

基于地理语义的空间关系查询和推理

杜冲^{1,2}, 司望利^{1,3}, 许璐¹

(1. 中科院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101; 2. 首都师范大学
三维信息获取与应用教育部重点实验室, 北京 100048; 3. 长安大学地质工程与测绘学院, 西安 710064)

摘要: 地理空间中的地物, 根据其本体类型具有不同的地理语义特征。受到地理语义的影响, 人们在描述几何特征和关系相似的物体之间的空间关系时, 所用的词汇是不同的。目前, 人们在空间关系的自然语言查询方面的研究主要集中在空间关系描述、空间关系自然语言形式化表达、空间关系自然语言查询语句的转换, 以及空间关系自然语言查询的查询接口等方面, 没有考虑到地理本体的语义差异, 这会使得空间关系的自然语言查询不能符合人类语言和认知的习惯。本文依据一些学者在空间关系的自然语言描述的形式化(主要是线和线、线和面等形状的空间地物之间的关系)研究的基础上, 针对不同地理本体类型的空间地物和空间关系自然语言描述词汇之间对应的语义关系有所不同的特点, 总结了自然语言词汇在描述空间关系时的对应规则, 提出不同特征类型地物和描述词汇之间的规则库设计方法, 并设计部分空间地物的规则实例。最后, 结合空间关系判断函数, 设计具体查询系统实践举例证明。

关键词: 空间关系; 自然语言; 地理本体; 语义

1 引言

随着各种智能化查询服务技术的提升, 要使GIS被更广泛的用户接受和使用, 就需要将更简单、更智能化的操作方式应用到GIS领域。因此, 建立以自然语言为输入的智能GIS查询系统具有很大的实用意义。通过理解自然语言词汇对应的空间实体之间空间关系的形式化描述, 得到人们用自然语言表述的查询结果, 使得GIS能够根据查询条件, 进行一定的空间关系推理, 这有助于提升GIS服务的智能化水平, 为更普通的用户提供服务, 从而利于实现PPGIS(Public Participation GIS)。

目前, 在面向自然语言的空间关系查询方面的研究主要集中在空间关系描述^[1-4]、空间关系自然语言形式化表达^[5-11]、空间关系自然语言查询语句转换^[12-17], 以及空间关系自然语言查询接口^[6, 8, 18]等方面。以往关于空间关系的研究, 往往将参与的空间对象抽象为不具有地理语义的空间几何体^[6-8], 很少考虑在空间关系自然语言表达

时, 不同的地理特征类型所引起的语义差别影响。因此, 在具体的自然语言查询实现时, 还难以完成面向现实中客观存在的空间物体的查询要求。

地理空间特征不仅仅是简单的点、线、面的组合, 而是具有一定的地理语义特征的地理实体。因此, 对地理特征空间关系的描述, 不仅受到物体之间几何关系的影响, 也受到地理本体所属的类型和语言环境等非空间属性的影响^[2-3, 19]。例如, 人们说一条支流汇入或流入干流, 却不能这样描述公路, 即使它们具有相同的几何空间关系。这就是由地理特征类型引起的语义差异。此外, 还有由个人的行为目的、行为方式导致的个体认知差异而形成的不同地理语义^[9-10, 20]。因此, 在空间关系的自然语言研究中, 需要结合地理本体的语义特征, 将地理语义应用到空间推理中去, 以获得更多的地理知识, 促使更加完备的知识推理。

本文针对不同地理本体类型的空间地物和空间关系自然语言描述词汇之间对应的语义关系有所不同的特点, 采用已有的空间关系的自然语言

收稿日期: 2009-03-12; 修回日期: 2010-01-06.

基金项目: 国家自然科学基金项目(40701151); 资源与环境信息系统国家重点实验室自主创新资助; 中国科学院资源与环境信息系统国家重点实验室开放研究基金资助项目(A0702)。

作者简介: 杜冲(1983-), 男, 硕士生, 山东淄博人。主要研究方向为地理信息系统应用及地理认知。

E-mail: duchong017@163.com

描述的形式化方法，结合地理本体和描述词汇之间的知识规则库设计方法，实现了部分地物的本体语义特征的空间关系自然语言查询和关系推理，以此提高 GIS 查询的自然语言空间关系理解能力和服务智能化水平。

