

# 街道的城市地址编码模型与实验

于焕菊<sup>1</sup>, 齐清文<sup>2</sup>, 李云岭<sup>1</sup>

(1. 山东科技大学 测绘科学与工程学院, 青岛 266590; 2. 中国科学院地理科学与资源研究所, 北京 100101)

**摘要:** 地址编码是指将地址映射成地理坐标的过程, 它提供了一种把描述性地理位置信息转换成GIS地理坐标的方式。本文分析了当前标准规范中的地址编码规则, 认为街道与小区等院落实体作为同级别的区域限定物弱化了街道的空间特征, 并且在现有地址编码规则基础上提出了将街道前置作为区域限定物的地址编码改进方案; 进一步, 为了体现街道作为各类地理实体的空间参考基础, 本文按照街道的前进方向, 定义了街道线性空间中的地理实体链表组织模型, 实现了将相互独立的地址编码实体在街道中的串联组织, 从而为今后地址查询中的参考定位提供了模型支持; 最后, 给出了街道两侧不同类型地理实体纳入街道地址编码实体链表的实例说明。

**关键词:** 数字城市; 街道; 地址编码; 地址匹配; 标准地址

**DOI:** 10.3724/SP.J.1047.2013.00175

## 1 引言

地址编码是指将地址映射成地理坐标的过程, 故也称地理编码。它提供了一种把描述成地址的地理位置信息转换成可被用于地理信息系统的地理坐标的方式<sup>[1-2]</sup>。

美国早在20世纪70年代就建立了全国的地址编码标准, 成功地应用于Tiger数据库系统, 在历次全国人口普查统计中发挥了巨大作用<sup>[2]</sup>, 地址编码技术已广泛应用于行政管理、商业分析与选址、疾病控制等领域<sup>[3]</sup>。在我国地址编码被看作是数字城市地理空间框架的基础之一<sup>[4-5]</sup>。

地址编码有点模型(Point Geocoding)、面模型(Parcel Geocoding)和街道模型(Street Geocoding)3种。点地址编码是关于兴趣点、标志物等点状地名地址的编码模型; 面地址编码是对院落、地块、行政区、邮政区块等面状空间特征的编码模型; 街道地址编码在世界上是应用最为普遍的编码模型, 旨在解决“街道名+门牌号”类的地址匹配问题。点模型和面模型的匹配精度高但效率较差, 街道模型采用插值技术, 有很高的匹配效率, 但精度较差<sup>[6]</sup>。

我国门址系统稳健性差及城市建筑变化过快, 门牌的缺号、跳号、重号现象比较严重, 故建立门址

的街道地址模型非常困难, 于是大家将地址地名数据库建设重点放在点、面地址模型两个方面, 并且需要花费大量的人力物力进行地址地名普查。鉴此, 须建立一套标准规范来指导全国地址地名数据库。这些标准规范定义了地址地名的三级层次描述规则, 将街巷与小区看作是行政区划之后的第二级区域限定物, 作者认为这样弱化了街道对地址地名的区域限定作用。为此, 以强化街道作为地址地名区域限定作用为出发点, 本文对现有地址地名描述规则进行了改进, 提出了4级描述规则, 将街道单独作为一级并且前置, 在此基础上又对地址地名数据的组织加入了街道元素, 从而将离散的点、面状地址地名数据用街道实现了空间上的关联。

## 2 传统的地址地名描述规则

为了保证不同区域相同名称的地理实体有唯一的编码, 通常要在编码中增加区域限定。区域限定除了行政区划外, 通常还有邮政区域、人口普查区域等。我国区域限定通常将省、市、县(区)、乡(街办)等逐级行政区划作为首级区域限定, 相关标准也将小区、街道同样作为第二层次的区域限定<sup>[7-9]</sup>。国家标准“数字城市地理信息公共平台地名/地址编码规则”中地址编码规则如图1所示<sup>[8]</sup>。

收稿日期: 2012-11-30; 修回日期: 2013-01-04.

