

遥感影像数据库引擎设计与实现

刘 鹏^{1,2}, 毕建涛¹, 曹彦荣¹, 何建邦¹

(1 中国科学院地理科学与资源研究所, 北京 100101; 2 山东科技大学信息科学与工程学院, 青岛 266510)

摘要:在参考国内外现有资料的基础上, 提出一个遥感影像元数据标准草案, 并利用该元数据标准对遥感影像数据进行描述; 用大型关系数据库对遥感影像数据及其元数据进行一体化管理, 实现遥感影像数据的结构化存储; 使用 XML 对遥感影像元数据进行描述, 便于实现影像元数据的网上查询与检索, 并最终实现遥感影像数据的网络共享。

关键词:元数据; 遥感影像管理; RDBMS; XML; 本体

中图分类号:P208

1 引言

随着当前空间信息技术的迅猛发展, 以空间数据为主的空间信息挖掘, 特别是对于遥感影像数据, 由于获取方便、周期短、信息量大等特点而成为空间数据的重要组成部分。然而, 由于遥感数据的数据量十分庞大, 特别是对于具有不同来源、不同分辨率与不同时相的数据, 其存储与管理均十分困难, 从而限制了数据的充分利用, 迫切要求对遥感影像数据进行有效的组织、存储、管理和共享研究。

基于上述目的, 本文提出了遥感影像元数据标准草案, 构建了影像数据的元数据库; 利用大型关系数据库(RDBMS)对遥感影像数据及其元数据进行组织与管理; 通过基于 XML 的影像元数据的发布, 实现用户通过网络对遥感影像数据的查询、检索与访问, 并为实现影像数据的共享奠定基础, 同时通过对信息本体的研究, 提出了建立遥感影像信息本体, 利用本体技术的优势, 以实现影像数据的互操作。

2 基于元数据的遥感影像数据管理与网络发布模式

元数据可以使数据生产者完整地说明数据, 从而帮助用户了解数据的基本特征, 更好地访问、评

价、购买和利用数据, 从而促进了数据的交流与共享。遥感影像数据集建立相应的元数据, 便于用户及时了解影像数据的相关信息, 同时利用元数据的导航服务, 快速定位感兴趣的影像数据, 以加强数据的交流与共享, 提高数据的利用程度与附加值。

对于影像数据目前存储方式, 我们采用基于 RDBMS 的影像数据存储与管理模式, 即利用大型关系数据库(RDBMS)作为后台支持, 将各种遥感影像数据经过相应地分割、压缩与重采样处理后, 直接存储到大型关系数据库中。

在影像数据入库的同时, 我们把影像数据对应的元数据也存入数据库中, 并建立影像数据与元数据的一一对应关系, 实现影像数据与元数据的一体化管理。

XML(eXtensible Markup Language)是国际组织 W3C (World Wide Web Consortium) 为适应 WWW 的应用, 将 SGML (Standard Generalized Markup Language) 标准进行简化形成的标记语言。基于 XML 的数据表示可以很好的在不同系统和平台间进行交互, 能很好地满足在网上描述、传输和显示元数据的需要。利用 XML 技术描述遥感影像元数据, 用户可以通过网络对元数据进行查询、检索, 了解遥感影像数据的分辨率、传感器信息、获取时间、光谱信息、覆盖范围、获取方式等信息, 并可以通过元数据导航服务直接浏览相应的影像数据。

收稿日期: 2005-02-19.

资助项目: 本研究得到国家“十五”科技攻关计划“中国可持续发展信息共享研制与开发”项目资助。

作者简介: 刘鹏(1981-), 男, 硕士, 山东淄博人, 主要方向为并行与分布式处理, GIS 平台软件开发。E-mail: lphpc@163.com

3 遥感影像元数据存储模型、库结构及其元数据描述

本文提出的遥感影像元数据标准草案是在参考了 ISO 19115.3 遥感影像元数据标准以及即将推出的我国地理信息元数据标准的基础上,结合项目的实际情况所制定的遥感影像元数据草案。该草案包括 7 个元数据集、6 个公共数据类型和 15 个代码表,从标识信息、数据质量信息、参照系信息、内容信息、覆盖范围、分发信息、遥感信息等方面对遥感影像数据进行了详细的表述。

