

数字地名词典中的类型表达和管理

陈凯晨,林星,袁一泓,李润强,刘瑜*

(北京大学遥感与地理信息系统研究所,北京 100871)

摘要:在数字地名词典中恰当地表达和管理地名的类型知识,有助于数字地名词典快速、有效地处理地名相关查询。为此,该文首先从数字地名词典的应用需求出发,分析类型在数字地名词典中的作用,进而设计一个地名类型本体模型。该模型表达了地名类型之间的继承关系以及对空间关系的约束。在本体模型基础上,提出了相应的查询处理策略并进行了系统实现。

关键词:类型;本体;地名库;地名;匹配度量

中图分类号:P208 **文献标识码:**A **文章编号:**1672-0504(2009)05-0006-06

0 引言

地名源于人们对于地理要素或场所进行概念化及命名。虽然地名库的历史可追溯到 17 世纪的中国和 19 世纪的英国^[1],但直到近几十年,地名库才得到高度重视和广泛使用。由于 WWW 的发展和 GIS 的智能化,大量基于地名等文本形式表达和传输的地理空间知识需要建模和处理,地理信息检索(Geographical Information Retrieval, GIR)^[2]、地理语义网(Geospatial Semantic Web)^[3]等成为地理信息科学领域的研究热点。其中,地名的表达和管理是其关键技术,数字地名词典(Digital Gazetteer, 或称数字地名库)通过对地名的管理并提供相关服务,广泛用于历史数据管理、地理编码(Geo-coding)、空间定位、导航服务、地理信息检索等领域。Goodchild 等^[4]认为,一条地名库记录可以形式化为三元组,即<名称、地理范围、类型>,Hodge^[5]也强调分类模式在作为知识管理系统(Knowledge Organization System, KOS)的地名库中的重要性。综上所述,地名相关的类型信息表达了地名所蕴含的地理语义,是数字地名词典的重要组成部分之一。

分类是人类普遍用于认识现实世界的 3 个基本法则之一^[6]。在现实地理世界中,离散对象(Discrete Object)和连续场(Continuous Field)是人们表达地理现象时主要采用的两种方法或视角,其中:人们从不同的场中根据语义特征识别相应的地理要素(Geographical Feature)^[7,8],进而对相似的地理要素进行概念上的抽象、聚类后产生地理要素类型。因

此,Mennis^[9]将分类作为地理空间知识管理的两个重要元素之一。而 Janowicz 等^[10]认为,地名库的相关服务需要表达类型。值得注意的是,地名不仅包括对地理要素(Feature)的命名,还包括对场所(Place)的命名。地理要素通常是指位置和边界都可知觉并且明确的对象,如山峰、河流;而场所则具有含糊的位置或边界,与人们的经历有关,并且其表达通常以地理要素作为参考^[11,4]。在实践中,人们通常不严格区分地理要素和场所,并且大部分地名源于对地理要素的命名,因此,笔者认为地理要素是地名描述的对象,在地名库中所管理的地名类型,可视同于相应地物或地理要素的类型。

地名类型在地名库中的表达、组织的规范可称为地名分类模式(Geotaxonomies)^[12],其组织方式已由无结构的地理辞典(Geographic Thesaurus)逐渐演化成有序的层次结构地名库(Gazetteer),甚至是建立在复杂的本体模型的基础上。这说明在地名库管理中,需要更好地形式化表达地名类型所蕴含的丰富的地理语义。然而,人们对地名的表达或对地名表达的认同,受语言文化、地理位置和本身专业知识影响较大^[13,14],而且,地名库的应用目标也会影响地名的入库选择及其相关的分类模式,因此,不同的地名库可能会采用不同的地名分类模式。目前,ADL(Alexandria Digital Library)(<http://www.alexandria.ucsb.edu/gazetteer>)的 Feature Type Thesaurus(FTT)是在美国及其全世界约 500 万地名数据的基础上抽象概括出来的,也是最通用的、可扩展使用的要素类型词典和内容参考标准^[15]。其它地名

