

基于 MCS 的空间元数据目录服务

付迎春¹, 袁修孝², 郭盛斌³, 张 过²

(1. 华南师范大学地理科学学院, 广州 510631; 2. 武汉大学遥感信息工程学院, 武汉 430079; 3. 广东南雄市国土局, 南雄 512400)

摘要: 针对元数据目录服务 MCS 对多源空间数据管理困难的特点, 该文提出一种新的分布式空间元数据目录服务 SMCS 的设计与实现方法。利用顾及空间关系的元数据分类与层级聚合方法, 构建具有双向推理机制的层级目录结构, 设计空间元数据的聚合操作和实现接口, 并基于分布式目录索引方法进行数据查找和定位。SMCS 的验证实验表明, 该方法能有效管理地理分布、动态异构的海量空间数据, 为网格环境下多源空间数据的高效集成应用服务提供技术支持。

关键词: 空间元数据目录服务; 空间关系; 聚合

MCS-based Spatial Metadata Catalog Service

FU Ying-chun¹, YUAN Xiu-xiao², GUO Sheng-bin³, ZHANG Guo²

(1. School of Geographical Science, South China Normal University, Guangzhou 510631; 2. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079; 3. Land Resources Bureau of Nanxiong, Nanxiong 512400)

【Abstract】 It is difficult for MCS to manage distributed multi-source spatial data. This paper introduces a new Spatial Metadata Catalog Services(SMCS). The spatial relation aggregation operations and catalog interfaces are defined. Thematic classical and nested aggregated collection referred to spatial relations are taken as governable method in the hierarchical catalog. And catalog index is created to find and locate distributed components which are efficiently proved to access, manage huge spatial data. The new application way enabled with SMCS can support efficiently run-time integration and share for huge multi-source spatial data under grid environment.

【Key words】 Spatial Metadata Catalog Service(SMCS); spatial relation; aggregation

地理空间目录(geographic data catalog)构成了空间数据框架基础设施(Spatial Data Infrastructure, SDI)的核心部分, 管理海量空间资源视图, 并发布数据^[1]。网格环境下基于元数据目录服务的多源空间数据管理和组织具有重要的意义。文献[2-3]分别利用轻量级目录访问协议(Lightweight Directory Access Protocol, LDAP)设计了网格资源信息服务, 美国物理网格GriPhyN(Grid Physics Network)开发出了元数据目录服务组件MCS, 但并未深入研究空间元数据的目录结构及服务体系。与空间分析一样, 空间数据的集成管理也需要深入分析空间元数据中的空间关系, 因此, 本文提出一种新的空间元数据目录服务的设计与实现方法。

1 网格元数据目录服务 MCS

地理空间元数据是属于以领域为中心的元数据, 描述如空间数据标识、空间参照信息和范围、遥感数据分辨率、平台等空间信息的元数据, 这些元数据中隐含着丰富的空间语义知识及空间关系。

MCS针对分布式环境中的元数据及其本身的描述性元数据, 提供注册、发布并根据数据属性值查询元数据目录^[4-5]。为了使不同用户进行一致的、透明的访问, MCS使用以下的逻辑数据模型:

其中, 逻辑数据项(Logical file, Data Item)是 MCS 最基本的数据项, 可以是文件和关系数据库表类的数据集实体, 以唯一逻辑名标识; 逻辑集合(Logical Collection)是容纳和管理逻辑文件或逻辑集合的实体; 逻辑视图(Logical View)由一个或多个逻辑文件、逻辑集合或视图组成。MCS 的逻辑数据结

