照事务控制机制(transaction buffer)批量构造字段数据数组,将字段数据逐行写入成果数据库(File Geodatabase 或 Enterprise Geodatabase)。写入前系统执行数据类型转换与空值处理逻辑(如 null padding、单位归一、编码格式转换),确保字段值满足数据库类型约束。

[0051] S3363字段一致性保障与异常写入防控机制

为防止字段写入过程中发生字段缺失、类型冲突、非法字符等异常,系统在写入阶段启用字段合法性校验器(Field Validator Module)。校验器基于规则库对每条记录的字段值执行正则规则检查、数值范围检测和时间格式解析,如检测“批准时间”字段是否为“YYYY-MM-DD”合法格式、面积字段是否为正数、行政区划字段是否为有效代码。若发现不符合条件的记录,则跳过写入并记录至“字段写入异常日志表”,标明失败原因与异常字段。

[0052] S337:异常字段标记与人工复核接口

本系统的核心目标是通过高效的异常检测、人工复核和自适应学习机制,提升数据质量,确保“三旧”改造项目的数据在字段提取和语义匹配中的高精度。整个过程实现了“模型-人工-模型”的闭环优化,不仅能自动标记低质量字段,还能通过人工反馈持续优化异常检测能力。

[0053] 其中,步骤S337包括步骤S3371-S3373:

S3371:异常识别与评分机制系统

步骤S3371包括步骤1.1和步骤1.2:

1.1 异常字段检测

异常字段检测模块基于多种算法,通过综合多维度的评分机制,识别字段中可能存在的错误或不一致性。系统采用以下几个子模块:

OCR识别置信度评估:OCR识别模块通过对图像中文字的识别,返回每个字段的识别置信度(如Tesseract OCR的OCR-Confidence)。如果OCR识别的文本在字符级别的置信度低于预设阈值(例如70%),则该字段将被标记为异常。

[0054] NLP语义匹配置信度评估:NLP模型(如BERT或LTP)对OCR文本进行语义解析,识别项目字段的正确性和语义一致性。模型根据识别出的字段类别和上下文语义判断是否匹配标准字段库。若匹配度低于设定阈值(例如80%),则该字段会被标记为异常。

[0055] 格式与规则校验:字段值与预设格式规则(如日期格式、编号格式、数值范围)进行

比对,若字段不符合规则,将自动标记为异常。例如,日期字段未按“YYYY-MM-DD”格式存储,或者面积字段包含负数等。

#### [0056] 1.2 异常字段评分机制

为了全面评估异常字段的质量,系统采用综合评分机制,结合OCR置信度、NLP匹配置信度、格式校验结果等多个评分指标。每个字段的评分按照以下公式进行加权计算:

$$\text{Field Confidence Score} = w_1 \times \text{OCR Confidence} + w_2 \times \text{NLP Confidence} + w_3 \times \text{Format Consistency}$$

[0057] 其中, $w_1$ 、 $w_2$ 、 $w_3$ 分别为OCR识别置信度、NLP匹配置信度评估和格式校验的权重,权重值根据实际需求进行调整,若最终分数低于设定的阈值(如80%),该字段被认为异常并进行人工复核与修复。

[0058] 在复核过程中,复核人员可以对字段进行标记,分类存储修正结果。标记包括:

“修复完成”:表示复核人员确认该字段已正确,系统自动更新字段数据库。

[0059] “待补充”:若复核人员认为该字段值不完整或无法确认,需进一步补充或等待其他信息源支持。

[0060] “无法修复”:若复核人员认为该字段的错误无法通过当前文档修复,标记为“无法修复”并提交进一步处理。

[0061] 修复后的字段将自动反馈给系统,通过API更新到主数据库,并生成复核记录表,包含复核时间、复核人员、修正内容及原因。

#### [0062] 样本积累与动态微调:

每次人工复核后的字段修正结果都会被记录并自动反馈给模型,作为增量数据加入样本库。通过不断积累人工修正过的数据,系统能够逐步优化OCR与NLP模型,提高对复杂文书(如手写档案、扫描不清晰的文档等)的识别能力。增量学习:系统定期从复核数据中提取训练样本,通过增量学习更新OCR与NLP模型,优化识别能力。此过程主要使用有监督学习方法,基于人工修正数据对模型进行微调。自适应模型训练:根据不同类型的文档(如批复文件、合同协议、供地审批等)进行模型训练优化,提升对特定文档类型的识别精度。