作者简介: 于焕菊(1977-), 女, 山东阳谷人, 博士研究生, 主要从事城市地址编码理论与应用、基础地理信息数据库建设及数字测图理论等研究。E-mail: xiaoyuer200178@163.com

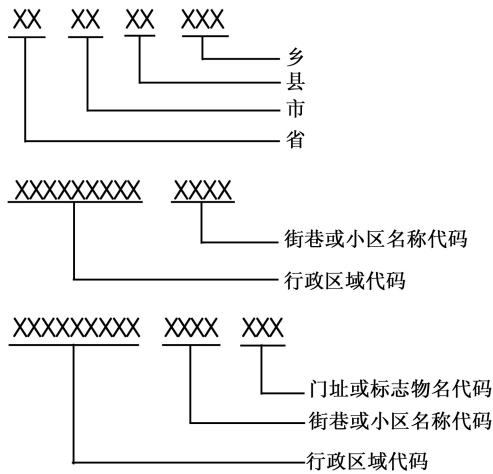


图1 门楼址或标志物名的编码结构

Fig.1 The current address code structure

在后续颁布的“地名地址数据规范(试行稿)”以下简称“规范”,进一步细化了地址地名数据库建设中的编码规则和属性结构。两个标准虽各有侧重,但在地址编码数据组织的做法基本一致,即将标准地址采用分段组合的方式描述,整个标准地址地名由“行政区域、基本区域限定物、局部点位置”3大类要素构成,描述规则<sup>[8]</sup>如下:

<标准地址>::=<行政区域名称>[基本区域限定物名称][局部点位置描述]

其中:

<行政区域名称>::=<洲级><国家级><省级><[地区级]<县级><乡级>[行政村级]

<基本区域限定物名称>::=<街>|<巷>|<居民小区>|<自然村>

<局部点位置描述>::=<门(楼)址>|<标志物名>|<兴趣点名>

可见,该编码规则中,将街道与小区作为同级别的区域限定物,二者是“或”的关系,对门址或标志物名的编码而言,前面属于某个小区或属于某条街道,二者取其一。街道作为贯穿城市的线状实体,有些不仅仅跨越几个小区,在城市中甚至跨越几个街办,在空间关系上与面状的行政区划有着交叉关系,若从小区的通达连接上看,小区出入口可以说是属于某条街道。

### 3 街道描述规则的改进

Clodoveu A 认为,加强区域限定可提升地址编码匹配的精度<sup>[10]</sup>。街道作为市民通行、商业活动、休闲娱乐的重要场所,承载着城市绝大部分地址地

名信息<sup>[11]</sup>。若将街道看作容器,那么街道两旁的地址编码实体就是这个容器中的一份子。为充分发挥街道作为区域限定物的作用,本文在原地址描述规则的基础上,提出了标准地名地址由“行政区域、有名街巷、其他区域限定物、局部点位置描述或名称”4大类要素构成的方案,定义如下:

<标准地址>::=<行政区域><有名街巷>[其他区域限定物名称][局部点位置描述或名称]

其中:

<行政区域>::=<洲级><国家级><省级><地区级><县区级>[街办、镇][行政村级]

<有名街巷>::=<街>|<巷>

<其他区域限定物名称>::=<大型单位>|<居民小区>|<自然村>

<局部点位置描述或名称>::=<门(楼)址>|<标志物名>|<兴趣点名>|<其他名称>

说明:

(1)<行政区域>中,由于街办和行政村可选,不再是必需的。

(2)<有名街巷>作为次于行政区划的第二级区域限定物,指那些具有名字的街道、小巷、胡同等。按照通达性原则,城市中院落、标志物、兴趣点都经由街道到达,因此,对那些不直接临街地址点,可选择主要连通的街道进行编码。

