

基于 Filter Encoding 的 GML 空间数据查询

苗立志, 张书亮, 伍 蓝, 阎国年, 焦东来

(南京师范大学虚拟地理环境教育部重点实验室, 南京 210046)

摘要: 随着地理数据网络化程度的提高和 GML 空间数据的应用越来越广泛, 如何从海量的 GML 空间数据中检索出符合特定条件的地理要素数据成为目前 GML 研究的焦点问题之一。该文根据 OGC 制定的过滤编码执行规程, 应用规程中的谓词作为查询关键字, 结合 GML 空间数据的特点, 通过查询转译器和优化器实现了一种应用于 GML 空间数据查询的可行性方法。

关键词: 查询转译器; 查询优化器; 过滤编码

GML Spatial Data Query Based on Filter Encoding

MIAO Li-zhi, ZHANG Shu-liang, WU Lan, LV Guo-nian, JIAO Dong-lai

(Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210046)

【Abstract】 Considering characters of GML spatial data, using the predications of OpenGIS® filter encoding implementation specification as query key words, transforming and optimizing the query expression by middleware, this paper presents an accessible method to implement query for GML spatial data according to filter encoding proposed by OGC. The objective is to provide one way that can help to achieve the query given data from massive GML spatial data, which is the focus of GML research along with geographic data transferring through Internet and GML spatial data application increasingly.

【Key words】 query transformation; query optimization; filter encoding

1 概述

随着GML核心模式的不断完善与越来越多的空间数据提供商和GIS厂商对GML空间数据的认同, GML格式的空间数据也越来越普及。这其中主要包括GML格式的空间数据交易以及GML数据为基础的面向Web的大众化GIS应用。例如, 英国的测绘机构Ordnance Survey Great Britain已经要求使用GML格式来存储全国的数据层, 并将其作为中间数据文件提供给用户; 荷兰地形和地籍部门、英国普查部门等都在数据生产中将GML作为其主要的格式^[1]。GML应用的发展使GML格式的数据文件的数量也随之越来越多; 用户对从海量的GML空间数据和复杂的GML空间数据对象中检索满足自身需求的信息的要求也越来越迫切。GML空间数据查询研究已成为GML空间数据在GIS应用研究中的热点问题之一。

国内外对 GML 查询的研究大多都是集中于扩展 SQL 和 XQuery, 增加可以支持空间操作的算子函数以实现 GML 数据的空间查询。文献[2]中提出基于 SQL 的 GML 空间查询语言, 由于 XML 的嵌套数据模型与传统的关系模型之间具有非常大的差异, 扩展 SQL 以支持 GML 查询也与当前 XML 查询标准不相一致, 并不是 GML 查询的最有效方法。文献[3]中提出了对 XQuery 语言进行扩展以支持 GML 查询的方法, 但文中并没有涉及如何实现以及实现了多少 GIS 空间操作支持等更深层次的问题。国内的研究主要有文献[4]中选择开源代码的 XQuery 引擎 XQEngine 进行扩展, 使其支持空间数据类型和空间操作算子来实现 GML 查询; 文献[5]中也扩展了 XQuery 使其支持空间操作以支持对 GML 文档数据的空间查询。上述无论是基于 SQL 还是基于 XQuery 的扩展都在一定程度上实现了 GML 空间数据的查询, 但查询语句语法相对比较复杂、专业化程度强致使难以在 GIS 应用系统

中使用, 查询功能谓词虽然较多, 但与过滤编码相比显得还不够丰富。

2 过滤编码

2.1 编码特点

OGC于2005年5月推出《OpenGIS过滤编码执行规范》, 它最初是作为网络要素服务(Web Feature Service, WFS)的一个组成部分, OGC制定过滤编码规范的目的是把它作为系统中立的查询谓词, 用来描述OGC CQL(Common Catalog Query Language)的XML编码。目前有大量的XML工具可以使用, XML的表示格式可以很容易地验证、解析或是转换成其他目标语言。例如, XML编码过滤器可以转换成SQL中SELECT句式的WHERE语句, 来获取存储在基于SQL关系数据库中的数据。类似地, 它也可以转换成XPath或XPathPointer来获取XML文档中的数据^[1]。

Filter 编码表达式的根节点是<Filter>, 它由以下模式片断定义:

```
<xsd:element name="Filter" type="ogc:FilterType"/>
<xsd:complexType name="FilterType">
  <xsd:choice>
    <xsd:element ref="ogc:spatialOps"/>
    <xsd:element ref="ogc:comparisonOps"/>
    <xsd:element ref="ogc:logicOps"/>
  </xsd:choice>
</xsd:complexType>
```

