

城市大数据的获取

Methods for retrieving urban big data



龙瀛，博士
清华大学建筑学院
2016年9月30日

健康步道起点



Approaching the Human City: Beijing Studio
September 11 - 23, 2016
COURSE SYLLABUS

*A joint workshop between the Human Cities Initiative at Stanford University
& Tsinghua University Academy of Art and Design and the School of Engineering*

We invite you to participate in this experiment while we are holding class. Allow yourself to be fully present in the room, so you can listen to your classmates and what they have to say and share with you. Experience the freedom of not having to have your attention diverted or your mood instantly altered by whatever email or text message should come your way. You may take a phone call if it seems particularly urgent. But for most circumstances, we urge that you give yourself permission to be in control of your own time and energy— to actively choose where you want to direct your attention, as opposed to a portable device making that decision for you.

Please turn your mobile devices to “silent” or “do not disturb” mode, and do not take them out for the duration of the class. We strongly encourage you to take notes using pencil and paper— as research shows that this helps with memory retention— but If you must have a laptop to take notes, **please do not check your e-mail or browse the internet at any time.**

There will be plenty of opportunities to plug in once you leave the classroom. Let’s treat our classroom as a sacred space to enjoy the moment.

《大数据与城市规划》教学大纲

1. 大数据与城市规划概论（整合进9月23日）
2. 大数据在城市规划中应用的研究进展（9月23日）
3. **城市大数据的获取（9月30日）**
4. 城市大数据的统计与分析（10月9日：包括部分GIS/SPSS操作介绍）
5. 城市大数据的可视化（10月14日）
6. 城市大数据挖掘：空间句法（10月21日）
7. 城市大数据挖掘：城市网络分析（10月28日）
8. 学生作业中期汇报与点评（11月4日）

9. 大数据与城市规划的结合（11月11日）
10. 数据增强设计（11月18日）
11. 战略及总体规划中的大数据应用（11月25日）
12. 控制性详细规划中的大数据应用（12月2日）
13. 城市设计中的大数据应用（12月9日）
14. 参与式规划中的大数据应用（12月16日）
15. 大模型：新数据环境下的城市研究新方法（12月23日）
16. 学生作业终期汇报与点评（12月30日）

本讲大纲

1. 城市大数据获取的总体思路
2. 北京旧城数据介绍（含典型城市大数据介绍）
3. 结构化网页数据采集
4. 基于API的数据采集
5. 大作业安排（含场地调研方法）

一、城市大数据获取的总体思路

传统数据的获取



城市大数据获取的总体思路

- 大数据不开放、开放数据不大
 - 关系
- 1:0→100:0
- 购买数据、购买服务、免费网络资源（如BCL）、合作、自学、上课、参加培训

二、北京旧城数据介绍

含典型城市大数据介绍

所共享的图层简要说明

• 边界与分区

- 街道办事处: Jiedaos
- 居委会: Juweihui
- 利用2014年路网生成的地块: Parcels2011AICP
- 公交分析小区: TAZs

• 开发

- 遥感影像: 2005、2010、2015image

• 形态

- 分等级路网: Road2009、Road2014、Road_all_attributes
- 建筑物: Buildings2010
- 土地出让: LandTransaction2013

• 功能

- 兴趣点 (points of interest、POI) : POI2009、POI2011、POI20014

• 活动

- 街道办事处人口: TownshipPOP
- 位置微博: Weibo2014_07_10
- 微博签到: Weibo_POI_Checkins1127jp
- 大众点评: Dianping2014jped