2 地理语义的空间关系操作算子

自然语言关于空间关系的描述是模糊的、定性的，要将其转换成计算机能够理解的内容，首先，要建立定性的自然语言描述与定量化的计算机模型之间的联系，即解决自然语言的形式化问题。目前，关于空间关系的形式化研究主要是 Shariff 等人关于面状和线状地理要素之间关系的研究^[5-6]，和 Nedas、许珺等人关于线状和线状地理要素之间关系的研究^[7-8]。它们的主要思想都是在关于空间关系的自然语言描述的认知实验基础上，通过结合反映拓扑关系的定量指标和反映各种类型的空间物体的度量关系的定量指标，使用统计或数据挖掘的方法，对实验的结果进行分析，得出可以表达定性的语言描述的定量化计算规则，利用这些规则进行空间关系的自然语言的查询。

本文基于 Shariff 和许珺的自然语言形式化方法，实现空间关系操作算子。空间关系操作是空间关系查询最终要执行的任务，空间关系查询操作最终由操作算子(函数)予以完成。空间关系操作算子完成的主要操作是：根据计算空间关系自然语言描述词汇对应的定量化指标，以此实现查询和推理的过程判断是否具有这个词汇所对应的空间关系语义。根据 Shariff 和许珺的理论，算子的设计主要分为两个部分，首先，是定量化指标的计算，然后，是具体算子的实现。

2.1 定量化指标的计算

定量化指标主要包括拓扑关系指标和度量关系指标，这些指标的作用在于形式化空间物体空间关系的自然语言描述，并最终作为定性的自然语言词汇蕴含空间语义的定量化判断条件。

(1) 拓扑关系指标的计算

选用 9-交集模型来表达空间物体的拓扑关系。9-交集模型采用点作为原始的空间实体，地理目标被定义为点的集合，表示二元的空间拓朴关系的模型^[21-22]。给定一个区域 Y ，与之相关的

有三个点集：内部点集(Y°)、外部点集(Y^{-})和边界点集(∂Y)。两个物体 A 和 B 之间的二元拓朴关系 R 是 A 的三个点集和 B 的三个点集的有序交集，可表示为一个 3×3 的矩阵。

$$R(A,B) = \begin{pmatrix} A^{\circ} \cap B^{\circ} & A^{\circ} \cap \partial B & A^{\circ} \cap B^{-} \\ \partial A \cap B^{\circ} & \partial A \cap \partial B & \partial A \cap B^{-} \\ A^{-} \cap B^{\circ} & A^{-} \cap \partial B & A^{-} \cap B^{-} \end{pmatrix}$$

如果空集和非空集分别用 0 和 1 表示，则理论上 9-交集模型可以区分 2^9 (即 512)种不同的拓朴关系，而实际上两个物体之间的拓朴关系不可能有这么。在 9-交集模型中，对于二维空间中的线状物体和面状物体之间的拓朴关系，可以实现的有 19 种，Shariff 将其记作 {LR11, LR12, LR13, LR22, LR31, LR32, LR33, LR42, LR44, LR46, LR62, LR64, LR66, LR71, LR72, LR73, LR74, LR75, LR76}；而对于简单线状物体(即没有分支、自相交等情况的线状物体)，由于两个线状物体的外部点集的交集($A^{-} \cap B^{-}$)总是非空的，因此，只用模型中的另外 8 个值作为表示线状物体之间拓朴关系的定量指标，Xu 将其分别记作 { $R_{11}, R_{12}, R_{13}, R_{21}, R_{22}, R_{23}, R_{31}, R_{32}$ }，由此可以实现的拓朴关系有 33 种。部分拓朴关系类型如图 1、图 2 所示(具体拓朴关系类型详见文献 [6]、[22])。

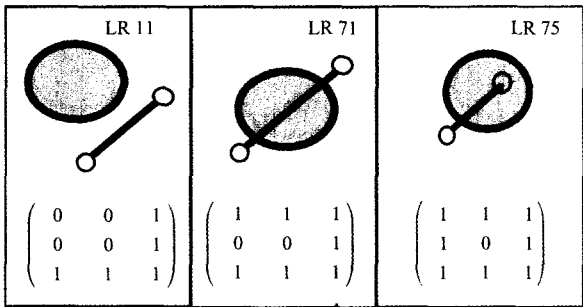


图 1 线面物体之间的拓朴关系类型示例^[6]

