



回家! GIS 温暖回家路

中国地质大学（武汉）

数据获取及处理

报 名 号：_____高级开发组_____K20175020

作品名称：_____回家_____

指导老师：_____郭明强 方芳_____

团队成员：_____余列冰 王俊珏 杨美娟_____

学 校：_____中国地质大学（武汉）_____

二零一七年 九月



目录

1. 失踪人员数据获取及处理	1
1.1 宝贝回家网数据获取与处理	1
1.1.1 数据获取	1
1.1.2 数据处理	1
1.2 全网数据获取与处理	2
1.2.1 数据获取	2
1.2.2 数据处理	3
2. 路网数据获取及处理	4
2.1.1 数据获取	4
2.1.2 数据处理	5



1. 失踪人员数据获取及处理

失踪人员数据主要来源于“宝贝回家网”以及 Bing 全网搜索，“回家”系统数据分析与可视化展示主要基于“宝贝回家网”数据，通过 Bing 进行全网搜索扩大了数据来源范围。

1.1 宝贝回家网数据获取与处理

1.1.1 数据获取

本模块通过后台定时调用爬虫脚本进行数据实时爬取与及时更新。数据来源于宝贝回家网（<http://www.baobeihuijia.com/Index.aspx>）（爬虫框架结构详见《回家——项目开发文档》2.1.3 架构体系中数据爬取层）。

1.1.2 数据处理

①对数据进行结构化处理，得到具有 13 个字段：寻亲编号、寻亲类别、姓名、性别、出生日期、失踪时身高、失踪时间、失踪地点、寻亲特征、其他资料、注册时间、注册者、状态（初始化数据共 34470 条）；

②通过对失踪地点调用 ArcGIS API 进行逆地理编码得到经纬度，数据结构相应增加经度、纬度两个字段；

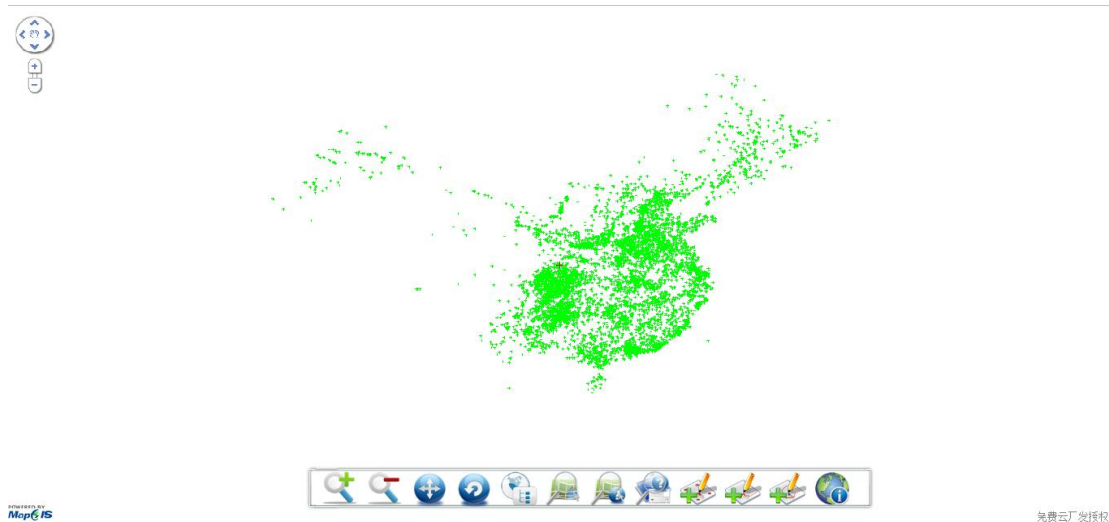
③在 MapGIS 桌面端建立一个空间数据库 gohome，数据转换工具，将数据转化为空间数据，预览如（图 1）；