构反映了存储网格节点的关系, 具有基于集合 Collection 的层级聚合机制。MCS 逻辑数据模型见图 1。

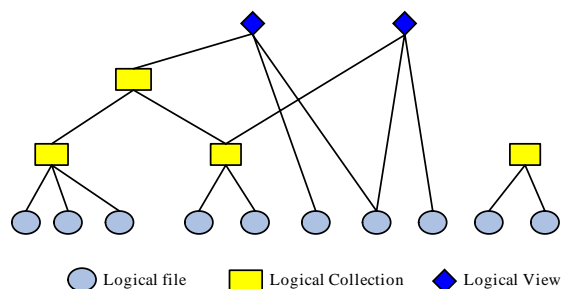


图 1 MCS 逻辑数据模型

MCS是独立的元数据目录中间件, 可以简单、灵活地与其他标准网格服务和自定义服务结合完成资源的管理^[4-5]。与网格平台资源管理部件MDS相比, 不需要很多的网格硬件、操作系统和编译环境支持; 与轻量级目录访问协议LDAP相比, MCS具有更为灵活简单的目录层级结构和容易操作的目录接口; 而与传统元数据目录相比, 提供网格开放标准的查

基金项目: 国家“863”计划基金资助项目(2003AA132080); 国家自然科学基金资助项目(40601084)

作者简介: 付迎春(1976 -), 女, 讲师、博士, 主研方向: 地理信息理论及应用研究, 网络 GIS; 袁修孝, 教授、博士; 郭盛斌, 工程师、学士; 张 过, 副教授、博士

收稿日期: 2007-07-08 E-mail: fyc266@163.com

询、数据发布和视图接口，容易进行扩展。因此，本文基于MCS组件进行扩展，设计空间元数据的目录结构和接口来管理地理分布、动态异构的空间元数据。

2 空间元数据层次目录结构

2.1 逻辑层次对象模型

本文将在 MCS 的元数据逻辑描述结构中考虑空间元数据的空间关系，从物理层、资源层和视图层 3 个层次描述空间元数据间的约束、映射与聚合关系，而每一层中的元数据进行聚类，再进一步嵌套聚合。层级元数据描述方案见图 2。

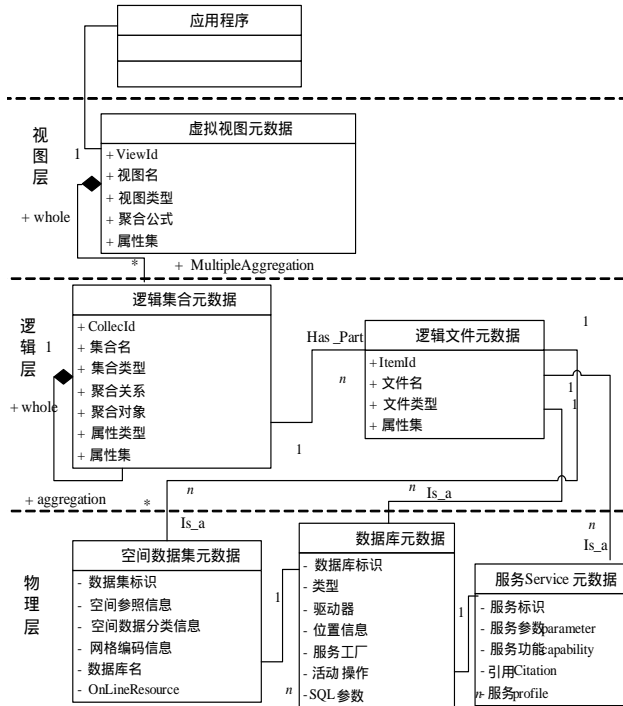


图 2 基于 MCS 的空间元数据层次信息模型

其中，物理层为最低层，主要是针对分布在不同地理位置上的空间数据，及各种应用相关信息集合(见图 2)，在物理层基础上抽象出逻辑资源层，形成空间数据聚合基础，其元数据主要包括两个方面：

(1)Item level metadata：描述逻辑层与物理层对象间映射关系,表现物理层各类元数据信息与数据项 Data Item 之间的映射。

(2)Collection level metadata：描述集合自身和相关的信息，又可以分为两个方面：

1)集合公共元数据(collec-common metadata)，描述集合标识、内容，及集合中所有对象的公共元数据，例如 TIFF 格式、WGS84 参考系等,其均为该集合公共元属性、减少 Item 项的冗余元数据信息。