Fig. 1 Examples of line-area topological relations

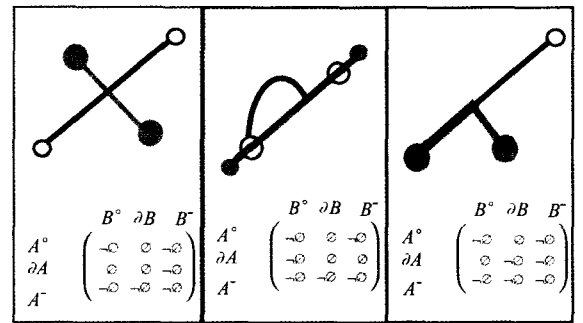


图 2 线线物体之间的拓朴关系类型示例^[22]

Fig. 2 Examples of line-line topological relations

(2)度量关系指标的计算

度量关系指标主要是设计两个空间对象之间的角度、距离、分裂度、走向等值的计算方法。由于实体类型的不同,所用度量指标也有

所差别。Shariff 采用的关于面状和线状物体之间的度量关系指标,以及 Xu 采用的关于线状与线状物体之间的度量关系指标,具体如表 1 所列。

表 1 度量关系指标
Tab.1 Metric indices

物体类型	度量指标	具体包括的指标
线状物体与面状物体	分裂度	内部区域分裂度(IAS)、外部区域分裂度(EAS)、内部穿越分裂度(ITS)、外部穿越分裂度(ETS)、边界重合度(PA)、线重合度(LA)、区域边界重合度(RBS)
	相近度	内部相近度(IC)、外部相近度(OC)、内部接近度(IN)、外部接近度(ON)
	大体走向	内部边界大体走向(IPA)、内部线性大体走向(ILA)、外部边界大体走向(OPA)、外部线性大体走向(OLA)
线状物体与线状物体	角度	局部角(LA)和全局角(GA)
	分裂度	交叉分裂度(IS)、重叠分裂度(AS)、内部穿越度(ITS)
	距离	最小距离-长度比值(SDLR)、最大距离-长度比值(LDLR)、最小距离-最大距离比值(SLDR)
	重合度	最小外接矩形的重合部分占两个地物最小外接矩形的比例(OR)

各指标所代表的具体含义为:

面状和线状物体之间的度量关系指标(详见文献 [6]):

- 内部区域分裂度(IAS): 线将面分成的两部分中较小部分面积占面的面积的比值
- 外部区域分裂度(EAS): 线和面组成的闭合面域和面的面积比值
- 内部穿越分裂度(ITS): 线在面内部的部分与线的长度比值
- 外部穿越分裂度(ETS): 线在面外部的部分与线的长度比值
- 边界重合度(PA): 线与面的边界的重合部分与面的边界长度比值
- 线重合度(LA): 线与面边界的重合部分与线的长度比值
- 区域边界重合度(RBS): 线和面相交成闭合区域的边界和面的边界重合部分与面的边界长度比值
- 内部相近度(IC): 面以线在面内的端点到面边界的距离进行缓冲缩小所形成的面与面的面积比值
- 外部相近度(OC): 面以线在面外的端点到面边界的距离进行缓冲放大所形成的面与面的面积比值
- 内部接近度(IN): 面以线到面边界的最小距离进行缓冲缩小所形成的面与面的面积比值

- 外部接近度(ON): 面以线到面边界的最小距离进行缓冲放大所形成的面与面的面积比值
- 内部边界大体走向(IPA): 线在面外的部分的长度与面以线到面边界的最小距离进行缓冲缩小所形成的面的边界长度比值
- 内部线性大体走向(ILA): 线在面外的部分与线的长度比值
- 外部边界大体走向(OPA): 线在面内的部分的长度与面以线到面边界的最小距离进行缓冲放大所形成的面的边界长度比值
- 外部线性大体走向(OLA): 线在面内的部分与线的长度比值
- 线状和线状物体之间的度量关系指标(详见文献 [8]):
- 局部角(LA): 两条线状物体相交处的局部形成的锐角
- 全局角(GA): 两条线状物体的最小外接矩形相交形成的锐角
- 交叉分裂度(IS): 被分割成的较短的一段物体的长度占物体总长度的比例
- 重叠分裂度(AS): 两条线状物体重叠的部分占物体总长度的比例
- 内部穿越度(ITS): 两个最靠近端点的相交点之间的物体长度占物体总长度的比例
- 最小距离-长度比值(SDLR): 两个物体间的最小距离与物体总长度的比值

