

分布式海量遥感影像编目检索机制

谢冰川, 陈 华, 赵 亮, 李莎莎

(国防科学技术大学电子科学与工程学院, 长沙 410073)

摘 要: 为了高效地组织、检索分布式海量遥感数据, 研究 FGDC 制定的遥感元数据标准草案, 提出一种具有良好可扩展性的遥感编目数据模型。基于该编目数据模型, 提出一种基于编目数据的分布式影像检索算法, 设计并实现一个遥感数据编目检索系统, 实验结果表明, 该系统是有效的。

关键词: 海量遥感影像; 编目; 数据模型; 分布式; 检索

Mechanism of Cataloging and Retrieval over Distributed Massive Remote Sensing Image

XIE Bing-chuan, CHEN Luo, ZHAO Liang, LI Sha-sha

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

【Abstract】 Focusing on the problem of effectively organize and retrieve the distributed massive remote sensing images, this paper presents a well extensible remote sensing catalogue data model based on the remote metadata standard draft which is proposed by the Federal Geographical Data Committee(FGDC). A distributed remote sensing images retrieval algorithm is designed based on the proposed catalogue data model. It accomplishes a remote sensing data catalogue and retrieval system which is successfully applied to a massive data management system.

【Key words】 massive remote sensing image; catalogue; data model; distributed; retrieval

1 概述

随着遥感技术的发展, 各种新型的传感器不断涌现, 遥感数据的类型、级别不断丰富, 呈现出多格式、多尺度、海量、分布存储的特征^[1]。分散存储于各个机构的不同种类的遥感数据在数据共享和使用上存在着“数据孤岛”现象。如何消除这些“数据孤岛”, 为不同领域的遥感数据用户提供快速、准确、方便的数据检索与浏览手段, 已经成为本领域的热点问题。遥感影像编目数据的建立为解决这些问题提供了有效的工具。

遥感数据编目是将遥感数据依照一定的规则加以系统的组织与整理, 将之编辑成目录, 供人们进行数据的检索^[2]。美国联邦地理数据委员会(FGDC)在 2002 年通过对原有 CSDGM 标准的扩展, 完成了遥感元数据标准草案的制定, 达到了提供遥感数据辅助信息的目的^[3]。标准草案以松散相关的层次状结构化信息提供了对遥感影像的通用性描述。目前针对元数据的描述模型大多采用关系图的方式, 使用关系表之间的关联关系描述元数据内部松散相关的层次状结构, 而关系数据表在存储松散相关的层次状结构化信息方面已经被证明其扩展性较差^[4], 不能便捷地加入新型数据字段。

本文基于 FGDC 制定的遥感元数据标准草案, 设计了一种扩展性良好的遥感影像编目模型, 提出了一种分布式遥感影像编目检索算法, 并应用于某海量空间数据管理系统中。

2 遥感影像数据编目模型

2.1 遥感影像编目数据模型

通过对 FGDC 制定的遥感元数据标准草案的研究, 针对实际应用需求, 提取核心元数据集 CoreMetaDataset, 设计如图 1 所示的遥感编目数据模型。

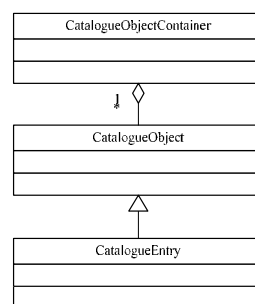


图 1 遥感编目数据模型 UML 描述

相关定义如下:

定义 1 CoreMetaDataset(核心元数据集)

CoreMetaDataset=(CategoryID, DatasetName, DatasourceName, MBR, SpatialResolutionX, SpatialResolutionY, SpatialReference, StartDate, EndDate, Time)

其中, CategoryID 表示遥感影像分类标识; DatasetName 和 DatasourceName 标识数据源中的遥感影像数据; MBR 表示遥感影像覆盖的地理区域的最小包围矩形框; SpatialResolution X 表示遥感影像 X 轴方向上的空间分辨率; SpatialResolution Y 表示遥感影像 Y 轴方向上的空间分辨率; SpatialReference 描述遥感影像的参考坐标系信息; StartDate, Time 表示遥感影像

基金项目: 国家自然科学基金资助项目(40601080); 国家“863”计划基金资助项目(2007AA12 Z208, 2008AA12A211)

作者简介: 谢冰川(1986-), 男, 硕士研究生, 主研方向: 地理信息系统, 数据库技术; 陈 华, 副教授、博士; 赵 亮, 博士研究生; 李莎莎, 硕士研究生

收稿日期: 2010-01-05 **E-mail:** nudtxbc@gmail.com

的成像开始时间; EndDateTime 表示遥感影像的成像结束时间。

定义 2 CatalogueObjectContainer(编目条目容器)

CatalogueObjectContainer=(CatalogueObject1,CatalogueObject2,
...,CatalogueObjectm)($m \geq 1$)

表示编目条目的聚集,其中,CatalogueObjectm 表示编目条目对象。

定义 3 CatalogueObject(编目条目对象)

CatalogueObject=(CoreMetaDataset)

表示编目条目对象,其中,CoreMetaDataset 为定义 1 中定义的核心元数据集。

定义 4 CatalogueEntry(编目条目)

CatalogueEntry=(CoreMetaDataset, SpecificMetaDataset)