④将该空间数据（MissingPeople）制作成地图文档并于 MapGIS IGServer 上发布，预览如（图 2）；

⑤在初始爬取发布后，之后实时爬取数据通过轮询存储插入至 IGServer 空间数据库中。



(图 1) 失踪人员数据矢量图层预览



(图 2) 失踪人员数据地图文档预览

1.2 全网数据获取与处理

1.2.1 数据获取

为了扩大数据来源范围,本系统通过调用 Bing 搜索引擎对各大寻亲网站对失踪人员进行搜索,并对返还数据进行相关度排序,列出前 20 条数据。



1.2.2 数据处理

① Bing 搜索引擎返还的数据为 json 数据（如图 3），通过 javascript 脚本对 json 数据进行解析与格式化处理；



（图 3）全网搜索 json 数据

②通过 html 与 javascript 动态将 20 条数据显示出来（如图 4），用户通过点击标题可以获得详细信息。



（图 4）全网搜索结果展示



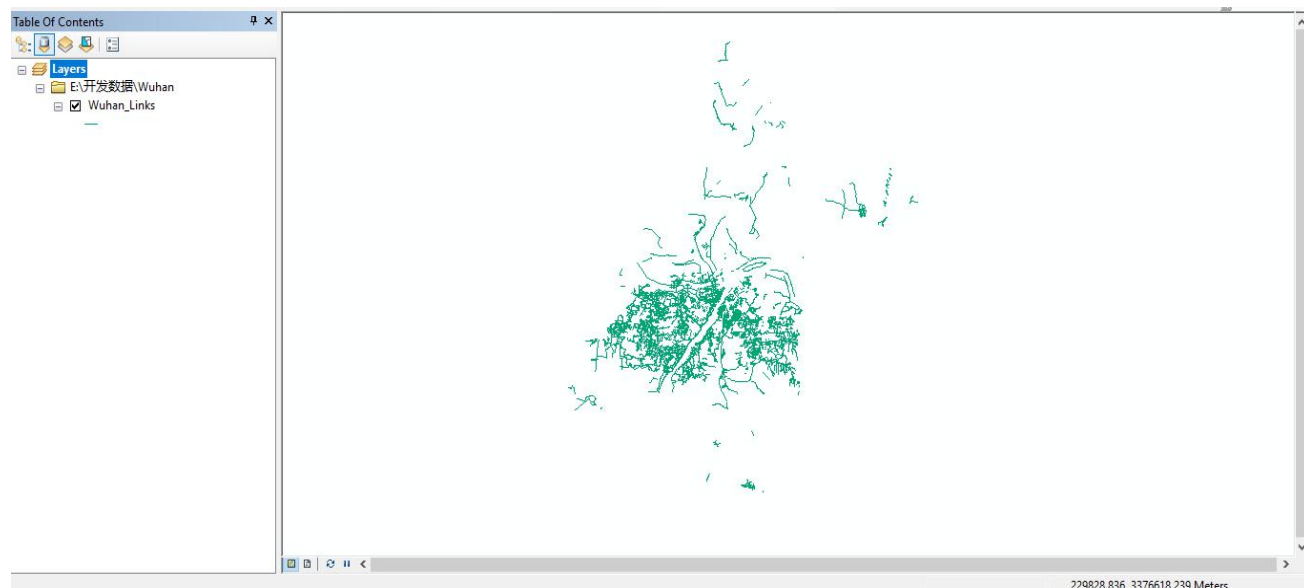
2. 路网数据获取及处理

2.1.1 数据获取

武汉市路网数据来源于 OSM（OpenStreetMap），一个免费自由开放的地理数据平台。路网数据为矢量格式，并且根据道路的等级与大小划分为不同类别。按照道路等级，降序排列为主要高速公路, 主要公路, 二级公路, 小路（即当地，邻里和乡村街道）。数据含有道路类型、道路长度等字段，如（图 5），于 ArcGIS 进行预览，如（图 6）

FID	Shape	OBJECTID	FID_Select	osm_id	name	ref	type	oneway	bridge	tunnel	maxspeed	Class	Shape_Leng
0	Polyline	2	6969	162957568			road	0	0	0	0	0	261.621702
1	Polyline	4	6969	162957568			road	0	0	0	0	0	204.100657
2	Polyline	6	6971	162957574			residential	0	0	0	0	0	10.506885
3	Polyline	7	7001	162957632			residential	0	0	0	0	0	23.475778
4	Polyline	8	6996	162957619			residential	0	0	0	0	0	7.373725
5	Polyline	9	7004	162957635			track	0	0	0	0	0	21.053451
6	Polyline	10	6994	162957614			residential	0	0	0	0	0	16.590402
7	Polyline	11	6976	162957580			residential	0	0	0	0	0	80.95026
8	Polyline	12	7005	162957636			track	0	0	0	0	0	23.992839
9	Polyline	13	6976	162957580			residential	0	0	0	0	0	30.280912
10	Polyline	14	6977	162957582			residential	0	0	0	0	0	22.215153
11	Polyline	15	6996	162957619			residential	0	0	0	0	0	17.57828
12	Polyline	16	6976	162957580			residential	0	0	0	0	0	90.127314
13	Polyline	17	6965	162957562			unclassified	0	0	0	0	0	49.866586
14	Polyline	18	6994	162957614			residential	0	0	0	0	0	13.878457
15	Polyline	19	7002	162957633			unclassified	0	0	0	0	0	20.288635
16	Polyline	20	6977	162957582			residential	0	0	0	0	0	13.620087
17	Polyline	21	6976	162957580			residential	0	0	0	0	0	69.052307
18	Polyline	22	6985	162957592			residential	0	0	0	0	0	24.843001
19	Polyline	23	6976	162957580			residential	0	0	0	0	0	12.065902
20	Polyline	24	6972	162957575			residential	0	0	0	0	0	44.367054
21	Polyline	25	6994	162957614			residential	0	0	0	0	0	14.296144
22	Polyline	26	6972	162957575			residential	0	0	0	0	0	13.269346
23	Polyline	27	6976	162957580			residential	0	0	0	0	0	17.434435
24	Polyline	28	7002	162957633			unclassified	0	0	0	0	0	22.864568
25	Polyline	29	7000	162957631			residential	0	0	0	0	0	43.758998
26	Polyline	30	6976	162957580			residential	0	0	0	0	0	24.490122