- 最大距离 - 长度比值 (LDR): 两个物体间的最大距离与物体总长度的比值
- 最小距离 - 最大距离比值 (SLDR): 两个物体间的最小距离和最大距离的比值
- 重合度 (OR): 两个物体的最小外接矩形的重合部分占两个物体最小外接矩形的比例

2.2 操作算子

根据 Shariff 和许珺等人通过认知实验数据分析出来的规则, 结合上面指标的计算, 设计出线状物体和面状物体之间, 以及线状物体与线状物体之间的空间关系算子。

例如, 对于线状物体与面状物体之间的空间关系算子 Pass 有如下规则:

LR11 = True And 0.18 < ON < 1.84 And
0.10 < OPA < 0.60 And
0.10 < OLA < 0.90

同理, 对于线状物体与面状物体之间的空间关系算子 Intersect, 满足的规则如下:

LR75 = True And
0.09 < ITS < 0.40 And
0.60 < ETS < 0.91 And
0.35 < IC < 0.99

Or

LR71 = True And
0.01 < IAS < 0.49

由上边规则就可以判定两个空间物体是否具有自然语言词汇 Pass 和 Intersect 所蕴含的空间关系语义。以此可以分别设计出线状物体和面状物体之间 Pass、Intersect 等自然语言词汇的空间关系判断算子。

3 基于本体的空间关系描述规则及其表达

在通常的空间推理研究中, 参与的对象被抽象为不具有地理语义的空间几何体, 如点、线或面。然而, 地理空间中的所有地物都具有特定的地理特征类型语义。由于地物空间属性和非空间属性在空间认知中的紧密耦合, 在地理空间推理中需要同时表达几何体的空间推理和地理语义推理, 才符合人的地理空间认知^[19]。人们在用自然语言描述空间关系时, 同样的空间关系所用到的

自然语言词汇也是不同的, 例如: 同样的空间关系, 河流会用汇入、流入或注入, 而道路之间却用到相交、交叉等来描述。地理本体自身的空间特征也影响着空间查询和推理。因此, 要实现空间关系的自然语言查询, 需要从面向客观存在的空间实体及其关系出发, 考虑地理本体类型所带来的空间关系语义影响。

本文采用产生式表达方式来表达空间实体之间的空间关系知识规则。这里将知识规则的表达分为两个阶段, 一阶段是将人脑的判断推理过程用于产生式规则予以表达, 另一阶段是将产生式规则用数据结构予以存储。第一阶段的实现有利于程序的编写、系统的开发; 而第二个阶段是如何将知识规则进行计算机存储的问题, 是充分利用知识库进行自然语言查询的关键, 知识规则的有效存储也是实现面向具体服务的知识规则自定义的基础。

为了满足空间关系的自然语言查询要求, 通过相关的专题咨询、资料查阅、经验收集等方式进行知识规则的获取, 收集到具体应用时所有可能的知识规则, 来构建知识规则库。由于地理特征类型较多, 本文通过下面的知识表达规则举例说明。

本文知识库主要面向地理特征类型所蕴含的知识规则, 因此, 这里的知识规则库的建立主要用于区分:

(1) 空间语义一样, 但是, 不同对象间对应的描述词汇可能不同。

例如: 同样的空间的关系, 河流用流入、注入, 而道路则会用到通向、相交;

(2) 有些地理特征类型的空间地物之间不存在某些空间关系。

例如: 河流之间多数是单侧相交的关系, 很少有交叉交叉 (cross) 空间关系; 公路和小粒度的面状建筑物之间不会有交叉通过 (intersect, cross), 而河流、公路与行政区划这样大粒度的面状地物主要存在的关系是相交 (cross)。

(3) 不同词语之间蕴含的空间关系语义也有可能不同。

例如: 河流之间关系在垂直时, 大都是相离的情况, 即非交叉垂直, 虽然通常人们不会查询两条分离的河流是否垂直。而公路之间垂直时, 相离和相交都有可能。

知识规则库中存储各特征类型之间进行空间操作所有可能的条件,该专题下不同的应用系统启用不同的条件,从而完成应用知识规则的定制。表2为与地理特征类型有关的部分空间关系自然语言规则库举例。