7 个元数据集:MD_元数据定义有关各种地理信息资源元数据的根实体;数据集标识信息描述了遥感影像数据集的基本信息;数据质量信息提供遥感影像数据质量的整体评价信息;参照系信息提供遥感影像数据使用的空间和时间参照系的说明;内容信息提供要素分类信息并说明数据层及影像数据特征;分发信息提供获取时间资源所需要的分发者和选项的信息;遥感信息提供获取遥感影像数据的平台信息、传感器信息、传感器口径、辐射校正、数据志信息以及遥感影像产品的相关属性信息。

6 个公共数据类型包括:覆盖范围信息指有关对象的空间和时间覆盖范围;负责单位信息指数据集负责人与单位识别和联系方式;联系信息指与负责人或负责单位联系所需信息;地址信息指负责人或负责单位地址;日期信息指参照日期及说明该日期的事项;在线资源信息指可以获取数据集、规范、共用的领域专用标准名称和扩展的元数据元素的在线资源信息;

15 个代码表包括:CI_日期类型代码、CI_角色代码、MD_字符集代码、MD_安全限制分级代码、MD_数据层内容类型代码、MD_进展代码、MD_限制代码、MD_空间表示类型代码、MD_专题类型代码、SC_大地坐标参照系统、SC_垂向坐标参照系统、MD_格网单元几何类型代码、MD_维名称类型代码、MD_影像条件代码、MD_像元定位代码。

对于该标准各数据集之间的逻辑结构和关系系采用 UML 静态结构类图来表示,如图 1 所示。

3.1 遥感影像数据存储模型

由于遥感影像的数据量十分庞大,直接以整幅图像为单位进行存储,不利于影像数据的后续处

理、提取、浏览与检索。故需要对影像数据进行存储前期预处理。主要包括采样、影像压缩与影像分割等内容。

影像分割是将遥感影像按照行列值将其分割为相同大小的数据块(tile),以 tile 为影像存储的基本单元,每个 tile 均以一条记录的方式进行存储,不同记录通过编号进行排列。对于不能够平分的,出现多余的行或列时,应将其单独存放。当用户对影像进行调用时,通过映射关系,只调用与用户有关的 tile 集合即可,从而优化了数据的存储、传输、浏览模式。

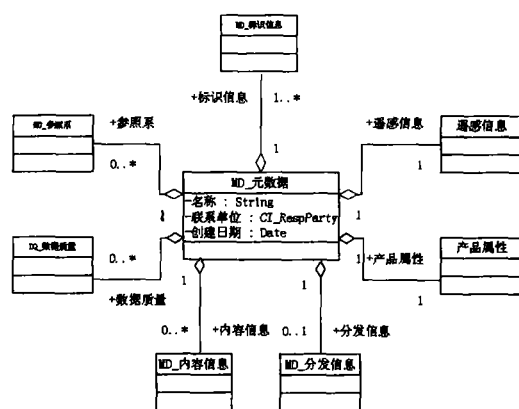


图 1 遥感影像元数据标准 UML 类图

Fig.1 Class diagram for remote sensing image metadata standard

为减小影像的传输数据量和优化显示性能,需建立影像金字塔,通过影像采样方法,建立一系列的不同分辨率的影像图层,每个图层分割存储,并建立相应的空间索引机制。常用的影像重采样方法有双线性差值、立方卷积等。

由于影像的数据量比较庞大,为减小影像的存储空间,还需要对影像进行压缩处理,然后进行存储。当用户调用数据时,首先对数据进行解压缩处理,然后再返回给用户。常用的图像压缩方法有 JPEG、LZ77 等。

遥感影像数据存储模型如图 2 所示。

3.2 影像数据库结构设计

遥感影像数据库主要可以分为影像元数据库和影像数据库两部分,如图 3 所示。影像元数据库用于对遥感影像元数据标准中的数据集进行存储与管理。影像数据库用于对影像数据进行存储和管

