

基于 XML 的土壤系统分类检索系统的元数据模型

汪善勤 周勇 于雷 舒宁

武汉大学遥感信息工程学院 武汉

华中农业大学资源与环境学院 武汉

摘要 土壤系统分类的检索系统是以量化特征为主 描述土壤的知识体系 具有严格的逻辑一致性和完整性 是建立土壤信息元数据模型的重要基础。是先进的网络语言标准 采用可扩展的标记 能够描述各种数据结构 是网络信息共享的重要工具。本研究分析了目前土壤信息共享中存在的问题 提出了以中国土壤系统分类检索系统为主体 参照中国土壤分类与代码的国家标准和 地理信息元数据标准 采用 为描述语言 建立的土壤信息元数据模型 为实现土壤信息共享提供解决方案。

关键词 土壤系统分类 元数据

中图分类号

文献标识码

地球科学信息共享是实施数字地球的基本条件 为此 国际标准组织 在 年推出了一系列地理信息的数据标准 涵盖了数据参考模型、一致性检验、时空规范、坐标参照系、地理标识、数据质量、元数据等方面 为实现地理信息共享奠定了基础。土壤信息是地球信息的重要组成部分 是国家基础地理信息建设的重要内容之一 在数字农业建设、生态环境保护和国土资源调查与评价中发挥着重要作用。土壤是一个具有明显时空特征的复杂系统 在信息采集和数量化表达上的数据量非常庞大。我国在第二次土壤普查中 获取了大量翔实的土壤数据 出版了《中国土壤图集》和： 万县级土壤图。 世纪 年代 土壤侵蚀数据库、红壤信息系统、水土流失信息系统等一批专题土壤数据库的建立 为土壤信息化提供了数据积累和应用经验。 年代 建立了以量化为主要特征的中国土壤系统分类 其土系数据库在基础研究 土壤发生与分类 和应用之间架设了桥梁。我国部分省市已建立不同比例尺的 土壤 地形体 数据库 可为非土壤学专业的使用者提供土壤资源的有效信息。： 万和： 万的土壤数据库作为中国科学院科学数据库的重要组成部分 为实现国家级和世界级土壤资源信息网络化和共享提供了基础条件。为了实现土壤信息的标准化与规范化

年和 年分别制定和完善了国家标准《中国土壤分类与代码 》。目前 有关单位参照 的系列标准 正在制定我国的地理信息标准 而元数据标准是其中首要的部分。

近几年来 数字土壤信息建设正向着全面化和多元化的方向发展 在土壤分类、土壤质量动态监测、土壤环境监测、土壤科普和区域农田生态等数据库的研究和应用取得了重要成果。但是与 世纪 年代和 年代相比 虽然在内容和形式上都得到了扩充 但是信息共享的程度却没有本质的提高 仍然存在以下几个急需解决的问题：①土壤信息存储和表达的文件格式与数据结构不统一 阻碍了信息的共享与应用。我国的土壤历史数据丰富 需要统一的表达方式将历史资料与现在获取的数据融合起来 ④目前国内通常使用 中国土壤发生分类系统 和 中国土壤系统分类系统 两种不同的分类体系 而且国际交流中还会使用其他分类体系 已有的信息模型中大多没有提供体系间的参比类型的描述 因此需要新的土壤信息模型提供这种描述 实现类型转换 ④传统的土壤信息模式对非土壤学专业用户而言不易理解 对土壤学专业的研究人员而言 缺乏属性数据获取方式描述、测试方法记录和数据质量评价信息等 ④土壤信息具有“分布式”的特点 大量的资料保存在地方部门和各个科

国家自然科学基金项目

资助

作者简介 汪善勤

男 讲师 博士研究生 主要从事土壤信息与遥感方面的研究工作。

收稿日期

收到修改稿日期

研院所集中收集、整理和挖掘这些信息相当困难而建立分布式的土壤信息系统需要统一的信息模式。

针对上述问题 本文以中国土壤系统分类的检索系统为主体 参照中国土壤分类与代码标准、土壤质量标准等国家标准和 地理信息元数据标准 采用 可扩展的标记语言 为描述语言定义和规划元数据 建立土壤信息元数据模型