表2 知识规则存储模式表

Tab. 2 The storage schema of the knowledge rules

特征类型操作条件 (if)		空间关系操作 (then)	
主体特征类型	目标特征类型	空间关系词汇	空间查询操作算子
线性河流	线性河流	流入、汇入、注入	Flow_ into
		平行	Parallel
		垂直	Perpendicular And R11 =0
线性公路	线性河流	穿越	Cross
		平行	Parallel
		垂直	Perpendicular
线性河流	面状水体	流入、汇入、注入	Flow_ into
线性公路	面状建筑物	附近	Near
		沿途	Pass
线性河流	行政区域	途经	Intersect
		发源于	Start_ in
		终止于	End_ at or End_ in
		在...内部	Inside

表2中操作条件(即if条件)是空间对象所对应的特征类型,进行的操作(即then)是不同特征类型的空间对象间的空间操作。通过建立这样的知识规则库的方式,一定程度地实现了应用规则库的用户自定义,提高了应用系统的通用性。通过上面的规则,可以在具体查询中先作规则选取,而后进行具体查询。例如,要查询线性河流和行政区域之间的空间关系,首先,可以利用规则筛选出只适合这两种特征类型之间的空间关系自然语言描述的词汇,然后,再根据具体的操作算子进行空间关系判断。

4 基于地理语义的空间关系查询流程和推理

4.1 空间关系自然语言查询流程

在进行空间关系自然语言查询时,面向的是自然语言语句,首先,应该进行自然语言的分词解析处理,大体的流程应该如图3所示。由于现阶段的自然语言的分词和解析技术还不完善,而

本文的研究主要在于空间关系查询,以及自然语言表达与地理本体的对应,因此,暂且忽略自然语言分词和解析过程,以下的查询都基于这部分工作已经实现的假设。

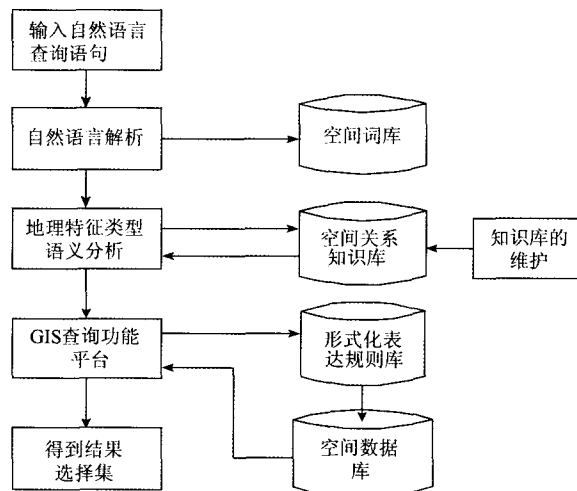


图3 空间关系自然语言查询流程

Fig. 3 Flow chart of natural language query on spatial relations

4.2 地理语义的空间推理

如前面所述,由于地理本体类型具有不同的空间语义,因此,在进行空间推理时可以结合空间物体的自身特性,将地理特征语义应用到空间推理中去,以获得更多的地理特征知识,促使更为完备的知识推理。在现实中,空间物体的地理特征所蕴含的知识有助于实现某些空间推理。例如,已知线性公路A、B、C,其中,A和B的空间关系是“平行”,A和C是“垂直”,则可以推出B和C之间的空间关系也是“垂直”的;已知一条线性河流和一个面状物体空间关系是“流入”,则可以推出这个面域是个湖泊、水库等;再例如查找一个城市的立交桥,可以在公路之间的“交叉”处查找;查找一条线性河流上的所有桥梁,即在这条河流和所有道路的相交处。

一般的,推理的过程可以表述为:已有的判断+事实⇒新判断。同理,与地理特征类型有关的空间推理的过程也可以表示如下:

已表示的事实+空间关系(自然语言描述的空间关系)和特征类型知识⇒新判断(包括事实和关系)。

对于一条产生式的空间关系推理可以如下表示:

例如:已知线状河流A(ID为FID1)和未知类型的地物B(ID为FID2)之间具有“流入(Flow_ into)”这样的自然语言空间关系,通过知识规则

库可知, 与河流 A 有“流入”关系的地物 B 的类型有两种: 线性河流和面状水体, 当 B 为线性河流时, 可以推断出该地物为河流; 当 B 为面状水体时, 可以推断出该地物为湖泊、水库或海洋。

此规则的产生式表达如下:

if FID1. type = “河流” and FID2. geometrytype = Line and Relation(FID1, FID2) = “流入” then FID2. type = “河流”

if FID1. type = “河流” and FID2. geometrytype = Area and Relation(FID1, FID2) = “流入” then FID2. type = “湖泊、水库”