收稿日期:2009-06-01; 修订日期:2009-08-05

基金项目:国家 863 计划项目(2007AA12Z216、2007AA120502);国家自然科学基金项目(40701134、40771171);香港、澳门青年学者合作研究基金项目(40629001)

作者简介:陈凯晨(1985-),女,硕士研究生,主要研究方向为地理信息系统。*通讯作者 E-mail:liuyu@urban.pku.edu.cn

库,如英国的 geoXWalk (<http://www.geoxwalk.ac.uk>),通常根据具体需求对 ADL 进行简化、调整、细化和扩展后形成适用的地名分类模式。然而,即便以 ADL 为基础,建立大而全、普遍适用于世界各地的地名分类模式仍然十分困难,而且对个别地名库而言,采用如此庞大的通用模式,将带来不重要甚至冗余信息,从而降低检索效率。因此,在数字地名词典中表达和管理类型知识十分必要。本文探讨了如何在地名库中管理地名类型,并在处理查询时选择恰当策略,使得地名查询结果更加符合人们的常识性地理空间认知。

1 类型在数字地名词典中的作用

地名库以词条或记录为单位存储地名实体的地理属性。在地名库的词条中,通常至少包括地名文本、类型、地理位置 3 部分信息^[4]。在地名库底层存储中,地名类型是词条的重要内容;在地名库提供的查询界面上,地名类型是可选的限定条件,因此,需要充分挖掘和利用地名类型中蕴含的语义。

1.1 度量相似性

如前所述,基本地名类型产生于对语义相似的地名进行聚类,对低层地名类型的再次聚类概括成更高层次的地名类别,这样逐层向上抽象,从而形成层次分明的类型树。不难发现:上层类别比下层类别的抽象程度高、所指称范围广,同属于一个上层类别的几个下层类别存在某种相近的语义。该类型组织结构可用来处理对用户查询的识别(Identification)、解读(Interpretation)和扩展(Expansion)。例如,“餐馆”与“咖啡馆”同为“餐饮场所”,具有相近的语义。当用户查询“餐饮场所”时,地名库需要将“餐饮场所”识别为“餐馆”和“咖啡馆”等,将相关结果同时反馈给用户;而当用户查询“酒店”时,地名服务可以将“酒店”扩展为“酒店”和“招待所”等,并标识不同的相似度,使查询结果更加丰富。

1.2 地名排歧

在地名库的应用中,地名二义性及其排歧是个重要问题,而类型信息可用于解决地名库应用中地名的部分二义性问题。地名的二义性,即相同的地名在不同的语境中指代的是不同的地理位置或范围,或者表达的是不同类型的地理要素。前者如“朝阳”(北京市朝阳区或辽宁省朝阳市的简称),后者如“黄山”(黄山市或黄山山峰)^[16]。在后一种情况下,类型可用于表达并区分不同的语境,从而消除地名的二义性。

1.3 约束空间关系

空间关系是地理空间知识表达中的重要元素。为描述地理位置或地理现象,人们需要用空间关系来关联有关地名,并在此基础上加工处理地理知识。常用的二元空间关系主要包括度量关系、方位关系和拓扑关系。类型对空间关系的约束作用有助于精确表达相关空间关系的地理语义,并在查询中利用类型简化查询操作。

(1) 尺度、类型、空间分布等语境信息对于定性空间关系而言尤其重要,影响人们对邻近关系的判断^[17]。隐含在地名类型中的相关信息,如点、线或面等几何属性以及其他地理语义,与人们地理空间认知紧密耦合而影响空间关系的表达和推理^[18]。类型决定了地物的拓扑维度,而某些类别的空间关系要求地物有特定的拓扑维度。如:行政区类地物一般为 2 维,不存在平行关系;山峰类地物为 0 维,不存在包含关系。而一些特定类型的关系,是在普通空间关系基础上增加类型约束。如“首都”关系,不仅蕴含了被包含拓扑关系,同时要求参与地物类别为城市和行政区。