简称 。 是一种先进的描述网络信息资源的语言 可以方便地建立元数据系统。本文就 的结构、功能和 的描述方法进行了研究和探讨。

1 的结构体系

土壤的分类检索体系详细地描述了类型的命名、判定和层次划分 包含了各类土壤的鉴别特性 是最便于应用和认识的土壤知识体系。我国目前常

用的有 和 的检索系统 前者注重对成土过程的性状进行定性描述 后者注重对鉴别特性进行定量化描述 与前者相比有利于信息的表达和应用。因此 按照系统分类的检索系统构造主体框架 补充发生层与发生过程、数据获取、测试方法等信息。 的整体结构按照 地理信息元数据标准进行设计 划分为 个部分 目录信息、标准部分和引用部分 图 。

目录信息

用于标示和查询一个完整的土壤信息集和单个信息元素序列 提供信息集描述、序列编号和约束条件 如用户使用权限等。

标准部分

标识信息 描述土壤数据元素的基本信息 与目录信息中的序列编码对应。标示信息应能反映出所属数据集、遵循的国际标准、国家标准、行业标准、语言等 还应包括对数据采集环境、用户访问权限设置等方面的描述。

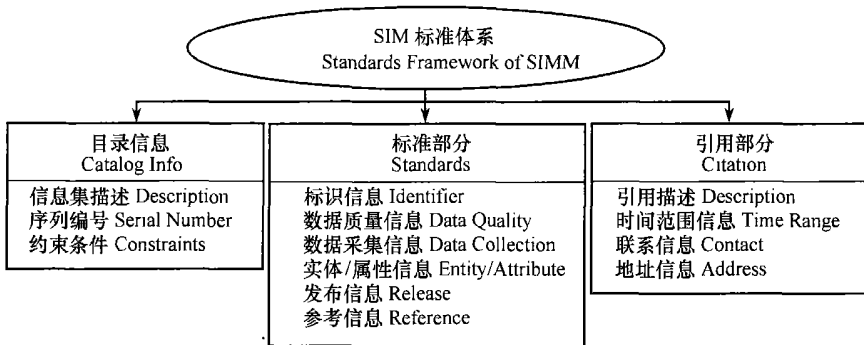


图 结构

④数据质量信息 记录数据集质量的评价信息。包括准确性、采集精度、时间和空间分辨率、比例尺、误差和不确定性等方面 根据各种应用要求 建立约束条件 用户可以借此检验数据的一致性和完备性 判断数据集的可用性。

④数据采集信息 描述数据采集的条件、工具、方法和过程等信息。便于使用者了解数据集建立的中间过程。例如描述获取土壤诊断层信息时 需要对使用的工具、分析方法、判断过程和应用目的等进行描述和记录。

¼实体和属性信息 是 的中心内容 由土壤信息元实体及其所包括的核心数据、特征信息和扩展信息实体组成信息模型。土壤信息元是对土壤信息的内容和数据进行描述的元数据 图 土壤

信息元实体则是指按照土壤信息元的定义 对一个具体数据集的描述 而得到的实例。

核心数据包括剖面描述和理化指标信息 是信息元实体的基本信息 具有强制性和通用型。 个土壤信息元实体包括多个剖面实例 记录多个理化指标数据。 个剖面描述对应 个样本信息 记录该样本在不同分类体系中的命名信息。

特征信息描述土壤类型的分布区域信息和主要鉴别特征 鉴别特征包括系统分类中的诊断层和诊断特性、发生分类中的发生层和发生过程 分别用实体表示。每个鉴别特征实体对应唯一的判定树实体 记录判定树编号和存储位置。使用判定树可以根据剖面实体记载的信息判断鉴别特征 用于确定土壤类型。每个剖面实体包括一个特征信息实体