(3)在该定义中,街道在城市内统一编码,不再受街办和行政村等低级别行政区划范围的约束;

(4)地址编码的唯一性由<行政区划>、<有名街巷>与<其他区域限定物>组合编码来保障;

(5)<其他区域限定物>在原来区域界定物的基础上剔除街道,保留小区,增加大型单位如学校、医院、大型企业等。

改进后的地址地名描述规则如图2所示。

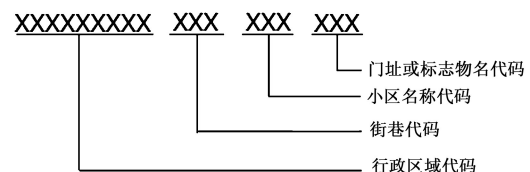


图2 顾及街道的门楼址或标志物名的编码结构

Fig.2 The revised address code structure

## 4 街道地址编码的组织模型

### 4.1 街道地址地名链表定义

空间参考系统是地理实体与几何坐标之间关

系的函数<sup>[12]</sup>,空间参照系统有多种,其中,线性参照系统是从特定参照点沿着线性要素的相对距离来描述空间位置的坐标系统<sup>[13]</sup>。传统地,不同类型的地址编码只是以地理坐标系为参考系统,编码实体相互之间并没有选择其他的地理实体作为参考系统<sup>[14]</sup>。街道作为典型的线性参考空间,参照原点可以是街道的起始点、交叉路口,也可以是分布在街道两侧的地址地名点,根据当前位置选择附近标志性地物作为参照物对寻址效率更加有效<sup>[15]</sup>。

将街道的区域限定地位提升到小区和单位院落区域限定物之前,解决了地址编码的合理性问题,但要充分发挥街道作为地名地址的空间参考和串联作用,需要构建街道的数据组织模型。我们将地址地名点位分为直接关系和间接关系两类。那些分布在街道两侧的点位属于某条街道,是直接的关系;另外,一些虽不沿街分布,但需要经由街道到达,即这些点在街道上有入口点,这是间接的关系。我们将直接的地址地名点和间接地址地名在街道上的入口点统一组织,沿着街道的前进方向分两侧贯续串联,便形成了街道的地址地名点位链表,组织方式如下:

```
<街道起点(道路交叉口)>
{
  [道路交叉口][门址、名称][人工标志物]
  [地形地物兴趣点]
}
<街道终点(道路交叉口)>
```

说明:

(1) 街道起点和终点:是沿街分布地址点序列的起止点,不可或缺。

(2) {}中内容:为街道中的地址点序列,沿街道起点贯序排列到终点。沿街地理对象直接取其几何中心作为定位点,不沿街的地理对象则需要向街道投影,投影点是经由街道抵达该对象的拐点。

当地址点与几条街道相连通时,若地址点处有门牌号,则一般选择门牌号中的街道进行投影;若地址点处没有门牌号,则可按照连通距离选择最近的街道作为投影街道。

(3) 门址、名称:门牌号与相对应的名称可并列出现,如“前湾港路579号”与“山东科技大学”同时出现。

(4) 人工标志物:指标志性的构筑物 and 建筑物,包括雕塑、大屏幕、公交站点等有名标志物。

(5) 地形地物兴趣点:指标志性地形地物在街道上的投影点,如广场、公园、胡同、桥梁、河流等。

## 4.2 地址地名链表表示实例

图3解释了某地东关大街上地理对象的分布,将间接关系地址地名的入口点用空心圆点表示,实心点是分布在街道两侧的地址地名点,  $A_n$  表示东关大街北侧的地址地名点、 $B_n$  南侧地址地名点。