(2) 地名类型隐式表达了相关要素的尺度、活动类型等信息,这些信息影响了对定性度量关系(如“远”和“近”等)的处理。例如,省级行政区划地名所蕴含的尺度通常大于县级行政区划地名的尺度,在这两个尺度中“近”等度量关系对应的绝对距离存在差异。除尺度因素外,定性度量关系还与地名类型中隐含的活动相关,如在“某地点附近(空间关系)的餐馆/度假村/公交车站(类型)”这样的描述中,邻近关系所涵盖的绝对空间范围不尽相同。通常,用餐活动对应的“近”往往小于旅游度假活动对应的“近”^[19]。因此,在地名查询中,需要根据给定的地名类型设定定性度量关系对应的绝对数值范围。

2 地名类型信息的 OWL 描述

在信息系统领域,本体是对共享概念模型形式化的明确表达^[20]。在地理信息检索领域,地理学者们提出基于地名或场所的本体模型^[16,21-23]。本体模型的优点在于:能够对本体之间的丰富关系进行显式的形式化表达,以用于相关推理。本文采用 OWL(Web Ontology Language)语言对类型本体进行形式化建模。OWL 是一种本体描述语言,具有简单且有概念定义的词汇表、以上下位定义的分类树以及描述能力齐全的特点。用 OWL 描述类型本体,有利于使用地名类型的相关知识结构或者规则、

提高地名信息检索能力。

广义地名是对一个场所的文本描述,可分为 3 类:地址、结构化广义地名和非结构化广义地名^[16]。地址是对地理空间位置或者范围用 IP 地址、邮政编码和电话号码等方式的直接编码;结构化广义地名是基于普通地名和空间关系的描述,包含参考地名、空间关系和类型约束 3 个要素;非结构化广义地名则是除地址和结构化广义地名以外的地名。与广义地名不同的是,普通地名是对实体的地理位置或范围的命名。因此,从地名分类的角度出发,普通地名类型相当于地理要素或地物类型(图 1)。

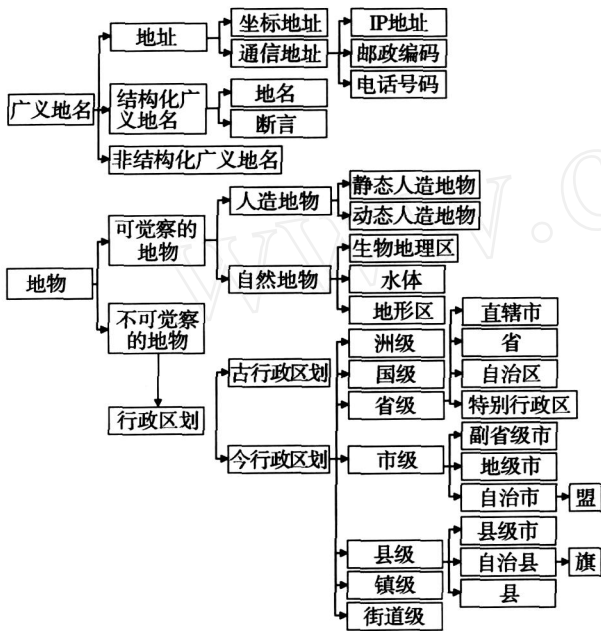


图 1 广义地名类型层次结构
Fig. 1 Hierarchical structure of the types of general place names

在地名类型本体模型(图 1)中,地名类型被组织成一棵类型树,叶节点类型代表直接从地名数据中提取出的基本地名类型,由叶节点类型概括产生非叶节点类型,这样的继承顺序形成相应的父节点和子节点。理想情况下,对于一个父节点,其子节点具有联合完备且互不相交的特点,那么父节点类型可以表示为所有子节点的并(owl:unionof)。