2)集合相关元数据(Collec-Specific metadata)：集合特有的，描述集合与对象，集合与集合间聚合和约束关系，聚合后属性类型等的元数据。

除了 MCS 自身提供的元数据外，扩展的 MCS 集合 Collection 的元数据字段及其详细取值范围如表 1 所示。图 2 表示的元数据层级信息充分描述了从物理层到逻辑层、视图层和应用程序的一种自下而上的层次体系，可以通过不同层元数据之间的转换和关联把资源实现的具体细节封装起来，使得应用程序可以按照自己熟悉的分类和组合方式对资源进行访问。

表 1 逻辑层集合元数据

元素	类型	取值范围
CollecId	集合标识	逻辑集合全局的唯一标识 ID
CollecName	集合名	逻辑集合名称
CollecType	集合类型	聚合对象所属类型，如空间、时间、时空、服务等
Agg_Relation	聚合关系	逻辑对象的交{generalization}并{extension}、差{substraction}
Agg_ReferingId	聚合对象	所聚合对象集合的 Id 标识
Agg_AttrType	属性类型	集合属性类型包括 Item level Common, Collec-Specific
Agg_Attribute	属性集	逻辑对象集合的元数据属性

2.2 空间元数据目录的双向聚合/推理机制

传统 MCS 通过记录 Collection 的 ID 与 Item 子项 ID 间的映射关系构建层级目录，因此，其缺乏处理复杂空间关系的能力，存在空间元数据冗余，备份复制和查询耗时的问题。

基于 MCS 并顾及空间关系的元数据目录构建机制应该考虑：(1)按照空间数据特征和专题聚合成 Collection，形成水平分类。(2)按知识语义关系形成纵向的层级结构，形成诸如空间索引 R 树的描述空间关系的元数据目录树，最低层叶节点 Item 项包含无冗余的元数据单元。其中的空间关系主要指空间范围上的包含、相交、相邻和方向顺序关系等，而知识语义则主要指自身知识内容的相互关系包括隶属(is-a)、聚合(part-of)和实例(instance-of)等关系。

元数据的聚合操作就是建立层级关系的关键，引入元数据的抽象数据结构^[1]，假设 a, b 是元数据记录， c 是由 a, b 经过交、并、差运算产生的新记录，引入以下定理说明：

定理 1 交运算(generalization) 取两个元数据记录 a, b 的交集即公共属性生成新元数据记录，即 $c = a \cap b$ 。

定理 2 并运算(extension) 取两个元数据记录 a, b 的并集生成新元数据记录，即 $c = a \cup b$ 。

定理 3 差运算(substraction) 取两个元数据记录 a, b 的差集，或指从元数据记录 a 中排除与 b 的共有属性，生成新的元数据记录 $c = a - b$ 。

