

遥感影像元数据处理技术研究

陈芃宇, 文成玉

(成都信息工程大学 通信工程学院, 四川 成都 610225)

摘要:文章提出了一种创建可配置的数据解析规则方法,用于解决从遥感影像数据头文件中提取元数据信息方法难统一的问题。经几种卫星影像数据头文件验证,该方法较好地解决了从不同遥感影像数据头文件自动提取元数据信息的问题。

关键词:遥感影像;元数据信息提取;解析规则

0 引言

种类繁多的遥感影像数据及其成果产品,其中大量的信息还以硬拷贝或光盘形式存储^[1],尤其是早期的影像数据,严重影响了数据使用的效率和价值。

实现遥感影像数据网络化的交换与共享,一方面需要建立符合元数据规范(如ISO19115-2)的遥感影像元数据^[2],另一方面,需要实现元数据处理过程的自动化。同时,新的影像数据头文件或头文件的格式、内容等发生变化,也都不能处理。为此,本研究提出了一种基于可配置的数据解析规则的元数据提取方法及生成元数据文件的模板,以实现遥感影像元数据的自动处理,可有效解决不同来源的遥感影像数据的元数据提取过程中的相关问题。

1 元数据及处理过程

遥感影像一般都有一个数据头文件描述该影像的信息,其中就包含了建立元数据的基础信息,但需要根据建立的元数据要素,从中提取出相关字段值作为相应元数据要素的值。

元数据目录仅包含遥感影像的简单信息,并且不能用于数据交换,而元数据文件则需要包含遥感影像元数据的详细信息且以可读的文件形式存在。综上所述,建立元数据的目的是要找一种适合组织管理网络资源的方法。随着元数据与网络技术的结合,元数据已经从最初的抽象描述数据对象转变为组织、管理、发现、交换和使用网络资源的重要工具和方法^[3],并贯穿于整个过程。

2 元数据的生成

2.1 数据解析规则建立

由于数据来源、传感器、平台载体、年代等的不同,这些文件包含的信息及格式都各不相同,归纳起

来可以分为3类:符合XML规则的文件、无规则的文本文件和Access数据库文件。一般为了解决解析数据格式困难可以根据已建立的元数据标准,为每种不同来源的遥感影像数据头文件建立解析规则。这样在提取元数据信息时,就可以根据映射关系从数据文件中读出需要的信息。采用可配置数据解析规则的优点是:将各种不同格式、内容的数据头文件的解析过程重组成统一格式的元数据要素与关键字映射的解析规则。数据解析规则文件中,在file元素中指定了头文件的格式(XML, TXT或DB)。在解析过程中,从DB类型的头文件中提取信息的方法与XML, TXT头文件是不同的。Node元素中的信息,将元数据要素和头文件中的主关键字联系起来;每个Node元素包含的5个子元素必须按序存放,不管是否有值。此外,解析规则文件中还将影像的共有属性和特有属性分开,分别对应着元数据库中一级元数据表和二级元数据表。

这样就把原来对不同头文件的处理过程统一为解析不同XML文件的过程,降低了设计提取元数据信息过程的复杂程度。

2.2 元数据提取

提取元数据信息的基本思想是:根据选定的影像数据包自动匹配一个数据解析规则文件,然后根据解析规则文件中指定的影像数据头文件的类型(DB与XML, TXT)选择不同的处理过程(DB文件用连接Access数据库方式,XML和TXT文件用文本处理方式),再依次提取出与元数据要素映射的主关键字的值,直到解析完XML的所有内容。其中,DB文件的方式中,由于数据库表中的字段名都具有唯一性,所以只需从XML解析规则中分别读出所有元数据要素

作者简介:陈芃宇(1993—),男,四川什邡人,本科生;研究方向:互联网应用开发。

标识和主关键字,并用主关键字组成SQL语句查询DB文件,从而就可以依次得到主关键字的值了。首先,从头文件只能提取到建立影像元数据的部分内容,而有些内容是需要人工输入。其次,元数据信息入库时,首先是写入一级元数据表;然后再写入二级元数据表、影像数据表及影像金字塔数据表等。此外,影像快视图也是一个重要的元数据要素。大多数影像数据包都已经有了JPEG/JPG格式的快视图。对于没有头文件的影像数据包,首先调用GDAL/OGR等工具读出遥感影像(TIFF格式)中的基本信息,再用手工方式建立元数据信息。

3 元数据文件生成

在基于元数据的遥感影像网络共享与发布系统中,用户不直接面对影像数据本身,用户是通过元数据目录及元数据文件和系统交互的。针对数字影像特征,ISO19115系列标准提供了核心的和扩展的元数据要素。但在具体实现时,由于不同的系统在元数据内容和应用策略等方面存在差异,因而提供的元数据内容及格式也不尽相同。为了保证各种格式的元数据文件能够被正确解析,提供的元数据需要通过ISO 19139系列的XML规范的验证。只有通过验证的元数据才能保证其内容是以ISO19115系列标准定义的结构来存放的。根据元数据模板生成的元数据文件以XML文件保存。通过分析其中的元数据要素与内容的对应关系,就能找到可转换的实质内容,这样就可以设计元数据文件转换程序实现不同元数据文件间的转换。这样的元数据文件是适合通过网络或离线方式进行影像数据的共享和交换。

从应用层需求考虑,不同的用户对元数据内容需求是不一样的。根据固定格式的元数据模板,生成内容统一的元数据文件,显然不能满足这样要求。通过向用户提供可配置的元数据模板生成方式,除必须的

元数据要素外,用户可以选择感兴趣的内容。这样根据元数据模板生成的元数据文件就是用户真正关心的内容,这也体现了元数据是面向应用的特征。生成的元数据模板是按相关规范组织的(可参考ISO 19139系列标准的附录,这里不再给出详细内容),是可验证的。此外,元数据模板需要将元数据要素组织为两级,分别对应影像的一级和二级元数据表中的内容。同时,需要为每类影像(如卫星影像,航空影像,正射影像,DEM等)设置元数据模板,才能生成各类影像的元数据文件。

4 结语

建设基于元数据的遥感影像网络共享与发布是遥感影像数据管理的必然趋势,这其中会涉及如信息存取完全等诸多重要问题。本文针对其中的元数据处理问题提出的元数据处理方法,在很大程度上实现了数据处理的自动化,尤其是元数据的自动提取。但是,由于不能完全得到现存的各类影像数据头文件的内容,因此,基于可配置的数据解析规则的元数据提取方法,针对未验证过的头文件内容,可能还需要进一步的完善和优化。

参考文献

- [1] 邢永超,马荣华,何隆华.基于ArcGIS Server的遥感影像检索系统的设计与实现[J].微计算机信息,2012(6):113-115.
- [2] 张龙,汪新庆.基于数据字典的空间数据库通用建库技术[J].国土资源遥感,2014(1):173-178.
- [3] 杜晓,晋佩东,郭大海,等.遥感影像快速入库和元数据提取系统的研究[J].国土资源遥感,2006(3):80-84.

(责任编辑 王宁)

Research on remote sensing image metadata processing technology

Chen Pengyu, Wen Chengyu

(Communication Engineering College, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: A method is presented to resolve the problems in exacting metadata information from the header file of remote sensing image by creating configurable data parsing rules. After testing with several kinds of satellites' header files, the results show that this method can exact the information from different header files properly and easily.

Key words: remote sensing image; metadata information extraction; parsing rules