• 活力

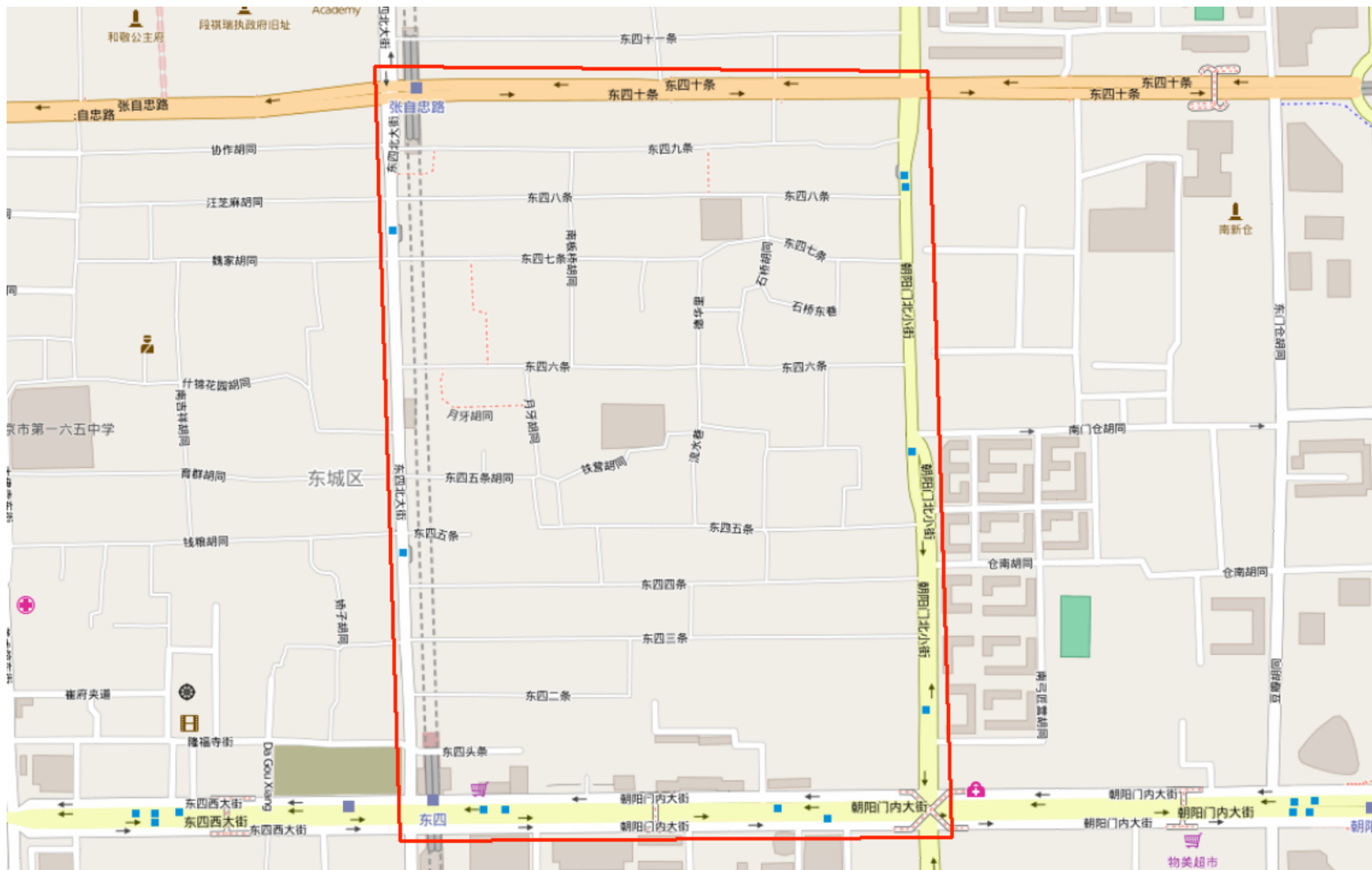
- 位置微博: 具体内容体现了情绪等 (尚未提供)
- 大众点评: 各POI的评价内容体现了品质、情绪等 (尚未提供)

• 其他