元数据元素和记录的交、并、差可针对上述的非空间属性和空间范围属性执行操作。

图 3 简单示例对比了传统模式和聚合模式下集合 Collection 与数据项 Item 效果和双向的聚合方法。

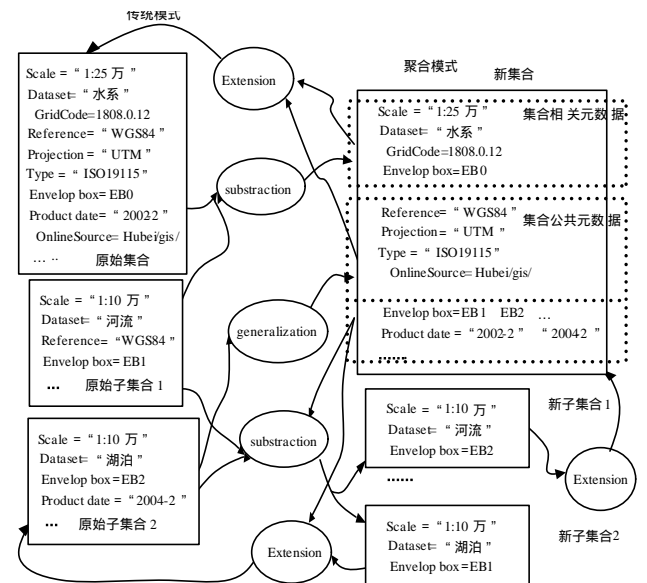


图 3 空间元数据聚合示意图

完成的操作分为：(1)原始集合与原始子集合求差(substraction)可得到新集合的相关元数据，求交(generalization)可得到新集合的公共元数据；原始子集合与新集合求差(substraction)可以得到新子集合；(2)反之，新集合求并(Extension)可以逆向聚合得到原始集合，而新集合和新子集合求并(Extension)，可以逆向还原得到原始子集合。

显然，聚合操作降低了数据冗余并形成有效的元数据分类和推理机制。

2.3 分布式空间元数据目录的索引命名和定位

除了空间关系，海量空间数据还具有地理分布、动态异构的特点，因此，分布式空间元数据目录必须具备较强的命名和定位功能，具体而言就是能够整合分布式目录提供完整的元数据视图，以供快速的空间查询。文献[2-3]研究了将LDAP各个分支目录节点和目录子树对象结合形成倒转的树型层次结构——目录信息树(Directory Information Tree, DIT)，由每个节点有的区分命名 Rdn 组成该节点的完整区分名 Dn，提供目录中定位的命名依据。而从图 1 可以看出，面向对象的 MCS 逻辑数据模型与 LDAP 同样具有良好的松散耦合结构和分布式管理的框架，但 MCS 并未提供额外的目录命名功能，而是将目录的映射和包含关系定义在逻辑对象结构中。鉴于复杂的空间关系和约束映射关系，MCS 原有的目录定位功能是不够的。因此，本文提出针对分布式空间元数据目录建立目录索引结构(见图 4)。

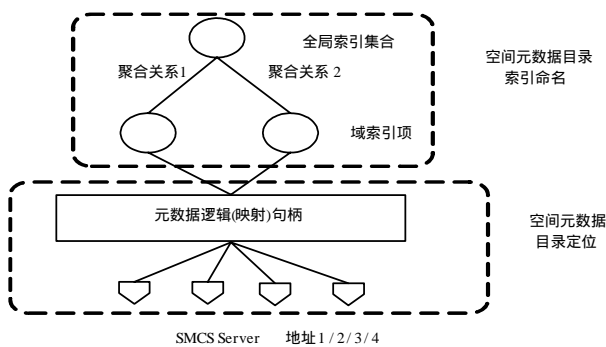


图 4 分布式 SMCS 索引结构

分布式 SMCS 索引的原理是：对各节点元数据对象取并集，构建全局索引集合项，对各节点的元数据对象依空间关系取交集和差集，构建层级域索引项，取全局和域索引对象的一个专题分类属性命名，并记录各索引节点间的聚合关系；同时将抽取各子节点上映射的物理地址，建立目录索引中各域与底层的逻辑(映射)句柄。当需要元数据目录服务查找时，先粗略定位到目录索引中的域，缩小查找的范围，进一步确定并返回属性所属的域，最后可按域的逻辑(映射)句柄去获取实际的地址和数据。因此是一种针对不同空间关系数据聚合与目录定位的有效方法。