基金项目: 国家自然科学基金资助项目“GML 空间数据存储索引机制研究”(040401045)

作者简介: 苗立志(1981 -), 男, 博士研究生, 主研方向: GIS 应用, 地理信息共享与空间数据互操作; 张书亮, 副教授、博士; 伍 蓝, 硕士研究生; 阎国年, 教授、博士生导师; 焦东来, 博士研究生

收稿日期: 2007-02-10 **E-mail:** miaolizhi@126.com

```
<xsd:element ref="ogc:_Id" maxOccurs="unbounded"/>
</xsd:choice>
</xsd:complexType>
```

元素<logicalOps>, <comparisonOps>和<spatialOps>分别是逻辑、空间和比较操作符的替代组。在过滤表达式中,逻辑操作符可以用于联合空间操作符和比较操作符。<_Id>元素是对象标识符替代组的头结点。

空间操作符用来判断几何属性值是否满足指定的空间关系,如果满足空间关系则值为 TRUE,否则值为 FALSE。通常包括以下几类:

(1)<BBOX>:它是一种非常方便和简洁的基于 gml:Envelope 的矩形空间约束编码元素。相当于空间操作<Not><Disjoint> ... </Disjoint></Not>,返回在空间上与矩形边框相交的所有几何。

(2)Equals, Disjoint, Touches, Within, Overlaps, Crosses, Intersects 与 Contains 等空间操作谓词,它们在文献[6]中均有定义。

(3)DWithin 和 Beyond:分别用于判定几何属性值是否在一个指定距离的范围之内或范围之外。距离值用元素<Distance>的内容表示,units 属性用来指定测量单位。

比较操作符用来构成由两个数学表达式组成的比较表达式。除了包括基本的=, <, >, >=, <=和<>等,还包括<PropertyIsLike>, <PropertyIsBetween>和<PropertyIsNull>。<PropertyIsLike>元素用模式匹配方法进行字符串比较操作;<PropertyIsBetween>元素用一种简单的方式来进行范围检查;<PropertyIsNull>元素用来检查内容值是否为 NULL。

逻辑操作符用来连接一个或多个条件表达式。<And>, <Or> 和<Not>用来连接数量表达式、空间表达式或逻辑表达式,形成更复杂的混合表达式。

对象标识符为网络服务上下文中的对象实例指定唯一标识。

查询表达式可由<Add>, <Sub>, <Mul>, <Div>, <PropertyName>, <Literal>和<Function>等操作符组成。它们都是 expression 的替换组,可以通过联合这些元素形成 XML 查询片段。

2.2 过滤编码与空间查询的结合点

空间数据查询是 GIS 的一个重要功能,一般定义为作用在 GIS 数据上的函数,它返回满足条件的内容。空间查询大致可以分为 3 类:(1)仅包含空间关系和空间属性的查询;(2)仅包含非空间属性的查询;(3)混合查询(包含非空间属性、空间关系和空间属性的查询),其中仅包含非空间属性的查询可以使用过滤编码的比较操作符,或者联合逻辑操作符来完成。空间查询请求使用过滤编码的空间操作算子联合比较操作符和逻辑操作符可以实现。过滤编码规范提供的操作算子支持 GML 表达的嵌套空间数据模型,满足空间数据查询的要求,易于实现由 GML 表达的地理实体空间数据的检索与查询。

3 基于过滤编码的 GML 数据查询实现

3.1 查询实现框架

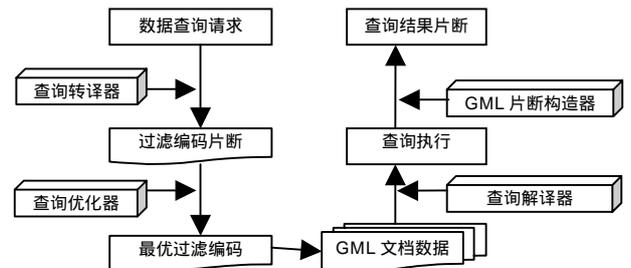
GML 应用文档数据作为一种空间数据,其空间查询同样主要包括以上 3 类。但 GML 数据模型跟传统数据模型又有区别,是一种特殊的空间数据,在查询的过程中需要支持 GML 数据模型和 GML 模式中定义的数据类型,包括几何、要素和拓扑等,而且查询结果要求也能够采用合法的 GML