图 1 的本体模型是通过本体之间的继承关系组织而成,如我国省级行政区主要包括直辖市、省、自治区和特别行政区,其中省级行政区与直辖市之间的层次关系采用 OWL 语言描述(图 2)。其中图 2a 表示直辖市与省级行政区的子类关系,图 2b 表示我国的省级行政区只包括 4 种情形。

OWL 不仅刻画了地名类型的层次组织关系,还可以描述与之类型相关的空间关系,如“首都”关系(图 3)。图 3 表达了“首都”关系的以下语义:1) 蕴含

了包含关系;2) 定义域和值域分别为国家和直辖市;3) 该关系是 1 1 的函数关系,即一个国家有且仅有一个首都。

```
<owl Class rdf ID="直辖市" >
  < rdfs subClassOf rdf resource = "# 省级行政区"/ >
  < owl disjointWith rdf resource = "# 省"/ >
  < owl disjointWith rdf resource = "# 自治区"/ >
  < owl disjointWith rdf resource = "# 特别行政区"/ >
</owl Class >
```

(a)

```
<owl Class rdf ID="省级" >
  < owl equivalentClass >
    < owl Class >
      < owl unionOf rdf parseType = "Collection" >
        < owl Class rdf ID = "直辖市"/ > < owl Class rdf ID = "省"/ >
        < owl Class rdf ID = "自治区"/ > < owl Class rdf ID = "特别行政区"/ >
      </owl unionOf >
    </owl Class >
  </owl equivalentClass >
</owl Class >
```

(b)

图 2 中国省级行政区类型的 OWL 语言描述
Fig. 2 OWL descriptions of relations among types of Chinese provincial units

```
<owl ObjectProperty rdf ID="包含" >
</owl ObjectProperty >
<owl FunctionalProperty rdf about = "# 首都" >
  < rdfs range rdf resource = "# 直辖市"/ >
  < rdfs domain rdf resource = "# 国家级"/ >
  < rdfs subPropertyOf rdf resource = "# 包含"/ >
</owl FunctionalProperty >
```

图 3 对空间关系的 OWL 语言描述
Fig. 3 OWL descriptions of spatial relations

3 地名查询中的类型处理策略

为了在查询处理时充分考虑地名语义语境信息,更好地满足用户对地名信息检索的需求,需要在建立类型本体模型的基础上,制定恰当的查询处理策略。

3.1 类型匹配度计算及相关的检索策略

通常,在地名数据库的一个数据表中存储同一类型的地名数据。因此,在理想的情况下,在地名数据库中存储全部的叶节点类型的地名数据,可以最大限度地避免出现地名数据存储的冗余现象。然而在实践中,有些叶节点类型(如内流河、外流河)在日常生活中不常用;而有些非叶节点类型较常用。例如,人们通常使用河流而非内流河、外流河等更细致的专业类型交流地理空间信息。本文称这些非叶子节点对应的地名类型为基本类型。在地名数据管理中,通常依据基本类型组织地名记录。因此,当在查

询中指定的类型是比基本类型更粗略或者更精细的范畴时,需要在类型树的基础上,计算查询指定类型与地名库中表达类型的匹配度(Fitness)。

类型树中上下层节点的父子继承关系揭示了其间的语义包含关系。该语义关系可用于待查询类型对应记录缺失时的查询处理。在图 4 所示的类型树中,当待查询地名类型为“公园”时,如果地名库中显式存储了公园对应的地名实例,可以直接返回,并且匹配度为 1。而当公园类型对应数据表缺失时,可以得到其子类(如植物园)、父类(如休闲场所)、兄弟类(如度假村)、甚至更远的亲戚节点对应的类型(如舞厅),根据类型计算匹配度并返回相应的地名记录。