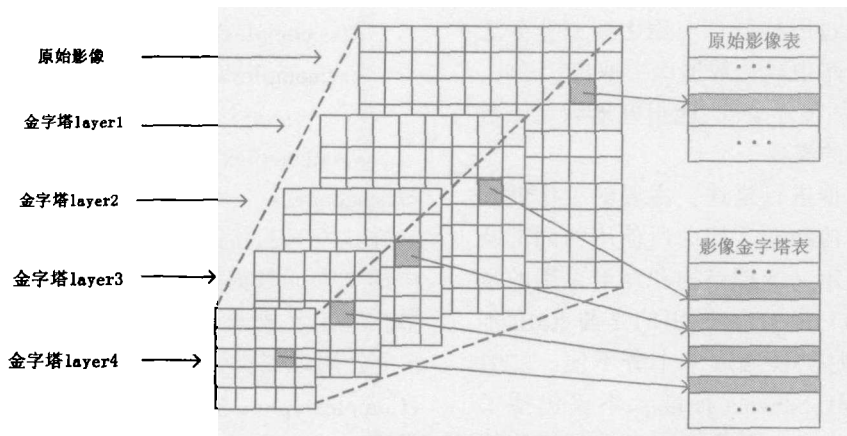


图 2 遥感影像数据存储模型图

Fig.2 The storage model diagram for remote sensing image

理。对于每一景加载的影像我们为其提供一个 ID 号,该 ID 号在系统中唯一标识该景影像,我们称之为影像 ID。通过影像 ID 把每景的影像数据与元数据关联起来。

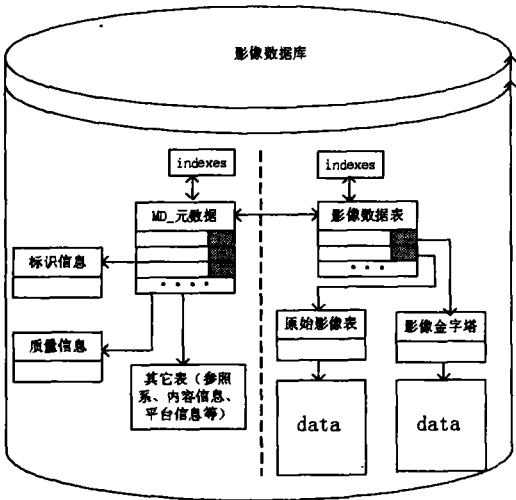


图 3 遥感影像数据库结构图

Fig.3 The structure scheme for remote sensing image database

元数据的组织应该清晰地映射出元数据 UML 类图(图 1)中各个类及其之间的关系。对于 UML 类图与关系数据库之间的映射,本文采取以下策略:将类映射成表,将类的属性映射成表的字段,通过在表中建立主键和外键来实现类之间的关联。这样,标识信息、质量信息、参照系信息、内容信息、分发信息、遥感信息等数据集分别映射成表,并为表中每条记录都设置一个 ID (记录号),用作主键。MD_元数据除了包含自己的元数据项外还用于标识影像的影像 ID 和用于与标识信息等 6 个数据集建立关联的外键。这样对于加载的每一景影像通过

MD_元数据就可以把该景影像的所有元数据集组织起来。

遥感影像的组织应该清晰地体现遥感影像数据模型的组织策略,这主要是通过影像数据表、原始影像表、影像金字塔表来实现的。原始影像分割后各 tile 块存储在原始影像表中,每个 tile 一条记录,同时在该记录中记下每个 tile 块的影像 ID、行号、列号等相关参数,以备提取数据时进行检索和访问。对于金字塔数据的 tile 块存储在影像金字塔表中,每个 tile 一条记录,并在该记录中记下每个 tile 的金字塔层号、行号、列号等相关参数。每加载一景影像就在影像数据表添加一条记录,该记录除包含唯一标识影像的影像 ID 外,还用于与原始影像表、影像金字塔表建立关联的外键。从而通过影像数据表就可以把每景影像的原始影像和影像金字塔组织起来。

对于每景加载影像,只要给出了它在系统中的影像 ID,通过 MD_元数据就可以得到该景影像对应的元数据,通过影像数据表就可以得到原始影像数据和影像金字塔数据,从而实现了影像数据及其元数据的一体化管理。

