

# 城市地理编码数据库的研究与实践

张悦丽, 张立涛

(青岛市勘察测绘研究院, 山东 青岛 266032)

**摘要** 地理编码数据库是地理编码技术的重要组成部分,它直接影响到地址匹配的效果。本文在青岛市地址体系的分析研究的基础上,详细阐述了青岛市地理编码数据库的建设流程,包括地址数据获取、地址数据标准化、地址数据入库和地址数据库组织的建设,并对地址数据的更新提出了建议。

**关键词** 地理编码 地址模型 地址标准化 地址元素

中图分类号: P208

文献标识码: A

文章编号: 1672-4097(2011)02-0022-03

## 1 引言

随着城市信息化进程的不断推进,地理信息系统技术在各行业的应用日益广泛。但由于大量信息资源的空间位置特征并非直接以地理坐标的方式记录而是通过文字地址表述方式存在,所以无法直接在地理信息系统中应用。利用地理编码技术能够将这些潜在的信息资源快速、准确地空间化,在统一的时空框架下与地理空间数据进行集成与融合。

国外地理编码技术发展比较成熟,如美国通用地理编码软件 TIGER 系统和 Mapinfo 的 Mapmarker、Arcgis 的 Geocoding 地理编码模块等,都已在各行业得到了广泛应用。我国地理编码技术起步较晚,在技术标准、服务模式、数据采集、管理、更新等方面尚未形成统一的解决方案。目前国内关于地理编码数据内容方面研究有一些,但仍存在一些问题:如我国的地名和地址体系相对比较混乱,各地的地名和地址使用不统一,因此各城市地理编码数据库建设应在现有相关研究基础上,探讨适合本地实际的地理编码数据库,以获得最佳的地址匹配效果。

## 2 地理编码定义

地理编码又称为地址匹配(Address Matching),是建立给定地址与真实地理坐标一致性的过程,简单来说就是给某一地址指定地图坐标,从而使得该地址所代表的地理实体在地图上得以空间化显示。它可分为正向匹配和逆向匹配两种方式,分别表示地址文字描述与空间地理坐标之间的双向转换过程。

## 3 青岛市地理编码数据库研究分析

地理编码技术的实现依赖于两个方面:完备、准确的地理编码数据库和准确、高效、可靠的地址匹配引擎。地理编码数据库建设主要集中在三点:一是明确地理编码对象,二是确定地址模型,三是确定编码规则。

### 3.1 地理编码对象

青岛市地理编码对象是地名和地址,包括政区名、区片、村、街巷、门址、楼址、标志物八大类。其中,政区名包含市、区、街道、社区名称的信息;区片名包含地片、区片、居住小区名称信息;村包含自然村名称信息;街巷名包含有地名标牌的街巷等;门址是指门牌地址;楼址是指楼牌地址;标志物包括具有地名意义的自然地名、经济地名、文化地名、单位、院落等。

### 3.2 地址模型

#### 3.2.1 地址元素内容

地址元素是构成地名、地址的最小语义单元,可以划分为 30 类地址元素类型,即:洞、公路、规划、河流、湖、潭、公园、环岛、纪念地、建筑物、交通站场、街巷、开发区、名胜古迹、桥梁、泉、山峰、山脉、水库、水渠、隧道、体育设施、铁路、政区、住宅区、村、地片、楼名、门址、广场、单位。

#### 3.2.2 标准地址的形式

标准地址是一个或多个地址元素组成的集合,各类型地址元素之间存在复杂的层次关系。标准地址由三个部分组成,即:

<标准地址> := <行政辖区> <基本区域限定物> <局部点位置描述>

其中<行政辖区>为政区类地名。<基本区域限定物>可以是道路、小区、公园、村、广场等层

次低于政区地名, 但可包含基本地址元素的地名。  
 < 局部点位置描述 > 是地址最后的地址元素, 包括门牌、楼号、单位、建筑物、自然地名、经济地名等点状地址元素类型。

### 3.2.3 标准地址层次组合规则

[ < 市 > | < 区 > | < 街道 > / < 地片 > | < 社区 > | < 村 > | < 街巷 > / < 区片 > | < 门(楼)牌 > / < 标志物 > / < 方位 > | < 兴趣点 > ]