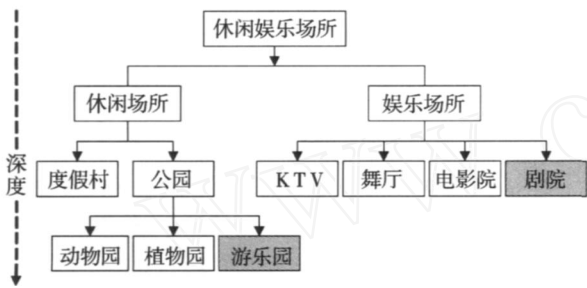


图 4 一个类型树示例
Fig. 4 An example of type tree

根据人的认知常识,在上例中,匹配度从大到小应为:子类 > 父类 > 兄弟类 > 远亲节点对应的类型。根据该顺序,可以看出匹配度具有非对称性,即 $MD(x, y) \neq MD(y, x)$, 其中 MD 表示两个类型的匹配度,其取值范围为 $[0, 1]$ 。假定待查询类型对应节点为 m , 拥有地名实例的类型对应节点为 n , 为了计算匹配度,一种可行的策略是在匹配度的计算中,引进与类型继承方向相关的权重因子。类型树上, m 和 n 之间存在深度最小的共同祖节点 (Least Common Ancestor) 和最短路径。例如,在图 4 中,动物园类型和度假村类型的最小共同祖节点为休闲场所,从动物园类型到度假村类型的最短路径为:动物园 -> 公园 -> 休闲场所 -> 度假村。为了基于该路径计算匹配度,首先定义两个相邻的具有子类/父类关系的节点相似度 (Similarity) 函数 SD , 满足 $SD(x, x) = 1$ 和 $SD(x, parent(x)) = SD(parent(x), x)$, 即相似度满足自反性和对称性。

在相似度的基础上,考虑最短路径上两个节点之间继承关系的方向性,两个节点的匹配度可以定义为:

$$MD(m, n) = [(1 - \alpha) \cdot SD(m, x_1)] \cdot [(1 - \alpha) \cdot SD(x_1, x_2)] \cdot \dots \cdot [(1 - \alpha) \cdot SD(x_{j-1}, x_j)] \cdot [(1 - \beta) \cdot SD(x_j, x_{j+1})] \cdot [(1 - \beta) \cdot SD(x_{j+1}, x_{j+2})] \cdot \dots \cdot [(1 - \beta) \cdot SD(x_k, n)]$$

其中: $MD(m, n)$ 为类型 n 对类型 m 的匹配度, $m, x_1, \dots, x_j, \dots, n$ 表示从 m 到 n 的最短路径,规定从 m 到 x_j 为“上行路径”,从 x_j 到 n 为“下行路径”。

α, β 为体现类型继承方向对匹配度的影响而设立,在 $(1 - \alpha) \cdot SD(x_1, x_2)$ 中, $(1 - \alpha)$ 表示节点 x_1 是节点 x_2 的子节点时的权重;而在 $(1 - \beta) \cdot SD(x_j, x_{j+1})$ 中, $(1 - \beta)$ 表示节点 x_2 是节点 x_1 的子节点时的权重影响因子。根据前面的讨论,通常有 $\alpha < \beta$, 即匹配度不具有对称性。假定 $SD(x, parent(x)) = \beta$, 则根据匹配度公式以及“动物园”到“度假村”的最短路径,当待查询地物类型为动物园时,度假村类型地名的匹配度为 $\alpha^2 \cdot \beta \cdot \beta^3$ 。

3.2 对类型名的处理

类型名特指在地名文本中,类似于后缀词,起到标示地名类型作用的文本,如“广东省”(省)和“黄河”(河)。并非所有的地名类型都有上述的类型名,通常,邮编、电话号码和 IP 地址等广义地名类型不存在这样的类型名。在某些地名表达,尤其是行政区划类型的地名表述中,通常省略类型名,如用“广东”表达“广东省”而不会产生歧义。因此,在地名服务中,必须将这些需要特别关注的类型名作为地名知识,存储在类型本体中。在处理“地名为湛江、类型为行政区划”的查询时,首先判断该地名类型不属于邮编、电话号码或 IP 地址类型后,再判断“行政区划”类型是否带后缀名;由于在类型本体库中存储了“行政区划”类型的后缀名(“市”、“省”、“县”等),于是为获得“湛江”的全称和简称而进行后缀增删,处理结果为:“湛江市”、“湛江省”、“湛江县”等;最后,在地名库中,对“湛江”、“湛江市”、“湛江省”、“湛江县”等进行检索可得结果。具体处理流程如图 5。