3.3 基于 XML 的遥感影像元数据描述

由于遥感影像元数据数据量大、结构复杂,要在网上传输、共享、交换元数据,数据库是不能满足需求的。XML 是一种元语言,是用于描述其他语言的语言,用户可以根据需要,利用 XML Schema(或者 DTD)自行定义标记和属性,从而可以在 XML 文件中描述并封装数据。XML 是数据驱动的,这使得数据内容与显示相分离。XML 可以在类似于 Netscape Navigator 或 Microsoft Internet Explorer 的

浏览器中显示,并通过因特网在应用之间或业务之间交换,存储到数据库中或从数据库中取出。因此,XML 是元数据最好的描述方式,能很好地满足元数据在网上传输、交换的需要。

用 XML 对元数据进行描述,主要的工作就是定义 XML 文档的结构和在文档中所使用的词汇及其数据类型。最初,定义文档结构和词汇采用的是文档类型声明(DTD),但 DTD 采用的是与 XML 不同的语法,数据类型少,使用起来十分不便。2001 年,W3C 提出了 XML Schema 标准,不仅包括了 DTD 能实现的所有功能,而且它本身就是规范的 XML 文档。最重要的是,它提供了一系列新特色,可以弥补 DTD 的不足,比如:丰富的数据类型、用户可以自定义数据类型、属性分组、Namespace 的支持等。为此,本文采用 XML Schema 来定义标识。遥感影像元数据 XML Schema 如下(主要以 MD_ 元数据为例,其他以“.....”省略):

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs = "http://www.w3.org/2001/
XMLSchema" elementFormDefault = "qualified" attributeFormDefault="unqualified">
  <xs:element name="MD_ 元数据">
    <xs:annotation>
      <xs:documentation> 遥感影像元数据</xs:documentation>
    </xs:annotation>
    <xs:complexType>
      <xs:sequence>
        <xs:element name="文件标识符" type="xs:string"/>
        <xs:element name=" 文件创建日期" type="xs:date"/>
        <xs:element name=" 联系单位" type="联系单位 Type"/>
        <xs:element name=" 数据标识信息">
          .....
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:complexType name="联系单位 Type">
    .....
```