根据青岛本地情况主要有以下几种组合:

表 1 标准地址层次组合规则示例

序号	组合规则
1	政区名+ 地片
2	政区名+ 片名
3	政区名+ 住宅区/ 标志物
4	政区名+ 村
5	政区名+ 村+ 门址
6	政区名+ 村+ 住宅区/ 标志物
7	政区名+ 村+ 住宅区/ 标志物+ 楼址
8	政区名+ 村+ 街巷+ 住宅区/ 标志物
9	政区名+ 村+ 方位+ 住宅区/ 标志物
10	政区名+ 街巷
11	政区名+ 街巷+ 门址
12	政区名+ 街巷+ 门址+ 楼址
13	政区名+ 街巷+ 门址+ 住宅区/ 标志物
14	政区名+ 街巷+ 住宅区/ 标志物
15	政区名+ 街巷+ 方位+ 住宅区/ 标志物

### 3.3 编码规则

由于青岛市社区的变动相对小, 所以以社区为基础进行地址数据的编码, 共 19 位, 依次是 6 位市辖区码、3 位街道办事处顺序码、3 位社区顺序码、1 位所属类型代码、6 位流水顺序号。结构如下:

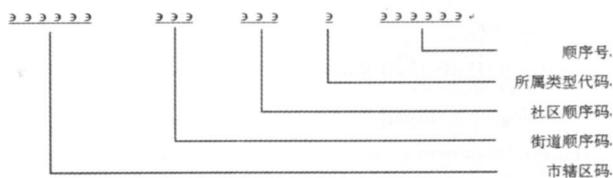


图 1 编码规则

市辖区码目前按照 GB/T 2260—2002《中华人民共和国行政区划代码》和 GB/T 10114《县以下行政区划编制规则》执行。街道办编码和

社区编码按从左到右、从上到下的顺序进行编码。

## 4 地理编码数据库建设

### 4.1 建设流程

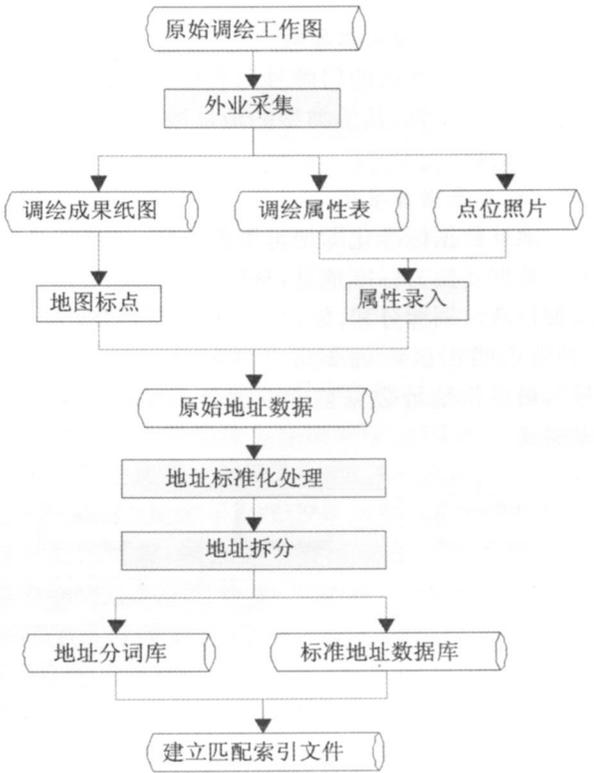


图 2 地理编码数据库建设流程

### 4.2 地址数据获取

#### 4.2.1 地址数据外业采集

利用 1: 500 地形图或影像图制作原始调绘工作图, 通过外业实地调查或走访问, 在工作图上标注地址点位, 填写外业地址数据属性调查表并拍摄照片, 通过流水号建立地址点位和属性信息的关联, 建立一一对应的关系。

地点描述应优先使用门(楼)牌地址, 优先级见表 2。

表 2 地点描述优先规则

优先级	地点类型	地点子类优先顺序
1	门(楼)牌地址	无
2	标志物地址	纪念地与建筑物、公共设施、交通运输设施等
3	兴趣点地址	沿街店铺名称
4	街巷名称	无
5	地片、区片	小区名称、地片名称
6	行政区划	市、区、街道、社区