（图 5）原始武汉市路网数据结构



(图 6) 原始武汉市路网数据预览

2.1.2 数据处理

①对获得的路网数据道路类型进行英汉转换，英汉转换表如下：

道路名称	中文实际意义
gondola	高空缆车
bridleway	骑马道路
cycleway	主要为自行车或自行车专用道路
footway	主要为行人车或行人专用道路。这包括步行小径和砾石路
living_street	行人有优先权，儿童可以在街上玩，最高速度较低的街道
trunk	主干道路/城市快速路
unclassified	未分类道路（最普通的道路，通常很窄）
motorway	高速公路
motorway_link	通往高速公路/快速公路的道路/出入口
path	一个非特定的或共享使用的路径
pedestrian	行人专用街道（步行街）或行人专用区
primary	主要道路。连接主要区域 / 城镇的道路
primary_link	连接主要道路与其他主要道路或较低级道路的连接路



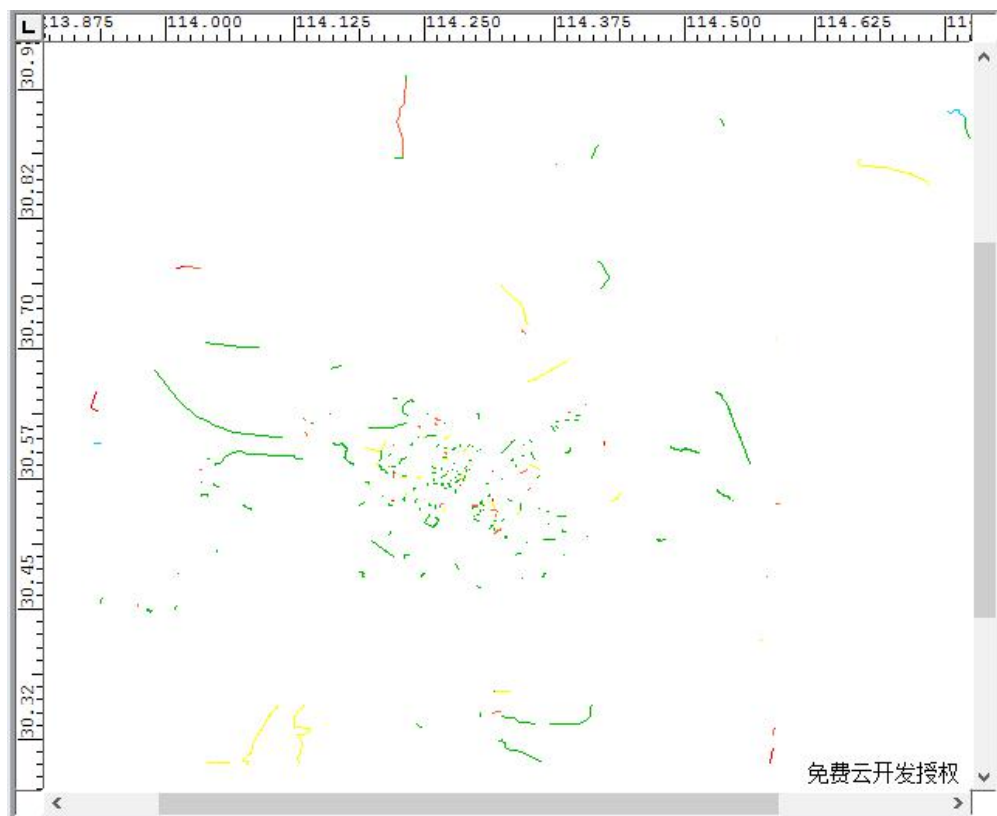
residential	进入或靠近住宅区的道路
secondary	次要道路。连接主要区域的道路，但不及主要道路重要
secondary_link	连接次要道路与其他次要道路或较低级道路的连接路
service	高速公路 / 快速公路上的服务区
steps	在人行道的楼梯(阶梯)/去庙里上的阶梯
tertiary	普通街道
tertiary_link	连接普通道路与其他普通道路或较低级道路的连接路
road	未知其分类的道路
taxiway	滑行道，飞机在跑道和停机坪之间的制动区域
bridge	桥梁
runway	跑道
roundabout	环形道
track	小径（通常未铺设路面）

（表 1）道路类型英汉转换表

②通过核密度模型计算每条路段失踪风险指数（模型详见《回家——项目开发文档》3.1 路段失踪风险评估模型），并通过调用 Arcpy 将失踪风险指数添加至每条路段属性中。

③在 MapGIS 桌面端，通过数据转换首先将 shp 文件导入空间数据库 gohome 中，在利用投影转换工具将坐标系转化为 WGS84 数据使其与失踪人员数据与地图吻合；

④通过根据 MapGIS 中属性选择器筛选剔除高速公路等道路等级较高的路段，并根据属性修改参数使得武汉市易失路段根据失踪风险指数分级显示，如（图 7）；



(图 7) 易失路段分级显示图