文档或者片断的形式。

过滤编码规范的出现解决了上述难题,其基础模式引用了 GML 核心模式中的几何模型定义 geometryAggregates.xsd 以支持 GML 的几何数据类型;过滤编码的空间操作是从 Simple Feature Specification(SFS) For SQL 衍生发展而来,而 GML 的数据模型也是在 SFS 的基础上扩展完善起来的,所以过滤编码的空间操作算子也适用于 GML 的数据模型的查询分析操作。

本文中 GML 空间数据查询的流程框架如图 1 所示。

图 1 GML 空间数据查询流程



一个 GML 数据空间查询请求语句(由一个或多个查询表达式构成)首先经过查询转译器进行转译成 Filter 过滤条件片断,检查该片断是否符合 Filter 基础模式的词法和语法,如片断中的类型是否有定义、嵌套结构是否正确等;若存在错误,提示错误信息,查询不再往下执行。

对检查无误的 Filter 编码片断进行优化,将其内嵌到查询请求客户端的过滤编码 XML 文档;由 WFS 服务器端接收和解析查询请求;然后对指定的 GML 数据文档过滤得到符合查询条件的地理要素,返回查询结果 GML 数据片断或文档;最后,通过 SVG 图形显示系统,将查询的结果数据以地图的形式显示出来。

3.2 查询实例

本例应用美国明尼苏达大学开发的开源软件 Map Server4.4.2 作为解译过滤编码查询片段的工具,以实现 GML 文档数据的查询。数据源为 river.xml 文档,查询名称为 Great Bear 的河流,即 NAME="Great Bear"。根据查询条件,应用查询转译器和查询优化器,将查询条件转化为如下最优过滤编码片断:

```
<Filter>
<PropertyIsEqualTo>
<PropertyName>NAME</PropertyName>
<Literal>Great Bear</Literal>
</PropertyIsEqualTo>
</Filter>
```

将过滤编码片断内嵌到支持过滤编码的 WFS 查询请求语句中,借助 MapServer 软件进行解析,获取满足查询条件的地理空间数据,并以 GML 数据文档片断的格式返回,如下所示。同样,可以实现对 GML 数据的空间查询和地理要素间空间关系的判定。

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<wfs:FeatureCollection
xmlns:ms_ogc_workshop="http://localhost/ms_ogc_workshop"
xmlns:wfs="http://www.opengis.net/wfs"
xmlns:gml="http://www.opengis.net/gml"
xmlns:ogc="http://www.opengis.net/ogc"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/wfs
```