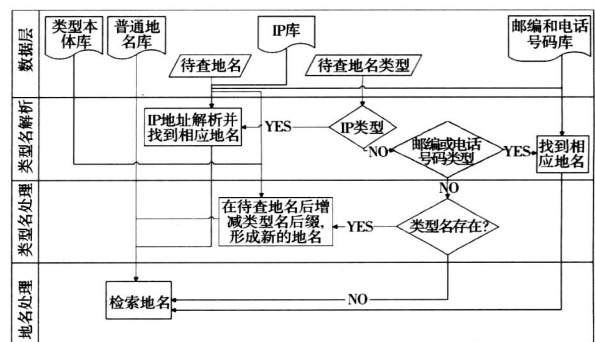


图 5 考虑类型名的地名检索
Fig. 5 Processing diagram of dealing with type postfixes

3.3 对特定空间关系的处理

类型对地名有关的空间属性有约束作用,这种约束分拓扑维度约束和语义约束两方面。前者如包含与被包含关系,后者则包括“首都”、“流入”等关

系。在 本体模型中,需要显式表达这些关系所具有的维度约束,如包含关系定义域对象的维度约束为 2;或直接描述关系所蕴含的类型约束,如“首都”关系的定义域和值域类型分别为“城市”和“行政区”(图 3)。因此,当地名查询条件包含了上述关系时,可以直接根据关系限定符合条件地名的类型,或者直接判定为非法查询,如“流入湖北的山峰”。

此外,定性度量空间关系取决于描述关系的语境有关的空间尺度,如校园尺度上的“远”未必比城市尺度上的“近”的绝对距离更大^[18,23]。因此,邻近关系随参考地名类型和目标地名类型而变化,不便显式存储在数据库中,应采用即时计算的方法,除了要自身地理空间范围的大小以及要素间绝对距离之外,在计算时引入类型影响权重因子。在数据库中显式存储的是类型影响权重因子,例如可以规定当参照地物类型为“组织机构”时,“附近”关系对应的绝对距离为 1 km,而当参照地物类型为“城市”时,相应的绝对距离为 10 km。

4 系统实现与应用实例

4.1 系统实现

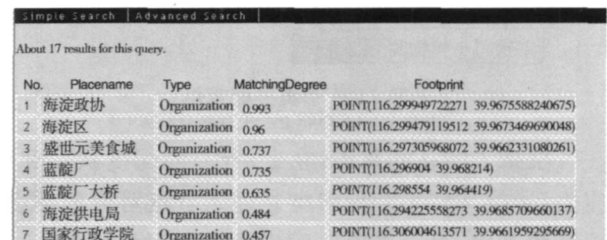
地名服务系统中除了管理大量的地名记录外,还需要管理地名类型及对应的地理语义。考虑到存储地理范围及其进行相关空间计算的效率问题,本文采用 Post GIS 存储具体的地名信息。在 Post GIS 一个数据表中,存储同一种类型的地名数据记录。此外,在 Post GIS 地名数据库中,建立并维护元数据信息表,记录数据表对应的类型等信息。对类型本体,本文通过 Protégé 用 OWL 对地名类型及其类型名进行建模、表达和存储,具体如下:1) 通过简单继承关系,建立基本的地名类型树;2) 采用对象型属性为地名类型建立更多空间或非空间的关系,并且定义关系所蕴含的约束;3) 采用数值型属性存储地名类型的类型名,如“省”(地名类型)有“省”、“直辖市”和“自治区”等具体类型名。