#### [0063] 知识图谱自进化:

系统通过知识图谱与历史数据的结合,自动进行数据补全和错误修正,形成“模型-复核-训练-优化”的闭环。每次复核和修复后,系统会将新的实体及其修正结果写入到知识图谱,并通过智能推理补充缺失字段。例如,对于“项目编号”和“批准文号”等关键字段,系统会将复核后的修正结果和关联数据加入到知识图谱中,并使用图谱推理功能自动纠正类似字段中的错误,如拼写错误、同义词映射等。

#### [0064] 异常容错与智能补全:

结合知识图谱的规则推理和容错机制,系统能够对常见错别字、同义词进行智能修复。例如,“土地划拨”和“划拨用地”都统一映射为“划拨用地”标准字段。通过与领域库的结合,系统能够根据字段的上下文和语义关系,自动纠正常见字段的错误,并且对遗漏的字段进行补全,提高字段的完整性。

#### [0065] 实施例2

本申请实施例2基于实施例1,提供一种多源异构三旧改造项目数据治理系统,用于实施例1的一种多源异构三旧改造项目数据治理方法,该系统包括:

数据采集分类与预处理模块,用于执行所述数据采集分类步骤和数据预处理步骤;

数据治理引擎模块,用于执行所述数据治理步骤,包括图形治理单元、附件治理单元和属性治理单元;

知识图谱服务模块,用于存储三旧改造领域知识并提供语义匹配、容错修正和知识推理服务;

数据入库与管理模块,用于执行数据关联与持久化入库操作。

[0066] 所述属性治理单元包括:

OCR识别子单元,用于从附件中识别文本;

NLP解析子单元,用于从识别文本中提取语义字段;

知识图谱应用子单元,用于调用所述知识图谱服务模块进行字段修正与补全;

人工复核交互子单元,用于提供图形化界面供用户对系统标记的异常字段进行核查与修正。

[0067] 为确保低置信度或异常字段的精确修复,系统为复核人员提供了一个图形化的人工复核界面。此界面通过以下模块支持复核工作:

字段对比模块:展示OCR提取字段和原始图像的对比,便于复核人员检查识别文本与图像是否一致。

[0068] 建议值显示:基于NLP的语义匹配,系统会为每个字段提供修正建议,复核人员可以查看最匹配的字段值。例如,对于“项目编号”字段,系统会自动显示最匹配的编号,并标明匹配度。

[0069] 手动修正与字段选择:复核人员可手动选择正确字段或修正错误字段,修正过程可以实时更新到数据库,并可支持字段值的批量修复。

[0070] 实施例3

本实施例3基于实施例1和实施例2,对一个“HY202501001中山市某某有限公司旧厂房改造项目”数据治理完整案例进行说明。

[0071] S1 数据采集与分类

以广东省中山市“HY202501001中山市某某有限公司旧厂房改造项目”改造项目为例,从以下部门获取资料:

自然资源局:宗地图(DWG格式)、项目方案文件(PDF)、项目方案批复文件(PDF)

住建局:改造方案(CAD、PDF)

土地市场部门:供地合同(PDF)、土地面积统计表(Excel)分类保存于相应数据目录。

[0072] S2 数据预处理

S21 统一数据格式

在ArcGIS创建标准化FGDB数据库,包含四个阶段性子库:

HY202501001\_年度实施计划

HY202501001\_改造方案

HY202501001\_完善用地手续

HY202501001\_项目实施阶段

将DWG、CAD文件转化为SHP后导入FGDB;Excel数据导入表格;附件统一命名为“HY202501001\_附件类型.pdf”并存入附件库。

[0073] S22 统一坐标系统与精度

统一采用2000国家大地坐标系(CGCS2000),高斯-克吕格投影38度带,完成空间数据统一。

[0074] S23 附件数据清点与结构挂接

检查附件,缺失“补偿协议”、“改造前实地照片”、“完成改造照片”,自动标记后生成报告,通知责任单位补充。

[0075] S3 数据治理

S31 图形治理

S311 图形结构标准化处理清理冗余节点、多余环,转换为标准面状要素。

[0076] S312 图斑拓扑检查与修复自动修复未闭合、缝隙与重叠区域,形成标准化图斑。

[0077] S313 图形版本与阶段归类管理存储于FGDB“改造方案”阶段,标记为版本V1.0。

[0078] S32 附件治理

S321 附件清单梳理与统一命名附件统一命名如:“HY202501001\_方案文件.pdf”、“HY202501001\_方案批复文件.pdf”、“HY202501001\_供地合同.pdf”。

[0079] S322 主键挂接与目录结构生成生成附件映射表(Attachment Link Table),实现附件与属性数据自动关联。

[0080] S323 附件完整性检查与缺漏预警检测到“补偿协议”附件缺失,自动标记,并发送通知进行补齐。

[0081] S33 属性治理

S331 字段标准库构建

根据相关规范文件《“三旧”改造数据治理技术规范(征求意见稿)》建立“项目编号”、“批准机关”、“项目面积”、“改造类型细分”……等标准字段。

[0082] S332 PDF档案OCR识别与文本预处理OCR识别批复文件,结果:

项目编号:HY202501001

批准机关:中山市自然资源局

批准日期:2025年1月12日

项目面积:50亩

改造模式:旧厂区改造(拆除新建)

S333 NLP语义解析与字段提取NLP抽取字段,检测出OCR误差:

批准机关:“中山市自然资源局”(异常)

改造类型细分:“旧厂区改造(拆除新建)”(非标准术语)

项目面积:“50亩”(需单位转换)

S334 基于知识图谱的字段容错修正机制

(1) 知识图谱构建

标准实体:“中山市自然资源局”、“旧厂区改造(拆除重建)”、“亩→平方米转换关系”等。

[0083] 实体关系三元组存储:

<HY202501001项目, 批准机关, 中山市自然资源局>

<HY202501001项目, 项目面积, 33333.5平方米>

<旧厂区改造(拆除重建), 属于,改造类型细分>

构建同义词及错别字映射:

“自然资源居”→“自然资源局”

“拆除新建”→“拆除重建”

(2) 容错修正流程

初步匹配校验与异常标记:未匹配字段进入模糊匹配,如“自然资源居”、“拆除新建”。

[0084] 模糊匹配与相似度计算:通过BERT模型语义相似度和Levenshtein编辑距离确定:

“中山市自然资源居”自动修正为“中山市自然资源局”(相似度92%)

“拆除新建”自动修正为“拆除重建”(相似度95%)

上下文语义动态校核与推理:基于知识图谱关系链校核字段修正有效性。

[0085] 自动补全缺失字段:OCR未识别“项目位置”,但是通过改造名称主体为“中山市XXX玻璃厂”,知识图谱推理补全:“中山市人民路1号”。

[0086] 字段单位识别与自动换算:自动转换面积:“50亩”→“33333.5平方米”。

[0087] (3) 图属OCR面积不匹配情况的处理(新增细化)

OCR识别面积:“50亩(33333.5平方米)”

空间图斑计算面积:“33000平方米”

面积误差阈值设定为 $\pm 1\%$ (约333.3平方米),实际误差为333.5平方米(略超阈值),标记为异常。

[0088] 自动推送至人工复核界面,由人工确定:

人工复核流程:

检查原始批复文件扫描件与空间图斑。

[0089] 确认OCR识别面积“50亩”为正确,图斑面积计算误差较大。

[0090] 人工修改空间图形节点,调整图形精度,使计算面积达到33333.5平方米。

[0091] 反之,若人工复核发现OCR错误,人工确认并修正OCR面积,更新知识图谱容错规则。

[0092] (4) 实时反馈更新与动态优化

将人工复核结果及纠错实例加入知识图谱,实现实时知识更新与OCR模型再训练,提高系统整体识别精度。

[0093] S335 主键绑定与数据关联结构设计

主键HY202501001关联属性、空间与附件数据,确保数据链路完整。

[0094] S336 字段写入与数据库持久化机制

使用ArcPy.da模块批量、高效写入字段数据,确保数据Schema准确一致。

[0095] S337 异常字段标记与人工复核接口

OCR置信度低于阈值,自动推送人工复核系统。

[0096] 人工复核后修正结果自动更新至数据库,并反馈更新到知识图谱。

[0097] 实现数据质量的持续提升与模型闭环优化。

**[0098]** 案例实施后数据治理成果展示

经上述全过程治理后, HY202501001项目各类数据准确、完整、一致性高, 最终入库的部分数据如表1所示:

字段	数据治理后内容	数据来源及修正说明
项目编号	HY202501001	识别准确
批准机关	中山市自然资源局	OCR识别纠错
批准日期	20251201	标准格式转换
项目面积	33333.5m <sup>2</sup>	单位转换为平方米
改造类型细分	拆除重建	术语归一修正
项目位置	中山市人民路1号	自动推理补全
图形面积	33333.5m <sup>2</sup>	人工图形校正后精确匹配

**[0099]** 表1 部分入库数据表

案例实施效果与技术创新总结:

明确处理了图属面积误差的分类处理方法(OCR错或图形错), 引入人工复核, 实现更精细化数据治理。通过OCR识别、NLP解析与知识图谱容错纠错, 形成了完整、高效的数据治理闭环。整个流程有效提升了“三旧”改造数据治理效率与质量, 明显减少了人工介入并提高数据准确度。

**[0100]** 综上所述, 本发明提供一种多源异构三旧改造项目数据治理方法及系统, 通过多源异构数据采集与分类、标准化预处理及分类治理, 实现了三旧改造项目数据的一体化整合与深度规范化。其通过图形治理保障了空间数据的拓扑正确性与版本可控性, 通过附件治理实现了附件的标准化管理与智能检漏, 通过属性治理结合OCR、NLP与知识图谱技术实现了非结构化数据的语义识别、智能匹配与自动补全, 显著提升了数据质量与一致性。该方法有效解决了多源异构数据融合难、标准化程度低、附件管理混乱、属性提取效率低下等问题, 为三旧改造项目的科学决策、动态监管与业务协同提供了完整、准确、可靠的数据基础。

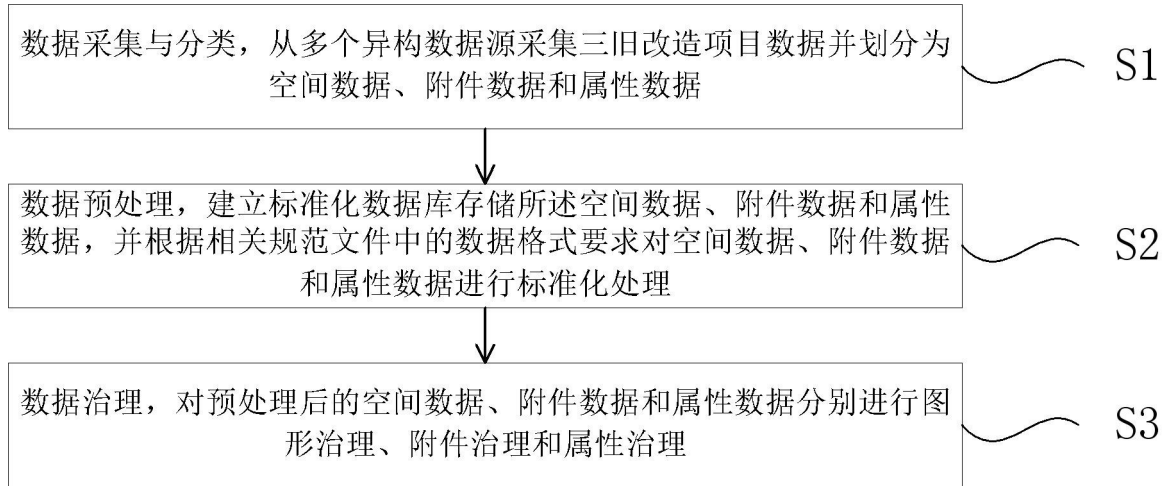


图1