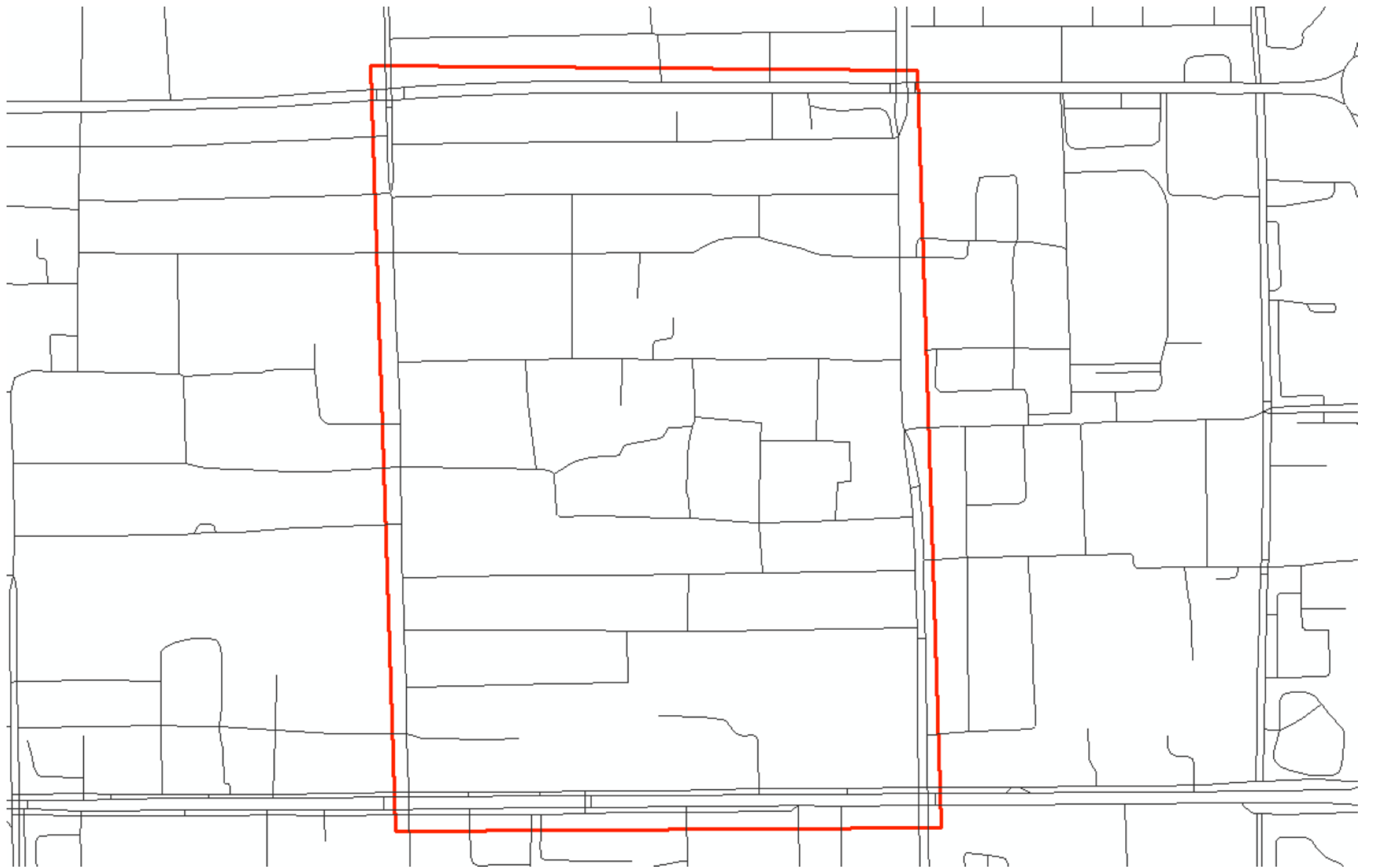
- 公交站点: BusStations201x
- 地铁站点: SubwayStations_new
- 街道步行指数: WalkScore2014
- 居住小区: ResidentialCommunity