对于不同时期获取的剖面数据 可能没有系统分类 或发生分类的鉴别特性信息 因此 一个特征信息实

体可能不包括任何鉴别特征实体 也可能包括多个 二者之间的关系用 对 来表示。

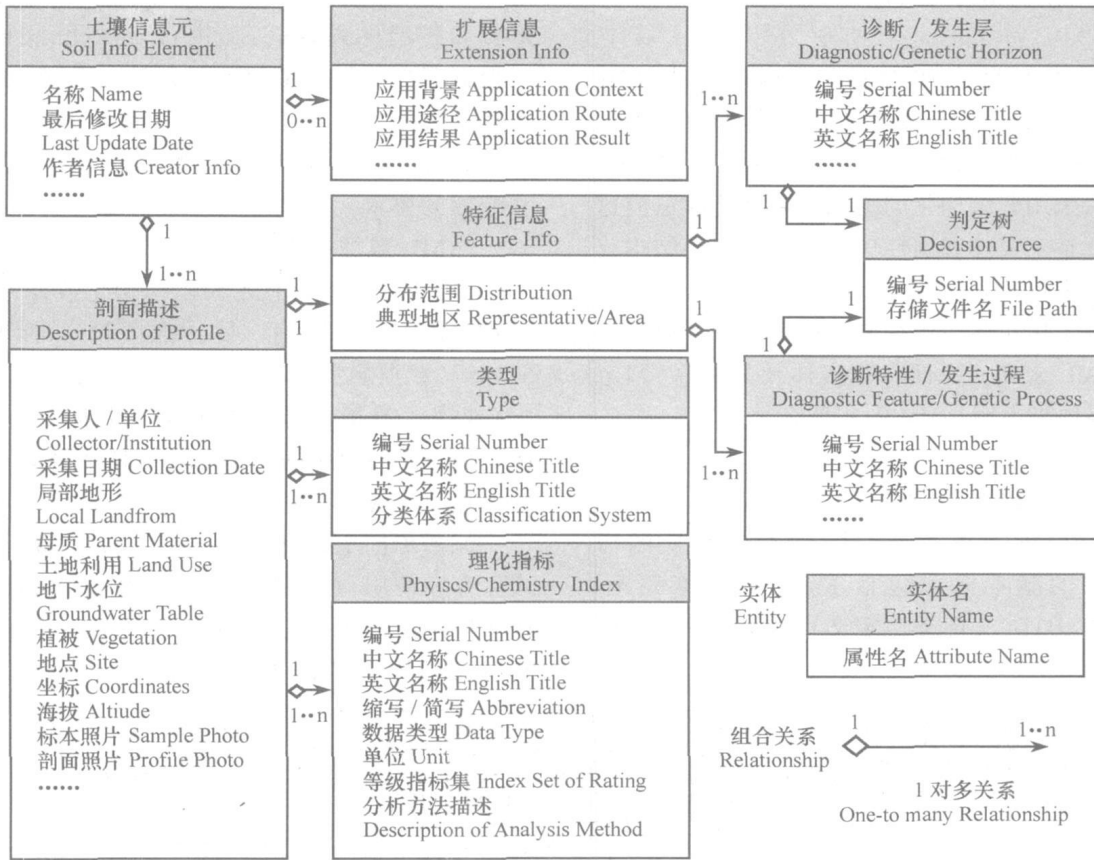


图 实体和属性信息

扩展信息记录信息的使用情况 是可选的 即每个信息元不一定包括扩展信息 可由用户根据需要添加。

发布信息 是关于数据集发行及获取方法的信息 包括发布时间、发布机构、使用数据集的技术参数等信息。通过发布信息 用户可以获得诸如数据集存储在何处、如何获取、获取介质等相关信息。