```

http://schemas.opengespatial.net/wfs/1.0.0/WFS-basic.xsd
http://localhost/ms_ogc_workshop
http://localhost/cgi-bin/mapserv.exe?map=/ms4w/apps/ms_ogc_w
orkshop/service/config.map&SERVICE=WFS&VERSION=1.0.0
&REQUEST=DescribeFeatureType&TYPENAME=rivers&OUT
PUTFORMAT=XMLSCHEMA">
<gml:boundedBy>
<gml:Envelop srsName="EPSG:4326">
<gml:coordinates>-125.383850,64.878967
-117.252357,67.112030</gml:coordinates>
</gml:Envelop>
</gml:boundedBy>
<gml:featureMember>
<ms_ogc_workshop:rivers>
<gml:boundedBy>
<gml:Envelop srsName="EPSG:4326">
<gml:coordinates>-125.383850,64.878967
-117.252357,67.112030</gml:coordinates>
</gml:Envelop>
</gml:boundedBy>
<ms_ogc_workshop:msGeometry>
<gml:LineString srsName="EPSG:4326">
<gml:coordinates>-123.263069,65.212860 ...
-125.383850,64.955276</gml:coordinates>
</gml:LineString>
</ms_ogc_workshop:msGeometry>
<ms_ogc_workshop:NAME>Great
Bear</ms_ogc_workshop:NAME>
<ms_ogc_workshop:SYSTEM>Mackenzie</ms_ogc_workshop:S
YSTEM>
</ms_ogc_workshop:rivers>
</gml:featureMember>
</wfs:FeatureCollection>

```

4 结束语

目前,对于 GML 空间数据的查询还没有统一的方法和语言,如何从大量的 GML 空间数据中检索得到满足一定条件的地理空间数据也成为 GML 研究的主要内容之一。由于

(上接第 23 页)

参考文献

- [1] Akyildiz I F, Sankarasubramaniam Y, Cayirci E. A Survey on Sensor Networks[J]. IEEE Communications Magazine, 2002, 40(8): 102-114.
- [2] 孙利民, 李建中. 无线传感器网络[M]. 北京: 清华大学出版社, 2005.
- [3] 任彦, 张思东, 张宏科. 无线传感器网络中覆盖控制理论与算法[J]. 软件学报, 2006, 17(8): 422-433.
- [4] Meguerdichian S, Koushanfar F, Potkonjak M, et al. Coverage Problems in Wireless Ad-hoc Sensor Networks[C]//Proc. of IEEE INFORCOM'01. [S. l.]: IEEE Press, 2001.
- [5] Meguerdichian S, Koushanfar F, Potkonjak M, et al. Worst and Best-case Coverage in Sensor Networks[J]. IEEE Trans. on Mobile Computing, 2005, 4(1): 84-92.
- [6] Veltri G, Potkonjak M. Minimal and Maximal Exposure Path Algorithms for Wireless Embedded Sensor Networks[C]//Proc. of the 1st International Conference on Embedded Networked Sensor Systems. Los Angeles, California, USA: [s. n.], 2003.

GML 数据不仅仅是结构化的 XML 文档,并且还包含空间、拓扑和时态等信息,这就使得原本可以用于 XML 文件的查询语言不能实现对 GML 数据包括空间查询的请求。本文应用 OGC 制定的过滤编码执行规程,借助其谓词,结合 GML 空间数据的特点,通过查询转译器和优化器实现了一种应用于 GML 空间数据查询的可行性方法,也是 GML 空间数据查询的初步研究成果。

本文主要是应用过滤编码实现了 GML 空间数据的查询,该方法的优点是空间操作算子丰富、语法简单、易于实现,复杂查询可以嵌套查询表达式实现等,但此方法是建立在第三方开源软件的基础之上的。为更好地实现对 GML 空间数据的查询还需要提出和实现具有相对独立性的、支持地理数据的诸如相交、叠加、包含等空间以及拓扑和时态操作的查询操作算子,或开发能准确解译过滤编码语义的解译软件等,这也将是今后 GML 空间数据查询的主要研究内容。

参考文献

- [1] OGC. OpenGIS® Filter Encoding Implementation Specification (Version 1.1)[Z]. (2004-09-09). http://portal.opengespatial.org/files/?artifact_id=8340.
- [2] Corcoles J E, Gonzalez P. A Specification of a Spatial Query Language over GML[C]//Proceedings of the 9th ACM International Symposium on Advances in Geographic Information Systems. [S. l.]: ACM Press, 2001.
- [3] Vatsavai R R. GML-QL: A Spatial Query Language Specification for GML[Z]. (2002-08-08). www.cobblestoneconcepts.com/ucgis2summer2002/vatsavai/vatsavai.htm.
- [4] 兰小机. GML 空间数据存储索引机制研究[D]. 南京: 南京师范大学, 2004.
- [5] Guan Jihong, Zhu Fubao, Zhou Jiaogen, et al. GQL: Extending XQuery to Query GML Documents[J]. Geo-spatial Information Science, 2006, 9(2): 118-126.
- [6] OGC. OpenGIS® Simple Features Specification for SQL Revision 1.1[Z]. (1999-06-06). <http://www.opengespatial.org/docs/99-049.pdf>.
- [7] Corts J, Martnez S, Karatas T, et al. Coverage Control for Mobile Sensing Networks[J]. IEEE Trans. on Robotics and Automation, 2004, 20(2): 243-255.
- [8] Li Xiangyang, Wan PengJun, Frieder O. Coverage in Wireless Ad-hoc Sensor Networks[J]. IEEE Trans. on Computers, 2003, 52(6): 115-121.
- [9] Zhang Honghai, Jennifer C Hou. Maintaining Sensing Coverage and Connectivity in Large Sensor Networks[J]. Wireless Ad-hoc and Sensor Networks: An International Journal, 2005, 1(1/2): 89-124.
- [10] Tian Di, Georganas N D. Connectivity Maintenance and Coverage Preservation in Wireless Sensor Networks[C]//Proc. of Conference on Electrical and Computer Engineering. Canadian: [s. n.], 2004.
- [11] Meguerdichian S, Koushanfar F. Exposure in Wireless Ad-hoc Sensor Networks[C]//Proc. of ACM International Conference on Mobile Computing and Networking. Rome, Italy: [s. n.], 2001.
- [12] Adlakh S, Srivastava M. Critical Density Thresholds for Coverage in Wireless Sensor Networks[C]//Proc. of IEEE WCNC'03. [S. l.]: IEEE Press, 2003.