开发 → 形态 → 功能 → 活动 *t*

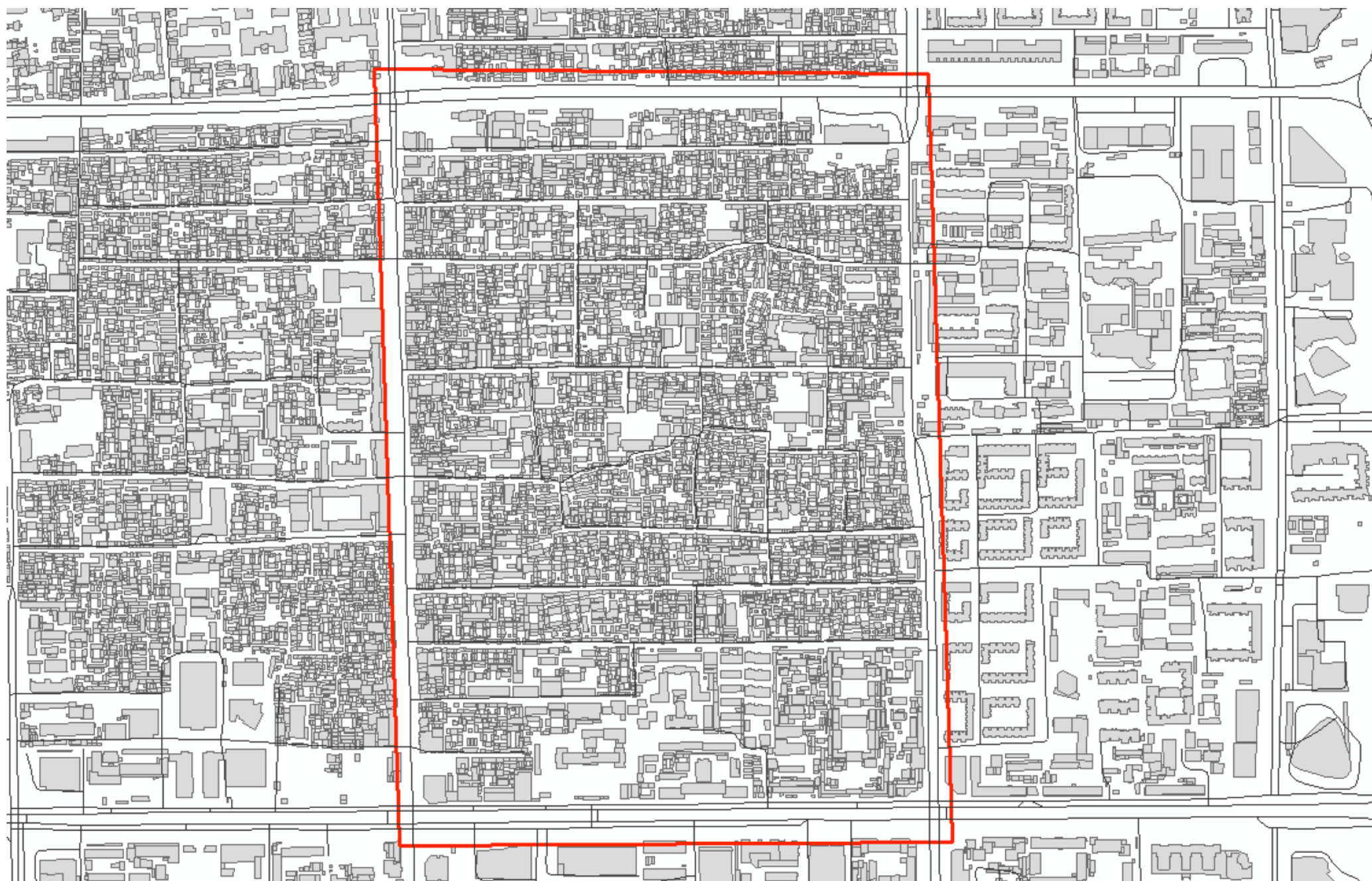