参考信息 包括数据集的安全信息、标准参照信息等 反映数据集的实时性等内容。

引用部分

其内容主要由使用者在使用时附加产生。统计数据集被引用的次数 记录引用人的相关信息、引用时间等内容。

2 的功能

提供土壤类型检索的基本功能。信息元中记录了每个剖面样本的类型信息 可以按照中国系

统分类和发生分类的命名检索样本。

能够按照 检索系统的检索顺序确定各级类别。 中的土壤类型判定树模型描述了鉴别性质 以及这些特性在确定土壤类型时的组合关系。对于一个未知类型的样本 按照检索顺序 将鉴别特性信息与判定树中的节点逐一匹配 确定土壤类型。

采用 描述数据和判定树 便于确定剖面的鉴别特征和类型。如果已确知一个剖面实体的 和 的类型 可以用来形成一条参比规则 如果只确知 或 的类型 则可使用鉴别特性判定树 根据实体信息和理化数据确定该剖面的诊断层 诊断特性或者发生层 发生过程 进而确定出归属 建立参比规则 。由于采用了 判定过程可以用现有的大多数开发语言实现。

为不同文件格式及数据结构存储和描述的土壤信息提供统一的数学模型。 继承了 的通用性、开放性和可扩展性的特点 不同用户和应

用程序也可以使用 转换和扩充数据格式。

3 的 实现

实体和属性信息是 的中心内容 也是 实现的重点。图 中各个实体自身是结构化的 但实体之间存在多种对应关系 它们的取值随实际情况的不同而不同 采用非结构化的组织方法更能够适应这些变化 如此整个实体和属性信息部分成为半结构化的。另外 判定树的形式多样 很难用统一的方式定义 因此在判定树实体中用存储文件名记录存放判定树实现的 文档。下面将就实体及其关系 判定树的 实现来说明 的实现途径。

3.1 XML 简介

是一种数据描述语言 语法简洁 具有很高的灵活度 已逐渐成为 上数据交换的标准 。 的基本概念包括 元素 、 属性 、 文档类型定义 、 模式和 命名空间 等 。 允许集成不同来

源的结构化数据 对分布式信息进行逻辑上的聚合和集成处理。 强调形式描述与内容描述的分 离 用 或 定义数据结构 而显示方式用 可扩展样式单语言 进行格式化 相同的数据 可根据用户的需要采用不同显示方式 使 数据 得到最大程度的重用。 具有通用性、开放性和 可扩展性的特点 能够满足各种行业对数据的组织 和处理的要求。

3.2 XML 数据的组织

描述的信息主要以结构化数据为主 它的 元数据内容非常明确。目录信息部分 标准部分中的 标识、数据质量、发布、参考 引用部分可直接引用 地理元数据标准中的方案。数据采集、实体和 属性信息是 的自有部分 需要根据土壤信息的 特点给出元数据定义。