```
</xs:complexType>
<xs:complexType name="覆盖范围 Type">
  .....
</xs:complexType>
</xs:schema>
```

在这个 Schema 中,我们把遥感影像元数据标准中的 7 个元数据集定义成元素,每个数据集中的元数据项定义成该元素的子元素。将元数据标准中的 6 个公用数据类型定义成 Schema 的复杂数据类型(Complextype),另外还使用了 Schema 内嵌的数据类型,如 date、string 等。

在定义了 XML Schema 后,就可以用 XML 准确无误地表达遥感影像的元数据了。如果用户需要扩展元数据,只需在遵守元数据标准中的扩展性规则的基础上,在 Schema 中增加新的元素或类型即可,从而达到元数据扩展的要求。

3.4 遥感影像数据库引擎(ImageDE)的设计与实现

基于上述技术,使用 Oracle 作为后台数据库,我们开发了遥感影像数据库引擎(ImageDE),用于实现遥感影像数据的组织、管理、快速查询与检索。

ImageDE 的模块图如图 4 所示。

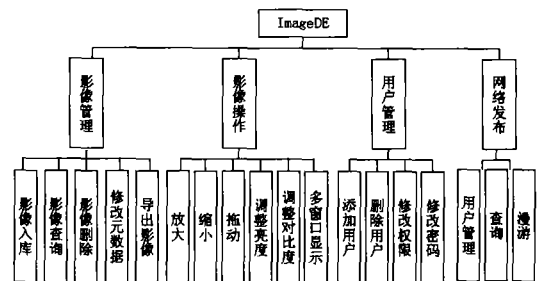


图 4 ImageDE 模块图

Fig.4 ImageDE modules diagram

ImageDE 主要有四大模块:影像管理、影像操作、用户管理、网络发布。影像管理模块主要是对影像数据库的维护操作,包括影像数据的入库、查询、删除、导出和修改元数据;影像操作模块主要是从数据库中将影像调出来浏览,同时可以进行放大、缩小、拖动、多窗口显示以及亮度和对比度的调整等;用户管理模块是对访问数据库的用户登记、设置权限,根据权限为用户开放不同的功能;网络发布模块指通过将元数据发布到 Internet 上,使用户可以利用浏览器对元数据进行查询,从而能够快

速、准确地从海量的遥感影像数据中查询、检索和了解所需的影像数据;对查到的影像可以通过浏览其元数据,了解影像的覆盖范围、内容信息、质量信

息、分发信息等。ImageDE 系统的运行情况如图 5 所示。

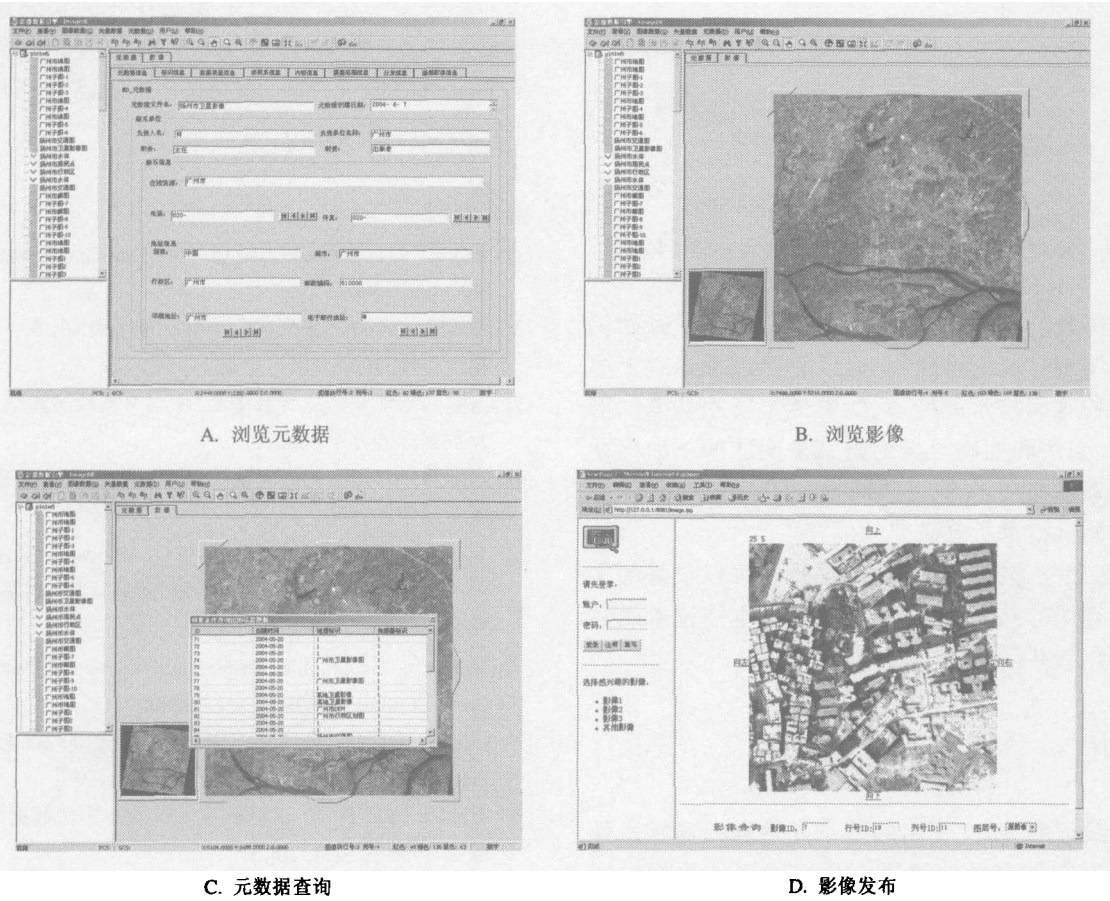


图 5 ImageDE 运行图

Fig.5 The diagrams of ImageDE

4 本体技术分析

本体(ontology)是从哲学的一个分支——形而上学中的本体论(Ontology)发展来的一个名词。本体论研究客观事物存在的本质,与认识论(Epistemology)相对。即,本体论研究客观存在,认识论研究主观认知。而本体(ontology)的含义是形成现象的根本实体。因而,本体是概念化的明确说明(Gruber, 1993)。最早把本体引入计算机领域的是人工智能领域。

地理信息本体与地理信息分类编码、地理信息标准术语表之间有着相似之处,本体论与分类学、术语学也存在一定的交叉。

然而地理信息本体并不是地理信息标准术语

表。地理信息本体提供了一组具有良好结构性的词汇,而且出现在本体中的词汇是经过严格选取,确保所选的词汇是本领域中最基本概念的抽象与界定。概念与概念之间的关系采用相应技术(如谓词、逻辑等)进行了完整而全面的反映。而正是这些关系的反映使得基于本体的系统实现后能够完成语义层面上的一些功能。而地理信息标准术语表仅仅是地理信息领域中各种词汇的集合,相对本体而言还是比较松散。

本体也不单纯是一个词汇的分类体系,即不是地理信息中的分类和编码表。本体和地理信息的分类非常相似,尤其是把本体的理论应用于地理信息分类编码时,这种相似性更为明显。总的说来,地理信息本体比分类编码表中所反映的词与词之间的关系要丰富。信息本体发展模式可参见图 6 所示。

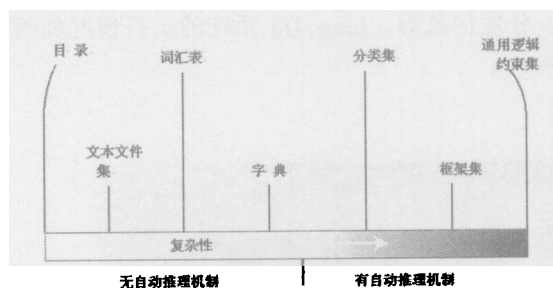


图 6 信息本体发展模式图

Fig.6 The development scheme of information Ontology