又如: 已知道路 A (ID 为 FID1) 和 B (ID 为 FID2) 之间的空间关系为“平行”, B 和 C (ID 为 FID3) 之间的空间关系为“垂直”, 则 A 与 C 之间也是“垂直”关系。

此规则的产生式表达如下:

if Relation(FID1, FID2) = “平行” and Relation(FID1, FID2) = “垂直” then RelationWord = “垂直”

4.3 查询原型系统实例验证

通过上述结构本文设计了一个示例空间关系自然语言查询系统。由于本文主要的研究在空间关系自然语言地理特征语义理解上, 因此, 先跳过自然语言分词和解析过程, 可以作为以后工作的补充。为了避开空间关系自然语言查询语句解析过程, 主要采用鼠标左键引导处理或下拉框选择等方式获得空间对象和空间关系词汇等。

(1) 空间关系自然语言查询

例如, 要查询“公路 1 沿途的建筑物”, 采用“Pass”这个空间算子, 在建筑物这个地理特征要素类中进行结果筛选, 最终得到下面的结果(如图 4 所示)。

再例如, 要查询“河流 5 途经的行政区”, 由于行政区相对于建筑物粒度较大, 因此采取的相应查询规则也不同于上例, 查询的结果如图 5 所示。

同样, 通过计算两个物体之间的定量化指标, 还可以推导出适合描述这两个空间物体的自然语言词汇。例如, 在实验中, 通过计算公路 0 与公路 2 之间的定量化指标来核对它们满足哪些空间关系词汇, 结果表明可以用“垂直”来描述这两条公路之间的空间关系。

(2) 空间关系自然语言推理

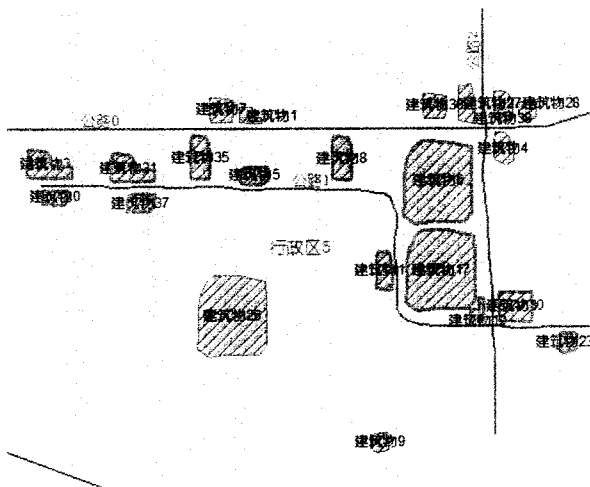


图 4 公路 1 沿途的建筑物

Fig. 4 Buildings that Road 1 “passes”

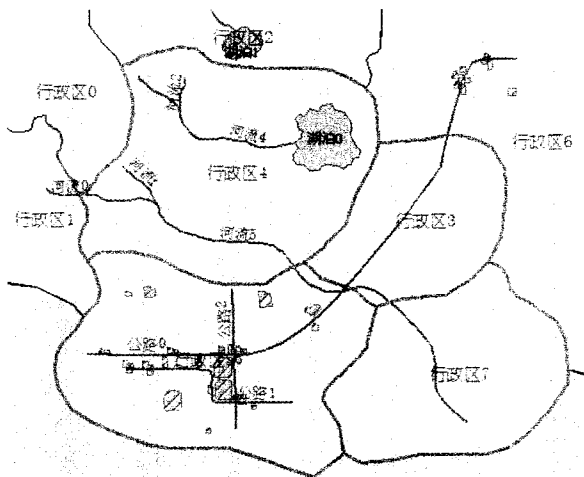


图 5 河流 5 途经的行政区

Fig. 5 Administrative regions which River 5 “intersects”

利用产生式的空间关系推理规则, 可以进行空间关系自然语言的推理。

例如, 在进行自然语言“河流 3 流入……”的查询时, 如图 6 所示, 已知查询主体为河流, 空间关系词为“流入”, 通过查询, 得出与“河流 3”满足“Flows_into”操作算子的地物类型只有“湖泊 1”。根据空间关系推理规则, 河流流入的地物(即查询谓体)包括线性河流和面状水体两类, 而“湖泊 1”是面状地物, 因此, 可以断定名为“湖泊 1”的地物是一个面状水体。