研究范围（仅供示意）



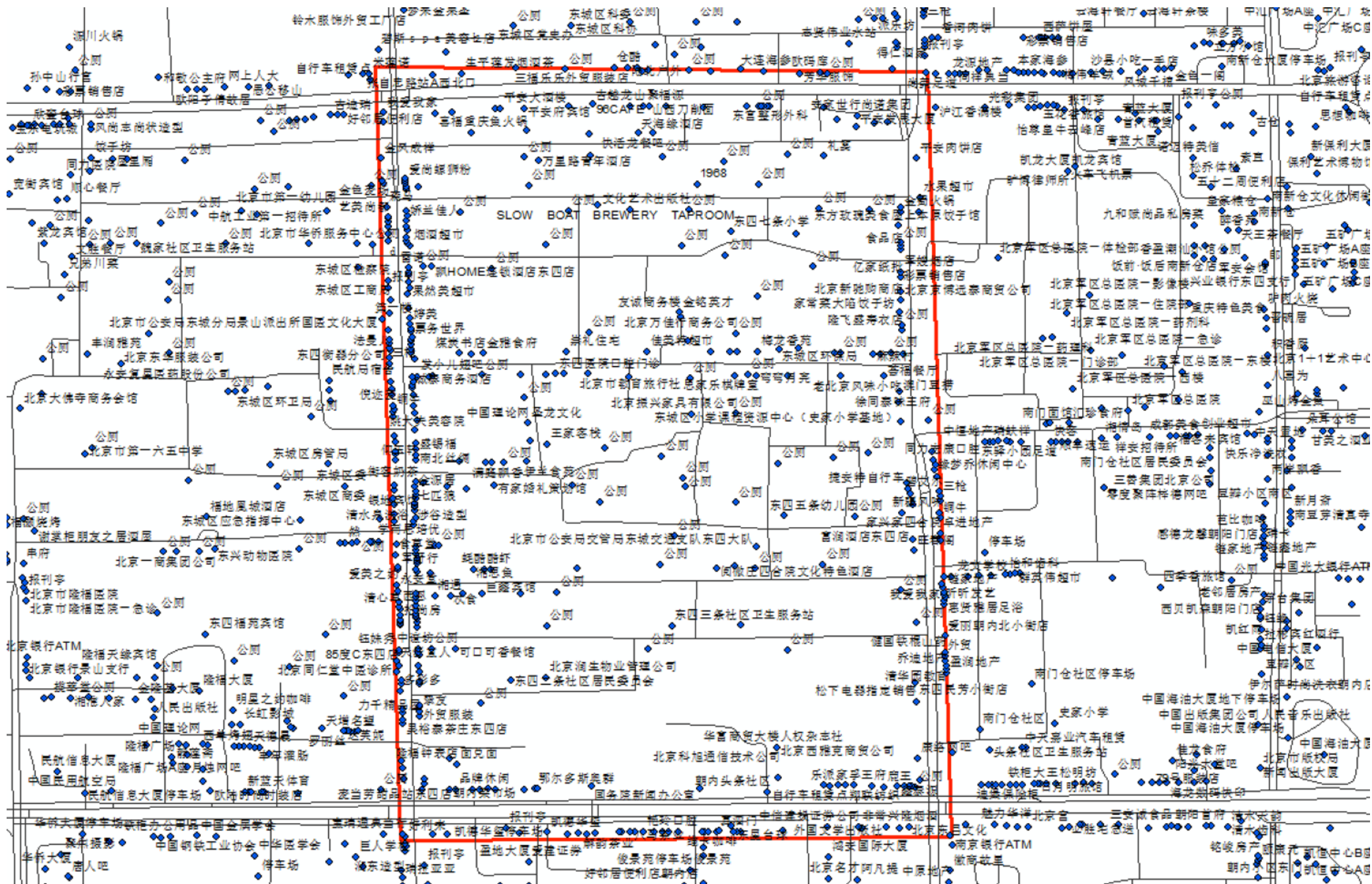
城市形态与设计 (道路网络)