3 空间元数据目录服务接口设计与实现

本文将各逻辑对象及其关系封装成图 5 所示的接口对象，元数据目录和目录索引服务的各项功能通过这些对象来实现。其中，PhyMetaData 是物理层元数据对象；mcsRegistryObject 是逻辑层注册的抽象基类，实现注册对象与逻辑资源层对象间的映射操作。mcsLgFile, mcsLgCollec, mcsLgView 对应文件、集合和视图 3 类对象，聚合对象 Agg_Operations 主要操作有：SetAggOperator 设置聚合运算算子；GetAgg_AttrType 得到集合的属性类型，如 Item level、Collec

level、View level metadata；GetReferringId 访问参与运算的逻辑对象 ID；GetAttribute 访问聚合对象属性。

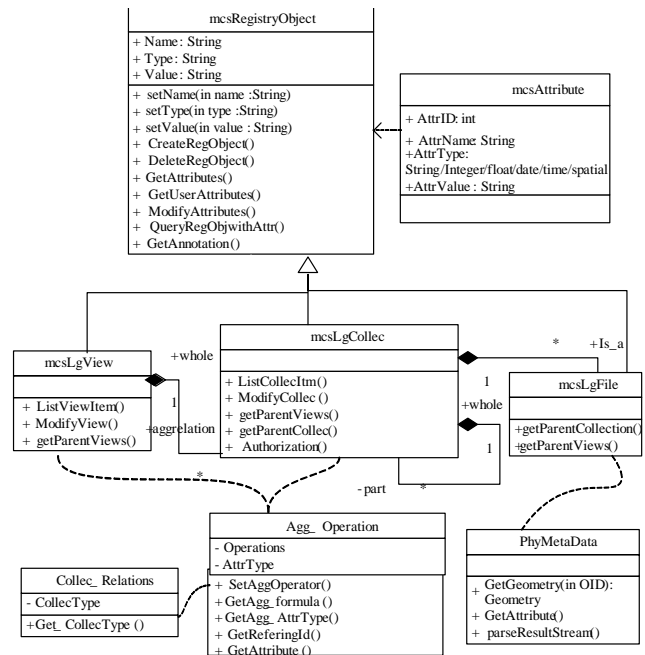


图 5 空间元数据目录层次对象模型及接口

服务端 SMCS 组件基于 Java 与 OGSA 的软件参考实现版本 GT3.2 编写，每个元数据服务类都将从 GridServiceImpl 类继承，实现基本的网格元数据的目录功能，如注册、查找、发布等，同时也需要实现相应服务的 PortType。在服务发布时，把服务的存根代理和 PortType 提供给客户，以进行服务的访问。采用数据访问与集成组件 OGSA-DAI 访问异构数据集和资源，开源数据库 MySQL 实现元数据中心的数据存储管理。

4 结束语

针对 MCS 管理空间元数据困难的特点，通过详细分析空间元数据的聚合关系和具体的交、并、差聚合操作，定义了合理的空间元数据分类和纵向聚合的逻辑描述对象和层次结构，在此基础上设计的空间元数据目录和目录索引的操作接口，能满足将来网格技术环境中对海量分布空间数据集成管理的要求，包括空间信息的动态获取、更新、快速查询、发布等。为网格环境下多源异构空间数据的实时集成提供技术支持。

参考文献

- [1] Noguera J. Geographic Information Metadata for Spatial Data Infrastructures[M]. Heidelberg, Berlin: Springer-Verlag, 2005: 31-125.
- [2] 肖 依, 任 浩, 徐志伟. 基于资源目录技术的网格系统软件设计与实现[J]. 计算机研究与发展, 2002, 39(8): 902-905.
- [3] 王 威, 刘卫东, 宋佳兴. 基于 Globe 的分布式 LDAP 元数据目录服务框架[J]. 2006, 32(11): 66-69.
- [4] Zhao Peisheng. Grid Metadata Catalog Service-based OGC Web Registry Service[C]//Proceedings of the 12th annual ACM International Workshop on Geographic Information Systems. [S. l.]: ACM Press, 2004
- [5] Deelman E. Grid-Based Metadata Services[C]//Proc. of the 16th International Conference on Scientific and Statistical Database Management. Santorini Island, Greece:[s. n.], 2004.