在地名及类型信息管理的基础上,可以向用户提供地名检索服务,这也是数字地名词典的功能层需要实现的重要目标。可以通过两种方式实现该目标,即应用程序接口(API)和 Web 查询界面(Web Interface)。其中 API 可用于查询数字地名词典和类型本体的描述信息、类型本体之间的继承关系和相似度等;而 Web 接口以地名、类型和空间约束等作为查询条件,可以处理各种含类型的检索;同时还利用本体库中的类型知识与用户进行查询时的交互,

便于用户随时选择合适的地名类型作为查询条件。

4.2 应用实例

笔者通过具体的查询实例(“海淀人大附近的组织机构”)说明广义地名的查询处理流程和类型信息的使用。在执行查询的过程中,系统需要对查询条件做如下处理:1) 对“组织机构”进行类型拓展,得到如“学校”、“政府机关”等子类型,组成类型列表;对于列表中的子类型,依据 3.1 节中算法赋予不同的匹配度值 MD_i 。2) 得到参考地物“海淀人大”的 Footprint 信息。3) 对类型列表进行循环,对于每一子类型的全部地名数据记录进行 Near 函数的返回值判定,判断其是否在北京大学附近,并返回“附近”关系的匹配度 MD_p ;其中附近的计算考虑了参照地物(即“海淀人大”)和待查询地名类型(即“组织机构”的类型特征)。4) 对于每个符合条件的地名记录,取 MD_i 和 MD_p 的最小值,得到记录条目的最终匹配度并返回查询结果集(图 6)。



No.	Placename	Type	MatchingDegree	Footprint
1	海淀政协	Organization	0.993	POINT(116.299949722271 39.9675588240675)
2	海淀区	Organization	0.96	POINT(116.299479119512 39.9673469600048)
3	盛世元美食城	Organization	0.737	POINT(116.297305968072 39.9662331080261)
4	蓝靛厂	Organization	0.735	POINT(116.296904 39.968214)
5	蓝靛厂大桥	Organization	0.635	POINT(116.298554 39.964419)
6	海淀供电局	Organization	0.484	POINT(116.29422558273 39.9685709660137)
7	国家行政学院	Organization	0.457	POINT(116.306004613571 39.9661959295669)

图 6 查询结果示例
Fig. 6 An example of query results

5 结语

地名类型蕴含了丰富的地理语义,在地名库分歧、扩展查询等方面发挥了重要作用。地名库中的类型知识可采用本体方法建模,因为本体方法能够形式化表达语义并且支持推理。本研究采用 OWL 语言对地名类型本体进行建模,显式表达了类型的继承关系、地名类型对于空间关系的约束等语义知识。实例研究表明,在数字地名词典中管理类型知识以及相关的地理语义,能够使得地名查询更加符合常识性地理空间认知。

由于类型知识源于人们的地理空间认知,在地名库中管理类型知识,既要考虑可实现性,又要考虑计算结果符合人们的认知习惯,应选择合适的地名匹配度计算方法。人工判读测试是检验匹配度计算结果是否符合人们认知结果的重要方法,本研究将基于该方法深入探讨和评判匹配度计算结果。除在数字地名词典和地名服务中表达和管理类型知识外,基于地名类型的互操作、自动推理乃至基于语义

网的地理信息检索等方面的应用将成为未来的研究热点。

参考文献:

- [1] HARGETT J M. Song dynasty local gazetteers and their place in the history of Difangzhi writing[J]. Harvard Journal of Asiatic Studies, 1996, 56:405 - 442.
- [2] JONES C B, PURVES R S. Geographical information retrieval [J]. International Journal of Geographical Information Science, 2008, 22(3):219 - 228.
- [3] EGENHOFER M J. Towards the semantic geospatial Web[A]. Proceedings of the Tenth Association for Computing Machinery International Symposium on Advances in Geographic Information Systems[C]. VA:McLean, 2002. 1 - 4.
- [4] GOODCHILD M F, HILL L L. Introduction to digital gazetteer research[J]. International Journal of Geographical Information Science, 2008, 22(10):1039 - 1044.
- [5] HODGE G. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files[M]. CO: Boulder, Council on Library and Information Resources, 2000.
- [6] COAD P, YOURDON E. Object Oriented Analysis[M]. NJ: Prentice Hall Press, 1999.
- [7] GOODCHILD M F, YUAN M, COVA T J. Towards a general theory of geographic representation in GIS[J]. International Journal of Geographical Information Science, 2007, 21(3):239 - 260.
- [8] LIU Y, GOODCHILD M F, GUO Q, et al. Towards a general field model and its order in GIS[J]. International Journal of Geographical Information Science, 2008, 22(6):623 - 643.
- [9] MENNIS J L. A conceptual framework for incorporating cognitive principles into geographical database representation[J]. International Journal of Geographical Information Science, 2000, 14(6):501 - 520.
- [10] JANOWICZ K, KEBLER C. The role of ontology in improving gazetteer interaction[J]. International Journal of Geographical Information Science, 2008, 22(10):1129 - 1157.
- [11] BENNETT B P, AGARWAL P. Semantic categories underlying the meaning of place [A]. COSIT 2007 [C]. Australia: Melbourne, 2007. 78 - 95.
- [12] HASTINGS J T. Automated conflation of digital gazetteer data[J]. International Journal of Geographical Information Science, 2008, 22(10):1109 - 1127.
- [13] EGENHOFER M J, MARK D. Naïve Geography [A]. COSIT 1995 [C]. Austria: Semmering, 1995. 1 - 15.
- [14] MARK D M, TURK A G. Landscape categories in Yindjibarndi: Ontology, environment, and language [A]. COSIT 2003 [C]. Switzerland: Ittingen, 2003. 28 - 45.
- [15] HILL L L. Feature type thesaurus [DB/OL]. In Alexandria Digital Library Project, Available online at: http://www.alexandria.ucsbc.edu/gazetteer/FeatureTypes/FTT_metadata.htm. 2002.
- [16] 刘瑜, 张毅, 田原, 等. 广义地名及其本体研究[J]. 地理与地理信息科学, 2007, 23(6):1 - 7.
- [17] YAO X, THILL J C. How far is too far? — A statistical approach to context-contingent proximity modeling[J]. Transactions in GIS, 2005, 9(2):157 - 178.
- [18] 刘瑜, 龚咏喜, 张晶, 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 - 7.
- [19] CAI G. Contextualization of geospatial database semantics for human-GIS interaction [J]. GeoInformatica, 2007, 11:217 - 237.
- [20] BORST W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse[D]. University of Twente, Enschede, 1997.
- [21] CHAVES M S, SILVA M J, MARTINS B. GKB-Geographic Knowledge Base[R]. Departamento de Informatica, Faculdade de Ciencias da Universidade de Lisboa, 2005.
- [22] NAGEL S. An ontology of German place names [A]. Corela: Cognition, Representation, Language, Numéros spéciaux [C]. Le traitement Lexicographique Des Noms Propres, 2005.
- [23] PURVES R S, CLOUGH P, JONES C B, et al. The design and implementation of SPIRIT: A spatially aware search engine for information retrieval on the Internet [J]. International Journal of Geographical Information Science, 2007, 21(7):717 - 745.

Representation and Management of Type Information in Digital Gazetteers

CHEN Kai - chen, LIN Xing, YUAN Yi - hong, LI Run - qiang, LIU Yu

(Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China)

Abstract: Digital gazetteers provide information about place names and corresponding services. Types of place names play an important role in digital gazetteers. With the representation and management of type information, a gazetteer would become more effective so that it can bridge the semantic gap between user requirements and the provided functions. According to the requirements of digital gazetteers associated with type information, an ontological model of type information in digital gazetteers is developed. The model explicitly represents the type hierarchy and constraints on spatial relationships. Following the ontological model, several execution strategies, such as measuring fitness, are designed to support queries about the digital gazetteer. A prototype software tool implements the strategies and demonstrates the feasibility of the proposed approaches.

Key words: type; ontology; digital gazetteer; place name; fitness measurement