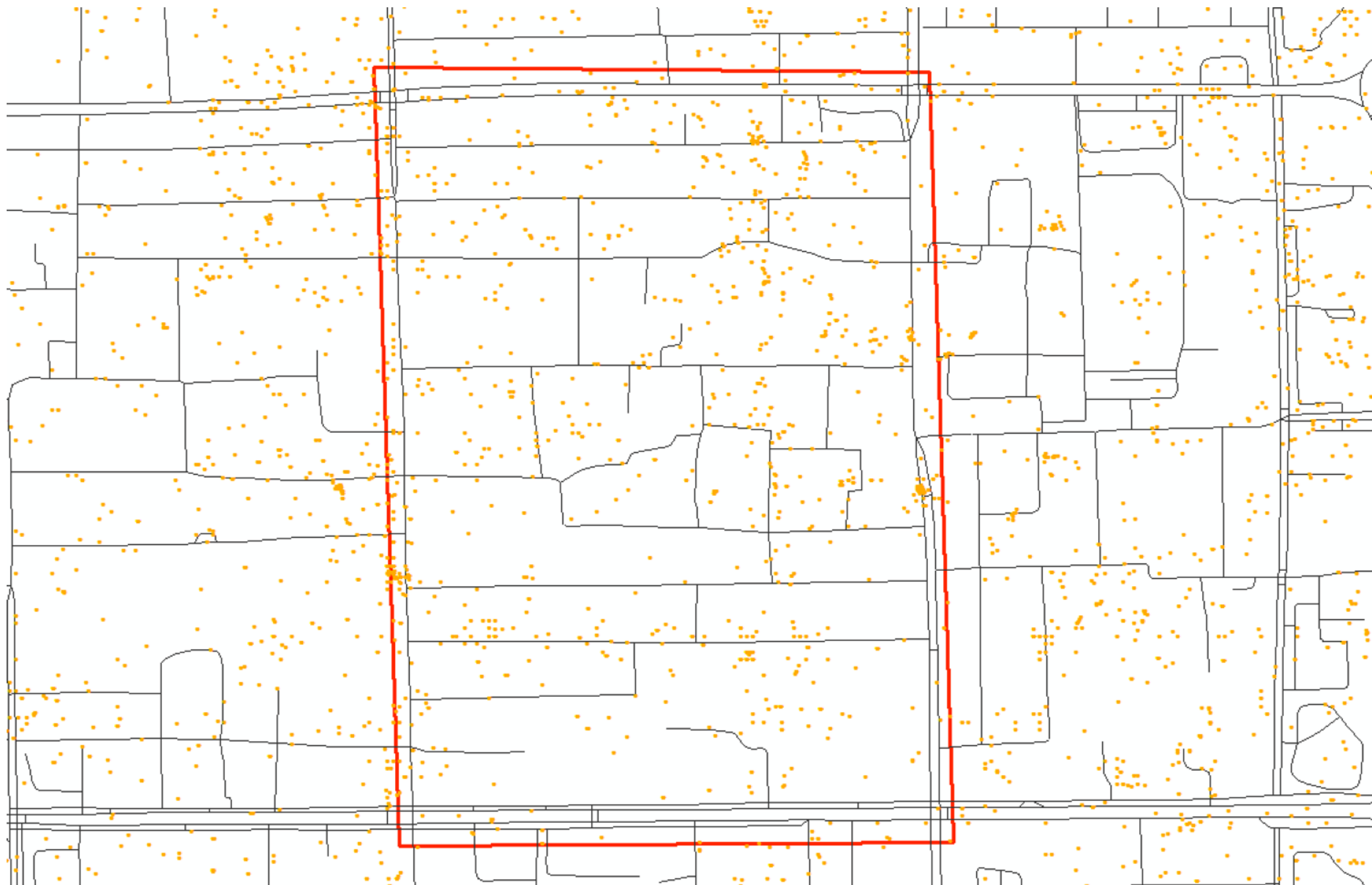
城市形态与设计（建筑物）



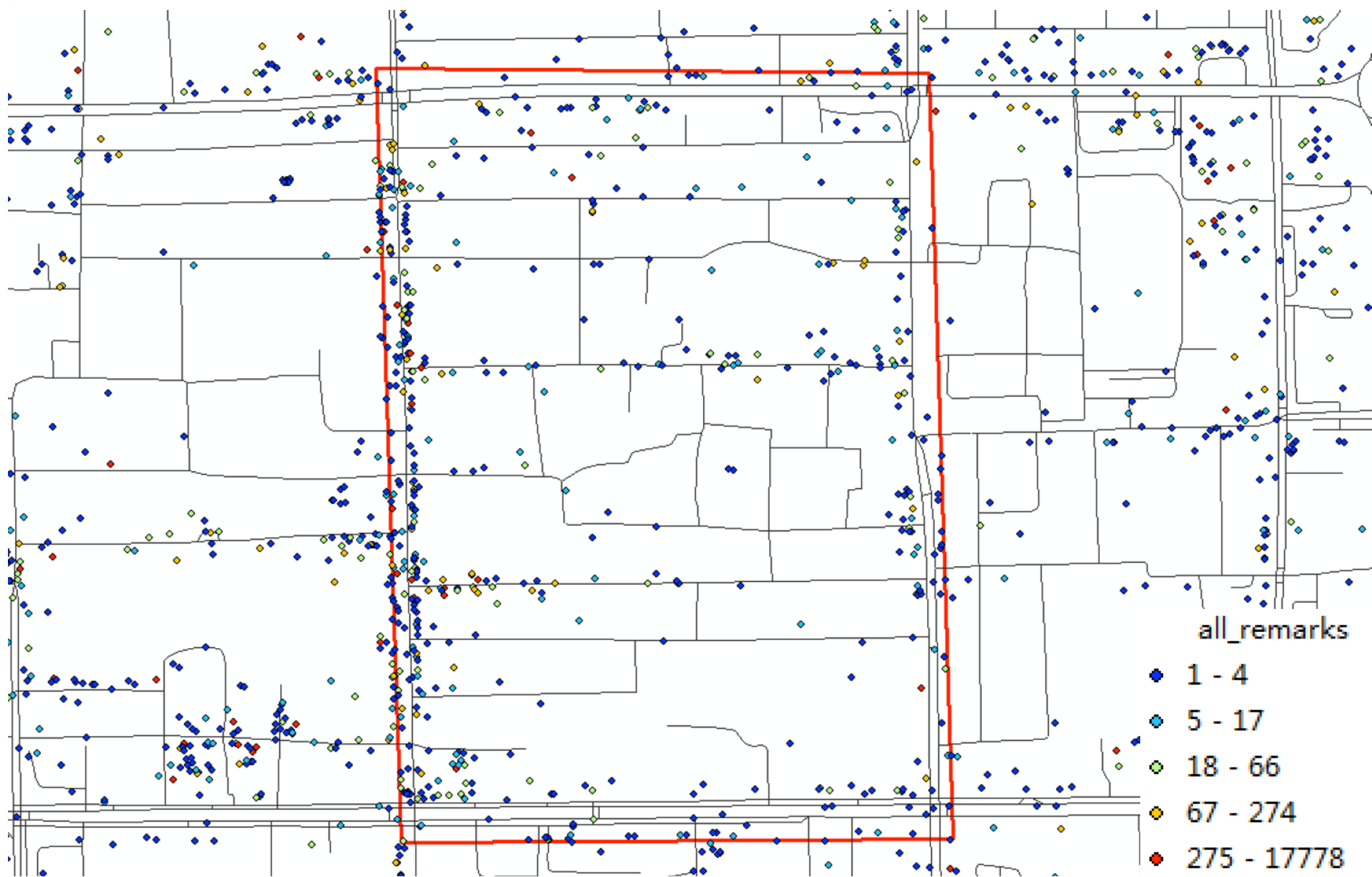
城市功能 (兴趣点)



城市活动 (位置微博)



城市活动 (大众点评)



新数据与不同尺度的城市研究的关系

表1 基于新数据支持总体城市设计的框架体系

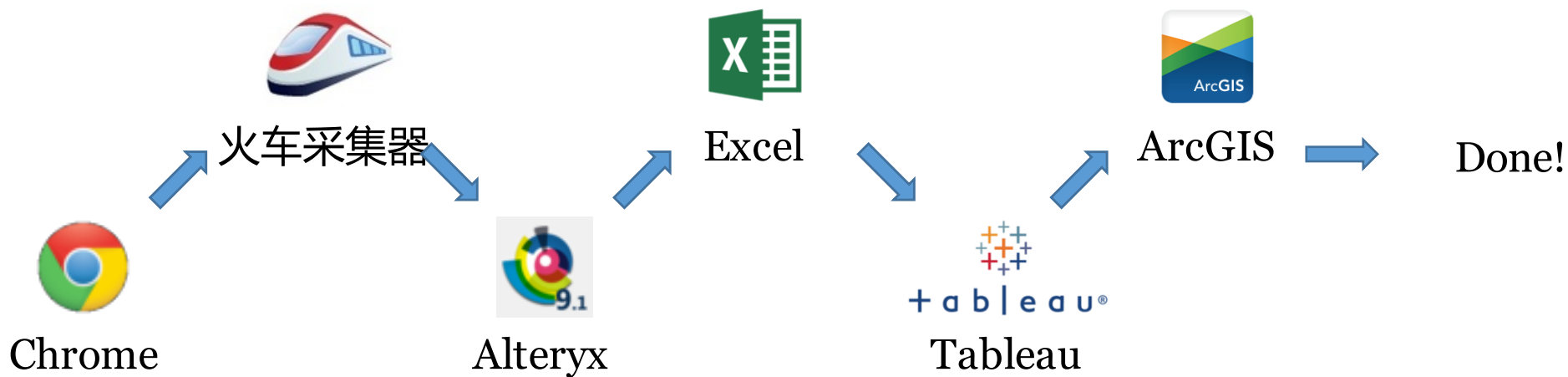
尺度/维度	区域/城市/片区/ 乡镇街道办事处	街区/地块	街区/地块内部	街道	街道内部
开发：遥感解译的土地利用、用地现状图（规划）、土地利用图（国土）	城镇用地面积、建设强度、生态安全格局、适宜开发土地 [城市扩张速度、城市扩张规模]	开发年代、是否适宜开发	肌理变化	角度变化	
形态：分等级路网、道路交叉口、建筑物、土地出让/规划许可、街景	基于道路交叉口的城乡判断、建筑面积、路网密度、交叉口密度、开放空间比例 [再开发比例、扩张比例]	尺度、紧凑度、基于建筑的城市形态类型、建筑密度、容积率、是否为开放空间、开放空间类型、可达性 [再开发与否、扩张与否]	是否有小路、建筑分布规律、是否有内部围墙 [历史道路构成]	长度、区位、直线率、建筑贴线率、界面密度、橱窗比、宽高比、可达性、铺装、建筑色彩 [历史上是否存在]	建筑分布特征