通过实验可知, 在现实世界中, 地理本体所蕴含的空间语义是不可忽视的。通过提取这些不同的语义知识, 可以为空间关系查询和推理提供服务。

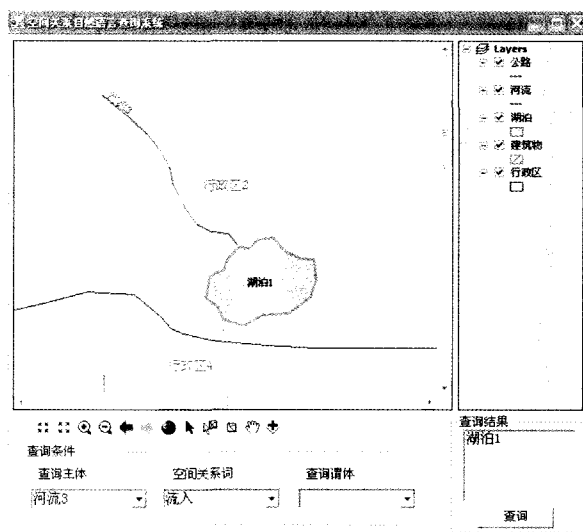


图6 自然语言“河流3流入……”的查询

Fig. 6 The query of “River 3 flows into ...”

5 结论与讨论

空间地物根据其本体类型的不同而具有不同的地理语义特征。地理语义对于地物之间空间关系的描述具有重要影响。人们在描述几何特征和关系相似的物体之间的空间关系时,受到地理语义的影响,所用的词汇也往往是不同的。本文考虑到地理本体的语义差异,总结了部分自然语言词汇在空间关系表达时所对应的特征类型知识规则,结合已有的表达定性的自然语言描述时的量化计算指标进行实验。实例研究表明,考虑地理特征类型的空间关系查询更加符合人们的地理知识表达特点。

由于空间关系的复杂性和丰富性,在进行地物之间空间关系的自然语言描述时,单一的空间关系词汇往往不能准确、完全地表达出两个物体之间的空间关系。本文中未考察模糊的空间关系表达方面的内容。在这方面,Xu 利用模糊决策树算法对线状物体之间空间关系的自然语言描述进行了研究^[23],可以作为本研究后续工作的参考。另外,本文关于地理特征类型的知识规则库,是在具体实验时加以实现的,不利于知识规则的显式表达和推理。目前,利用本体描述语言建立本体模型,在地物本体之间关系的形式化表达和推理方面具有很大优势,可以作为其下一步研究的内容。

参考文献:

- [1] Mark D M, Egenhofer M J. Calibrating the Meanings of Spatial Predicates from Natural Language: Line - Region Relations [J]. Proceedings of Spatial Data Handling, 1994, 9 (6): 538 - 553.
- [2] Mark D M, Comas D, Egenhofer M J, et al. Evaluating and Refining Computational Models of Spatial Relations through Cross-linguistic Human-subjects Testing. Frank A, Kuhn W. Spatial Information Theory-A theoretical Basis for GIS [C]. International Conference COSIT '95, Semmering, Austria: Lecture Notes in Computer Science, Berlin: Springer-Verlag, 1995, 988: 553 - 568.
- [3] Xu J, Mark D M. Natural Language Understanding of Spatial Relations between Linear Geographic Objects [J]. Spatial Cognition and Computation, 2007, 7(4): 311 - 347.
- [4] 许珺, 张晶, 司望利, 等. 线状物体空间关系的自然语言理解的双语比较[J]. 遥感学报, 2008, 12(2): 362 - 369.
- [5] Egenhofer M J, Shariff R. Metric Details for Natural-language Spatial Relations [J]. ACM Transactions on Information Systems, 1998, 16(4): 295 - 321.
- [6] Shariff R, Egenhofer M J, Mark D M. Natural-language Spatial Relations between Linear and Areal Objects: The Topology and Metric of English-language Terms [J]. International Journal of Geographical Information Science, 1998, 12(3): 215 - 246.
- [7] Nedas K, Egenhofer M J, Wilmsen D. Metric Details of Topological Line-line Relations [J]. International Journal of Geographical Information Science, 2007, 21(1): 21 - 48.
- [8] Xu J. Formalize Natural-language Spatial Relations between Linear Objects with Topologic and Metric Properties [J]. International Journal of Geographical Information Science, 2007, 21(4): 377 - 395.
- [9] Yao X, Thill J C. How Far is Too Far-A Statistical Approach to Proximity Modeling [J]. Transactions in GIS, 2005, 9 (2): 157 - 178.
- [10] Yao X, Thill J C. Neuro-Fuzzy Modeling of Context-Contingent Proximity Relations [J]. Geographical Analysis: An International Journal of Theoretical Geography, 2007, 39(2): 169 - 194.
- [11] 许珺. 关于线状地理特征空间关系的自然语言描述的形式化表达 [J]. 遥感学报, 2007, 11(2): 152 - 158.
- [12] 刘瑜, 高勇, 林报嘉, 等. 基于受限汉语的 GIS 路径重建研究 [J]. 遥感学报, 2004, 8(4): 323 - 330.
- [13] 马林兵, 龚健雅. 面向自然语言的空间数据库查询研究 [J]. 计算机工程与应用, 2003, 22: 16 - 19.
- [14] 吴静, 蔡砥, 王铮. 地理信息系统中自然语言查询的分