### 4.2.2 地址数据内业录入

数据录入包括属性录入和标点录入,其工作通过地址录入系统中完成。录入内容包括位置编号、名称、关键字、地址、电话、照片、重要度、采集时间、备注等,最后通过流水号关联整合获得 shp 格式原始地址数据。

### 4.2.3 地址数据定位原则

具有官方承认的门牌号码和楼牌号,点位标注在实际标牌位置,其他类型的地址数据定位在几何中心或门口。

### 4.3 地址数据标准化

地址数据标准化即地址数据整理,是指将原始地址数据变换为标准地址,进行规范化地址描述并添加行政区划部分和,如:“辽阳东路 9 号”标准化为“青岛市崂山区中韩街道办西韩社区辽阳东路 9 号”,最后将地址数据整理成拆分入库前的中间数据格式。

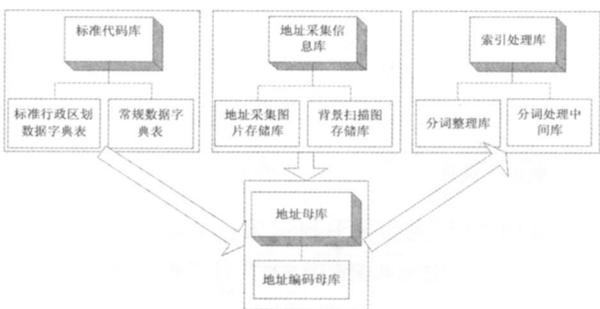


图 3 数据库逻辑图

### 4.4 地址数据入库

对中间数据格式的地址数据按照地址元素层次关系进行拆分,其中行政区域地名按[ < 市> | < 区> | < 街道> | < 社区> ]分段,保证类似“69 号 1”、“95 号 1”门址的完整性,确定每个地址元素的

分词级别、父地址、地址类型、同义词等,建立地址分词库和标准地址库。

最后建立匹配索引文件,利用中文分词技术进行地址匹配。

### 4.5 地址数据库逻辑结构设计

如图 3。

## 5 结语

青岛市地理编码数据库是在深入研究其他城市地址数据库建设情况和相关标准基础上,结合青岛市地址体系具体实际建设而成的,在青岛市信息资源的空间化上发挥了重要的技术支撑作用。它不但整合了政府各部门分散的政务信息资源,有效地支撑政府管理与决策,同时可以满足青岛市公众服务领域基于位置服务(LBS)的需求。目前的地理编码数据还有进一步改进的空间:一是需要与门牌号码和地名的政府管理部门合作,保证地址数据的权威性,二是保证数据的现势性,建立定期和实时更新机制,充分利用地址数据的编码提取一定范围地址数据实现局部更新。

### 参考文献

- 1 中华人民共和国建设部. CJ/T 213-2005 城市市政综合监管信息系统单元网格划分与编码规则[S]. 北京: 中国标准出版社, 2005
- 2 国家测绘局. CH/Z 9002-2007 数字城市地理空间信息公共平台地名/地址分类、描述及编码规则[S]. 北京: 测绘出版社, 2006
- 3 李军, 李琦, 毛东军等. 北京市地理编码数据库的研究[J]. 计算机工程与应用, 2004 (2): 4-6
- 4 李军, 彭凯等. 政务地理空间信息资源管理与共享服务应用[M]. 北京: 北京大学出版社, 2009: 256-257

## Research and Practice of Urban Geocoding Database

ZHANG Yue Li, ZHANG Li-tao

(Qingdao Geotechnical Investigation and Surveying Research Institute, Qingdao Shandong 266032, China)

**Abstract** Geocoding database is an important part of the geocoding technology. It directly affects address matching. Based on the analysis of the address system of Qingdao, this paper elaborates the establishment of Qingdao geocoding database, that including the acquisition of address data, the standardization of address data, importing address data into database, organization of the geocoding database. At the same time, some suggestions on the updating of address data are represented.

**Key words** geocoding; address model; address standardization; address factor