功能：兴趣点、用地现状图（规划）、土地利用图（国土）、街景	各种功能总量及比例、（城镇建设用地内）各种公共服务覆盖率/服务水平、职住平衡水平、产业结构/优势/潜力	用地性质、（各种）功能密度、功能多样性、主导功能、第二功能、各种公共服务设施可达性、市井生活相关的功能密度	（各种）功能分布特征（单面、双面、三面还是四面）、内部功能相比总功能（内部+临街）占比、界面连续度	（各种）功能密度、功能多样性、主导功能、第二功能、各种公共服务设施可达性、市井生活相关的功能密度、步行指数（walk score）、绿化、等级	（各种）功能分布特征（交叉口附近还是中间）
活动：普查人口、企业、手机、微博、点评、签到、公交卡、位置照片、百度热力图、高分辨率航拍图	总体分布特征、（城镇建设用地内）各等级活动所占面积比例、人口/就业密度体现的多中心性、联系所反映的多中心性、平均通勤时间/距离、各种出行方式比例	（不同时间段的）活动密度、微博密度、点评密度、签到密度、与之产生联系的地块、人口密度、就业密度、热点时段、通勤时间/距离	活动分布特征（内部还是边缘）、内部联系特征	（不同时间段的）活动密度、与之产生联系的街道、点评密度、热点时段、（各类型）交通流量、选择度与整合度、限速	活动分布特征（交叉口附近还是中间）
活力：街景、点评、手机、位置照片、微博和房价等	平均心情、整体意象、整体活力、幸福感	平均心情、平均消费价格、好评率、意象、市井活力、平均房价、居住隔离程度		平均消费价格、好评率、设计品质、风貌特色、活力、意象、平均房价	