- 词处理与应用[J]. 地球信息科学, 2005, 7(3): 67-71.
- [15] 徐爱萍, 边馥苓. GIS 中文查询系统中 SQL 语句的形成[J]. 测绘科学, 2006, 31(5): 110-113.
- [16] 徐爱萍, 边馥苓, 黄源. 基于范例的 GIS 中文查询的 SQL 推理[J]. 武汉大学学报(信息科学版), 2007, 32(4): 369-371.
- [17] 张连蓬, 刘国林, 江涛, 等. 受限自然语言查询在 GIS 中的应用[J]. 测绘学院学报, 2002, 19(4): 283-286.
- [18] Rauschert I, Agrawal P, Fuhrmann S, et al. Designing a Human-Centered, Multimodal GIS Interface to Support Emergency Management, Advances in Geographic Information Systems[M]. ACM GIS 02, Washington, DC, USA: ACM Press, 2002: 119-124.
- [19] 刘瑜, 龚咏喜, 高勇, 等. 地理空间中的空间关系表达和推理[J]. 地理与地理信息科学, 2007, 23(5): 1-6.
- [20] Worboys M. Nearness Relations in Environmental Space[J]. International Journal of Geographical Information Science, 2001, 15(7): 633-651.
- [21] Egenhofer M J, Franzosa R D. Point-set Topological Spatial Relations[J]. International Journal of Geographical Information Science, 1991, 5(2): 161-174.
- [22] Egenhofer M J, Herring J. Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases, Technical Report[R]. Department of Surveying Engineering, University of Maine, 1994.
- [23] Xu J, Yao C. Formalizing Natural-language Spatial Relations Descriptions with Fuzzy Decision Tree Algorithm[C]. In Proceedings of SPIE, Volume 6420, Geoinformatics, Wuhan, China, October 28-29, 2006, 64201E-1-10.

Querying and Reasoning of Spatial Relations Based on Geographic Semantics

DU Chong^{1,2}, SI Wangli^{1,3}, XU Jun¹

(1. State Key Laboratory of Resources and Environment Information System, Institute of Geographic Science and Natural Resources Research, CAS, Beijing 100101, China; 2. College of Resources Environment & Tourism, Capital Normal University, Beijing 100037, China; 3. College of Geological Engineering and Geomatics, Chang'an University, Xi'an 710064, China)

Abstract: The description of a spatial relation is the reflection of human's cognition of spatial objects. It is not only affected by the topology and metric, but also affected by geographic semantics, such as the categories of geographic entities and contexts. Currently, the researches about language aspects of spatial relations mostly focus on natural-language formalization, parsing of query sentences, and natural-language query interface. In these researches, spatial objects are usually abstracted into geometric figures. However, geographic objects are not simple geometric points, lines or polygons. Spatial relation queries that just consider geometry property of objects cannot always meet human's habits of language and cognition. So, in order to get a sound answer according with human cognition in spatial relation query, we have to take the semantic aspect of geographic objects into account. Based on previous work on natural-language formalization and human-subject tests, the functions of the natural-language terms which describe spatial relations are calculated in this paper. Then, we sum up some knowledge about how spatial relations between two different geographic objects with different geographic semantics are represented in natural language, and design a knowledge base to represent these knowledge rules. Finally, using the rules in knowledge base and GIS platform, the prototype of a query system is implemented. The results show that spatial relation queries which take geographic semantics into account can be more in accordance with people's cognitive habits of geographic knowledge.

Key words: spatial relation; natural language; geographic ontology; semantics