表示具体的编目条目对象。其中,SpecificMetaDataset 表示特有的元数据集信息。

从模型结构上分析,该数据模型结合面向对象的思想,采用继承的机制实现对核心元数据集的扩展,理想情况下可以管理任何类型的遥感数据,因此,对于将来可能出现的新类型影像数据具有良好的可扩展性。

2.2 遥感影像编目数据存储组织

基于遥感影像编目模型及其相关信息,设计如图 2 所示遥感影像编目数据逻辑结构,关系实体 Pro_Dataset_Category 描述了遥感影像的分类信息,通过该信息可以获取遥感影像数据及所属的编目分类,其中 Catalogue_TableName 描述了保存编目分类结构的关系实体的名称,CategoryID 是在指定的 Catalogue_TableName 中描述某项编目分类的唯一标识;关系实体 Pro_DataSource_Params 描述了各类型数据源的存储参数信息,其中 Storage_Type 描述支持的数据源类型,包括文件系统、空间数据库以及空间数据中间件等 3 种方式,Connect_Params 中描述了各分布节点影像数据库的连接参数信息;关系实体 Pro_Spatial_Metadata 描述了了遥感影像的元数据信息,其中,SpecificMetadata 描述了不同类型遥感影像的特有元数据信息。

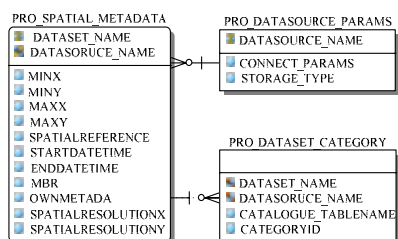


图 2 遥感编目数据逻辑结构

3 基于编目数据的分布式遥感影像检索

基于本文提出的遥感编目数据模型,设计全局编目数据库,存储各分布节点存储的遥感影像的编目数据信息以及相应的存储参数信息,提供统一的查询视图。根据存储网络中不同节点承担角色的不同,对网络中的节点划分为 2 种角色,分别为 Manage-Node、Storage-Node。

定义 5 Manage-Node(管理节点)

该节点表示存储网络的中心节点,不但具有存储遥感影像数据集的能力,而且承担管理整个存储网络中遥感影像编目数据的能力。

定义 6 Storage-Node(存储节点)

该节点表示存储网络中影像存储节点,只具有存储遥感影像数据集的能力。

用户请求遥感影像时,首先向 Manage-Node 提交查询请求,Manage-Node 组织查询,获取符合用户查询请求的编目

条目,获取关联的存储参数请求;然后由 Manager-Node 组织数据请求,根据获取的存储参数信息,向存储有符合用户查询请求的遥感影像的 Storage-Node 发送数据请求,并将返回的遥感影像提交给查询用户。

基于上述存储网络结构,设计一种基于编目数据的面向分布式遥感影像数据的检索算法,该算法对应的伪代码如下:

算法 Distributed-RSImageRetrieval

输入 待查询空间区域窗口 QW

输出 待查询空间区域覆盖的遥感影像

BEGIN

A=CatalogueDataQuery(QW);

//数组 A 表示与查询窗口相交的遥感影像的编目数据的集合

O=CreateOutputDataset(QW);

//O 表示符合查询条件的输出影像数据集

for $j \leftarrow 1$ to length[A]

Temp=GetImageDataset(A[j]);

//Temp 表示编目数据条目 A[j] 所代表的遥感影像

O.Insert(Temp);

//将影像数据集 Temp 插入到输出影像数据集 O 的合适的位置

O.FlushCache();

END

上述算法基于遥感影像编目数据,隐藏底层的数据存储细节,并提供了统一的数据视图,从而能够较好地完成分布式的遥感影像检索。

4 系统实现

4.1 编目系统设计

设计如图 3 所示编目检索系统,系统基于 C/S 架构开发,逻辑上分为 3 个层次:数据层,服务层以及应用层。在服务层,提供了遥感数据编目服务以及检索服务。基于全局编目数据库,构建遥感影像统一查询视图,屏蔽遥感影像的底层存储细节,完成遥感影像的检索。

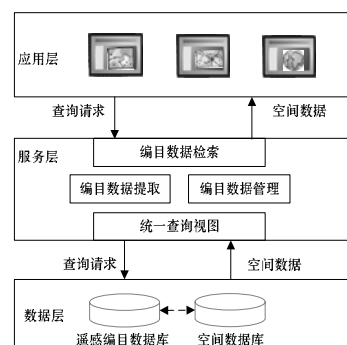


图 3 编目检索系统逻辑结构

在具体存储方案中,根据图 2 提出的编目数据逻辑模型,针对编目数据模型中提出的核心元数据集各数据项,采用数据表属性字段进行描述,其中对于 MBR 属性项使用 Oracle 提供的 Geometry 数据类型进行描述;对于特有元数据 SpecificMetadata 使用 Oracle XML Type 进行描述。

4.2 分布式遥感影像检索

基于 Visual Studio 2008 平台,实现了分布式遥感影像检索算法,完成了面向空间区域范围的分布式遥感影像检索。图 4 描述的查询结果表示查询窗口覆盖的遥感影像数据集分布在多个数据存储节点中。从用户角度来看,检索结果并无大的差别。系统实现了使用统一编目模型屏蔽数据分布存储底层细节的目标,支持分布式节点的扩展。

(下转第 285 页)