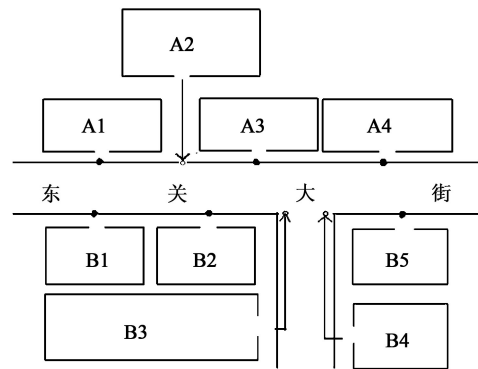


图3 直接和间接地址地名点位分布

Fig.3 The distribution of direct/indirect geocoding points

图4是沿东关大街两侧分别组织的地址地名点位链表,  $Start$  和  $End$  是东关大街的起、止点,在每个节点中除了节点自身的编码和属性外,还保留着邻接节点的编码作为指针。

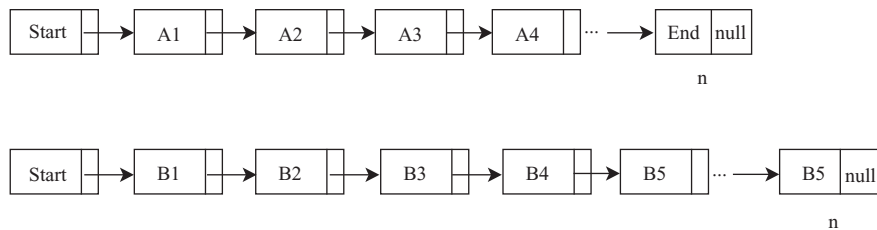


图4 街道地址编码链表

Fig.4 The address linked list based on stree

## 5 模型的实验结果

实验地点选取山东潍坊市城区,依托 1:500 比例尺数字地形图作为参考底图,采集了有名街道 200 余条,院落 4871 个,兴趣点(标志物)35 192 个,门牌 12 900 个。为了比较传统编码模型与街道编码模型的匹配效果,选取了注册地址中含有街道信息的私营企业 200 条经济户口数据,将这些地址数据进行地址分解和标准化,然后进行匹配实验。匹配时采取了注册地址与注册名称综合匹配的方式,试验结果见表 1。

表 1 两种编码规则的地址匹配结果

Tab.1 The address matching result of two geocoding rules

匹配类型	传统编码地址库		街道编码地址库	
	数量	精度	数量	精度
匹配到要素	71		82	
匹配到街道	53	2.1km	109	1.8km
街道链表中插值			35	15m
匹配失败	76		9	

实验结果表明:

(1) 匹配到要素:指依据名称和注册地址可与地址库中院落多边形、门址点或兴趣点要素匹配成功。由于考虑街道的地址编码中增加了街道信息,匹配结果增加了 11 条匹配到要素的记录。

(2) 匹配到街道:当依靠名称和地址不能准确匹配时,可依据地址中街道信息匹配到街道,点位取街道中点。量取企业实际位置与街道中点的距离,计算平均值得到匹配精度。从实验结果可见,街道编码后,可匹配到街道的记录增加约一倍。

(3) 由于一些企业注册地址采用了相对定位的地址描述方式,如“胜利西街百货大楼东 100m”等形式,因此,利用街道链表可精确计算出定位点。从表中可见插值计算后匹配精度大大提升。

本实验没有采用模糊匹配与候选打分机制,直接将既不能唯一匹配到要素也不能匹配到街道的记录视为匹配失败。

## 6 结论

本文提出的城市标准地址编码定义的修正方案,进一步明确了街道与行政区划、街道与小区和自然村对地址编码的区域限定关系。利用街道的

空间参考作用将门牌号码、点地址、面地址等有机地组织起来,形成了统一的街道地址编码模型,实现了街道地址编码与点地址编码和面地址编码的融合。在该模型中,地址匹配不再是单纯的地址字符串匹配,另外,还可以充分利用街道的方向性、地址单元的邻接关系等空间特征辅助进行地址匹配,从而大大提升了地址的匹配效率和匹配精度。