通俗地说,最根本的区别是 Ontology 一开始就致力于实现计算机可理解,所以它在表现形式上要有更为特殊而技术的处理,如本体是要用精确的形式语言、句法和明确定义的语义来阐述的。如何在现有的基础上做这些特殊而技术的处理就是我们结合计算机技术所要研究的内容。

通过建立遥感影像的领域本体,可以很好的实现遥感知识的共享与复用,实现基于本体技术的数据互操作与共享。

5 结语

实践证明,本文提出的基于元数据的遥感影像管理方案能够有效地解决海量遥感影像数据的组

织、存储和管理问题,通过遥感影像元数据的网络发布实现的遥感影像的网上查询、检索,使得遥感影像可以在互联网上交换、共享。这很好地解决了目前遥感影像生产部门中普遍存在的遥感影像数据量大,难于管理、检索,利用率低的问题,具有广泛的应用前景。目前,ImageDE 系统已在广州市城市规划部门投入使用,运行良好。

参考文献

- [1] ISO/TC 211, ISO/DIS 19115. Geographic information — Metadata. 2000.
- [2] 蒋景瞳,何建邦. 地理信息国际标准手册. 北京: 中国标准出版社, 2004.
- [3] 国家基础地理信息研究中心. 中国可持续发展信息共享元数据标准方案. 2001.
- [4] 萨师焯,王 珊. 数据库系统概论(第三版). 北京: 高等教育出版社, 2000.
- [5] Tom Barclay, Jim Gray, Don Slutz. Microsoft TerraServer: A Spatial Data Warehouse, <http://www.TerraServer.com>, 2002.
- [6] Managing Geographic Raster Data Using GeoRaster. An Oracle Technical White Paper. 2003.
- [7] Charles F Goldfarb. The XML Handbook (Fifth Edition). Prentice Hall, 2003.
- [8] 毕建涛,曹彦荣等. 遥感影像元数据与影像数据库研究. 测绘标准化, 2004, 20(2).

Research on Remote Sensing Image Database Engine

LIU Peng^{1,2}, BI Jiantao², CAO Yanrong¹, He Jianbang¹

(1 Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2 College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao 266510, China)

Abstract: Referring to the overseas and domestic data, this paper puts forward a remote sensing image metadata standard draft and describes remote sensing image data with the draft. It brings forward to manage remote sensing data and its metadata integratively with large RDBMS so that the structured storage of remote sensing image data can be realized. It presents remote sensing image metadata with XML and gives the XML Schema of remote sensing image metadata so as to implement the query and search of the metadata and the sharing of sense image data on the Internet.

Key words: metadata; remote sensing image database; XML; ontology