元数据的 或 定义以 文档的形式保存和使用 每一个文档可以用层次结构视图描述文档中各元素的层次关系。下面是剖面描述的 文档的层次结构视图 图 。

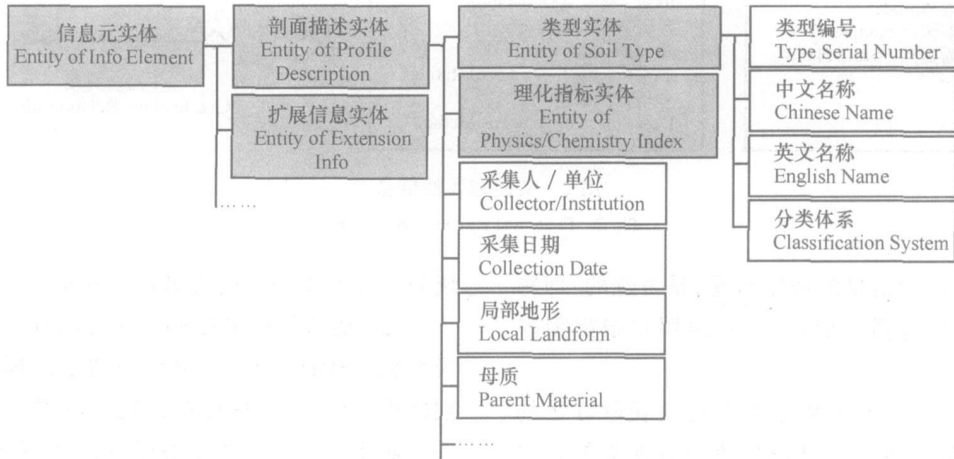


图 剖面描述的 文档的层次结构视图

实体之间的组合关系用数量约束表示 如信息 元实体包括一个或多个剖面描述实体 数量约束为 包括一个扩展信息实体 数量约束为 。每个 剖面描述实体在不同的分类体系中存在对应的类型

信息 并且至少有一种 因此类型实体的数量约束为 剖面描述实体包含一系列的理化指标信息 理化指标实体的数量约束也为 。下面是剖面 描述的部分 定义

”	”	”	”
”	”	”	”
”	”	”	”
”	”	”	”
”	”	”	”

” ”

” ” ” ” ” ”

” ” ” ”

” ”

” ” ” ” ” ”

” ” ” ”

” ”

” ” ” ” ” ”

根据上面的定义 泥碳质表层的一个数据实例可以表述如下

” 泥碳质表层”

” 是否符合水分饱和条件” ” ”

” 厚度” ” ”

” 水藓纤维体积比”

” ”

4 结 语

参 考 文 献

技术为网络信息应用带来了革命性的变化 土壤信息的表达与应用也将为之而改变。基于 是实现用于描述土壤信息的全新的网络信息模型 它针对土壤信息共享、组织、显示、管理和应用而设计。与传统的土壤信息模型相比 继承了 的开放性和可扩展性的特点 能够适应不同的应用目的和各种用户的要求。 用统一的语言表达了判定树 通过这种方式 能够很方便地从数据中解析出同一剖面的不同分类体系中的特征 形成参比规则。

是基于土壤系统分类的检索系统建立的模型 主要用于描述土壤的属性信息 不包括空间信息和时序信息。在进一步的研究中 将以 为基础建立时空数据模型 并且建立土壤信息交换 简称 的 语言 为土壤信息共享提供语言标准。

石元春 土壤学的数字化和信息化革命 土壤学报

赵其国 叶方 信息化与农业现代化 土壤学报

张学雷 张甘霖 龚子同 中国土壤系统分类土系数据库的建立、更新与应用 土壤

龚子同 张学雷 骆国保 等 的建立及其在世界上
的传播 地理科学
et al

龚子同 等编 中国土壤系统分类——理论·方法·实践 北
京 科学出版社 *et al*

et al

原立峰 谢永生 黄土高原土地信息系统数据库的建设 土
壤

史学正 于东升 “数字土壤”—— 世纪土壤学面临的机遇
与挑战 土壤通报
—

龚子同 张甘霖 陈志诚 等 以中国土壤系统分类为基础
的土壤参比 土壤通报
et al

周慧珍 中国土壤 信息共享研究——： 万中国土壤分布
式查询数据库 土壤学报

于东升 史学正 王洪杰 等 铁铝土的发生分类与系统分
类参比特征 地理学报
et al.

张定祥 潘贤章 史学正 等 中国： 万土壤数据库建设
中的几个问题 土壤通报
et al

中国土壤分类与代码

XML-BASED METADATA MODEL FOR RETRIEVAL SYSTEM OF CHINESE SOIL TAXONOMY

(College of Remote Sensing Information Engineering, Wuhan University, Wuhan , China)

(College of Resources and Environment, Huazhong Agricultural University, Wuhan , China)

Abstract

Key words