### 参考文献:

- [1] 薛明,肖学年.关于地理编码几个问题的思考[J].北京测绘,2007(2):54-56.
- [2] 王凌云,李琦,江洲.国内地理编码数据库系统开发与研究[J].计算机工程与应用,2004,(21):167-212.
- [3] 谢小惠.地理编码原理及方法研究[D].武汉:中南大学,2006.
- [4] 张鹤,陈倬,聂丹,等.北京市城市地理编码及可行性应用分析[J].测绘通报,2011(4):65-67.
- [5] 兀伟,高小萍.澳大利亚地理编码标准介绍[J].测绘标准化,2008(2):1-6.
- [6] Paul A. Zandbergen. A comparison of address point, parcel and street geocoding techniques[J]. Computers, Environment and Urban Systems, 2008(32):214-232.
- [7] 国家地理信息公共服务平台总体技术组.国家地理信息公共服务平台公共地理框架数据地名地址数据规范(试行稿-20100125版)[S].北京:国家基础地理信息中心,2010.
- [8] 国家测绘局. CH/Z 9002-2007,数字城市地理空间信息公共服务平台地名/地址分类、描述及编码规则[S].北京:测绘出版社,2008.
- [9] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.GB/T 23705—2009,数字城市地理信息公共服务平台地名/地址编码规则[S].南京:凤凰出版社,2009.
- [10] Clodoveu A, Davis J, Fonseca F T. Assessing the certainty of locations produced by an address geocoding system[J]. GeoInformatic, 2007,11(1):103-129.
- [11] 何冀飞.浅谈市场制度向街巷制度转型下的城市形态演变[J].城市建设理论研究,2011(18):20-22.
- [12] Paul Cooper. The open GIS abstract specification, topic 2: Spatial referencing by coordinates[R]. Open GIS Consortium Inc, 2010,12-14.
- [13] 高勇,张晶,张毅,等.一种基于地标的线性参照方法[J].北京大学学报(自然科学版),2010,1(46):109-114.
- [14] Lovasi G S, Weiss J C, Hoskins R. Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree[J]. International Journal of Health Geographics, 2007(3):6-12.



[15] Sadalla E K, Burroughs W J, Staplin L J. Reference-points in spatial cognition[J]. *Journal of Experimental*

*Psychology: Human Learning and Memory*, 1980,6(5): 516-528.

## Study on City Address Geocoding Model Based on Street

YU Huanju<sup>1</sup>, QI Qingwen<sup>2</sup> and LI Yunling<sup>1</sup>

(1. *Geomatics College, Shandong University of Science and Technology, Qingdao 266590, China;*

2. *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*)

**Abstract:** Address geocoding is a kind of address matching method based on spatial technique. It is the process of transforming a description of a location, such as coordinates, the address or the name of a place, to a location on the earth's surface. Address geocoding technologies started late in China, many cities just start capturing the geo coding address data. Many topics need to be explored, such as address code structure and geo coding data management. The authors analyzed the current address geocoding rules. The street is regarded as the same level distinct restriction as the courtyard in the current standards. At this point they think that the spatial characters of the street are weakened. On the basis of relevant national standards, they put forward the definition amendment of city standard address geocoding. In the amendment, as distinct restriction, street becomes independent and is in front of courtyard. Thus, an entity must belong to one street and the address code must contain the street's code. Usually it is difficult to do address matching with many types of address data because the address data store in different database tables by the entity types. Using streets as spatial reference system, according to the street onward direction, they well integrate the doorplates, address points and parcels together and put forward the new unified street geocoding model: street address chain model. In this model, geocoding is not simple address string matching any more, it also makes well use of the streets' direction, parcels' adjacency and other spatial characters to assist address geocoding process. This can provide model support for future address matching. The example about street address chain model of address geocoding is given. Predictably, this new address geocoding model will improve the match rate and accuracy greatly.

**Key words:** digital city; street; geocoding; address matching; standard address