注：表中 [] 特别给出了简单指标变化之外的指标；此表也适用于城市规划与设计方案的评价

三、结构化网页数据采集

感谢清华同衡李栋提供的本部分的资料

基本步骤



• 工具：获取/处理 → 分析 → 表达

数据的获取与处理

- 探查、定位数据源的网络地址
- 访问该网址，并将结果保存在本地
 - 自动化：持续、反复、定时.....
- 处理、清洗数据
 - 根据分析任务而定
 - 根据数据特征而定



Chrome



火车采集器



Alteryx



Excel



火车采集器

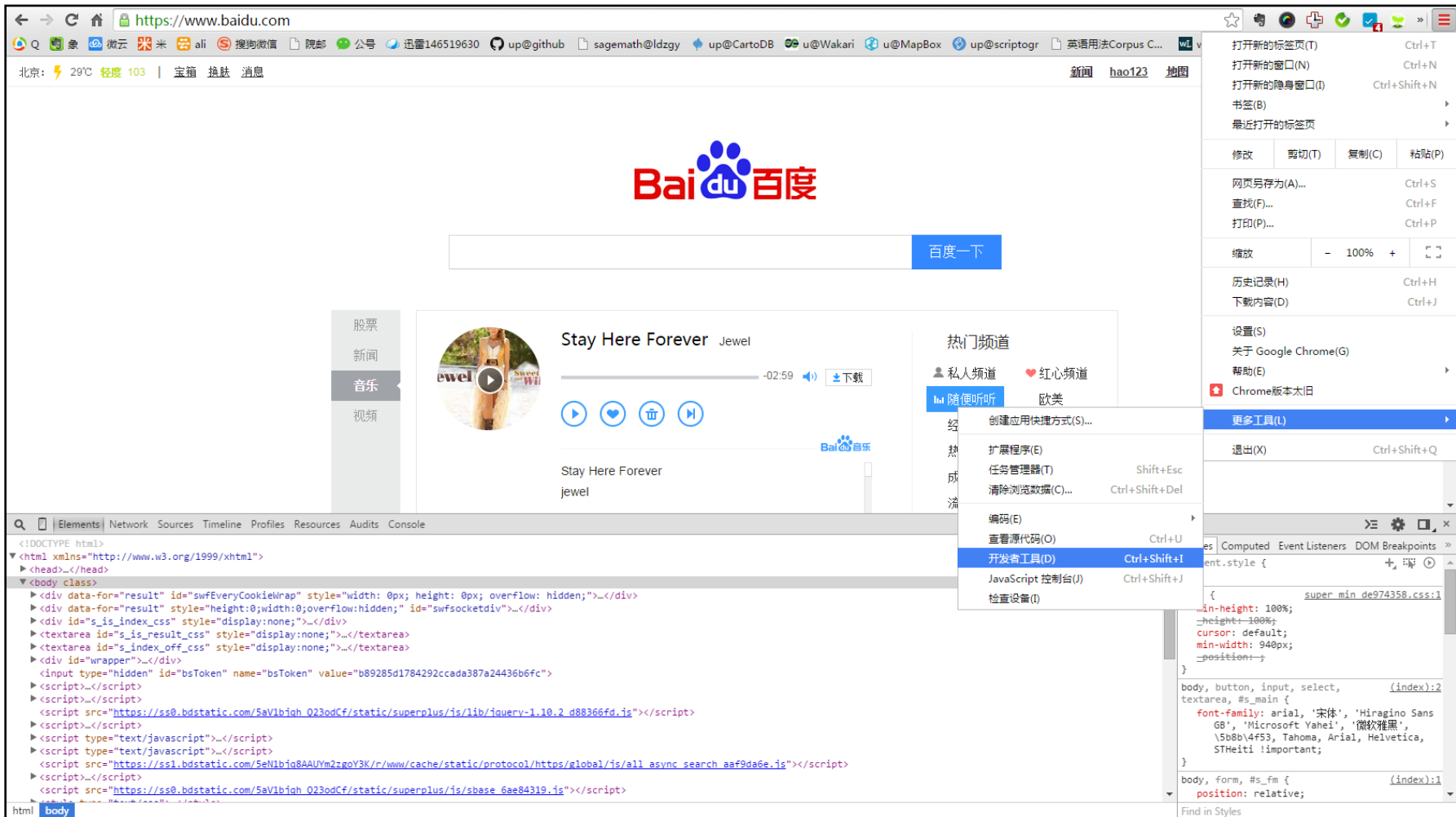


SAFE FME

数据的获取与处理

• Chrome开发者工具 (Chrome DevTools)

- 查看网页要素
- 抓包分析
- Javascript调试
-



数据的获取与处理

• 火车采集器

LocoySpider

火车采集器 8.6

文件(F) 分组(S) 任务(T) 视图(V) 工具(T) 高级(A) 扩展(E) 帮助(H)

新建 开始 暂停 停止 发布 入库 调度 代理 识别 管理

任务列表树

任务队列管理

任务名称 [ID]	网址成...	内容成败重
北京交通指数 [370]	1/0/0	1/0
定位api调用分布统计since201412...	1/0/0	1/0
全国城市air_stations_since2015...	1/0/0	1/0
全国城市air_cities_since201505...	1/0/0	1/0
北京景区舒适度since20141122 [280]	1/0/0	8920/223
迁徙统计json 迁入 since2014122...	1/0/0	365/0
迁徙统计json 迁出 since2014122...	1/0/0	365/0
消息sq07_more组二 [705]	0/0/0	0/27
消息sq07_more组二 [704]	0/0/0	0/65
消息sq07_more组二 [703]	0/0/0	0/0
数量复查 [706]	6/0/0	42/5756
数量复查2 [707]	6/0/0	251/5549
数量复查3 [708]	6/0/0	3/5797
数量复查4 [709]	6/0/0	2/5796
数量复查5 [710]	6/0/0	0/5800
数量复查6 [711]	6/0/0	0/5800
数量复查7 [712]	3/0/0	0/2693

任务运行完成

【totalnum】不符合不得为空设置, 被标记为未采, 记录Id2690, 记录网址
[@count=1&page=1&range=5000&starttime=1433520000&endtime=1434729600 @mapid=CSJ_R5k_8026@lat=32.6217780584@long=117.376895206 @access_token=2.00z_2idFt1d1sC14925229adBKe16E@turn2=80](https://api.weibo.com/2/place/nearby_timeline.json?sort=0&offset=0)
当前记录符合过滤要求, 已经被标记为未采: id2690, 网址:
[@count=1&page=1&range=5000&starttime=1433520000&endtime=1434729600 @mapid=CSJ_R5k_8026@lat=32.6217780584@long=117.376895206 @access_token=2.00z_2idFt1d1sC14925229adBKe16E@turn2=80](https://api.weibo.com/2/place/nearby_timeline.json?sort=0&offset=0)

【totalnum】不符合不得为空设置, 被标记为未采, 记录Id2691, 记录网址
[@count=1&page=1&range=5000&starttime=1433520000&endtime=1434729600 @mapid=CSJ_R5k_8027@lat=32.6197516489@long=117.451348681 @access_token=2.00z_2idFt1d1sC14925229adBKe16E@turn2=81](https://api.weibo.com/2/place/nearby_timeline.json?sort=0&offset=0)
当前记录符合过滤要求, 已经被标记为未采: id2691, 网址:
https://api.weibo.com/2/place/nearby_timeline.json?sort=0&offset=0

采内容 (2693/2693) 滚屏显示

文件	大小	完成	进度	剩余	速度	时间	状态	续传	地址	序号
----	----	----	----	----	----	----	----	----	----	----

分块下载 日志 文件下载平均速度: 0 KB/s

尝试	进度	已完成	分块大小	起点	终点	速度	剩余	状态	地址
----	----	-----	------	----	----	----	----	----	----

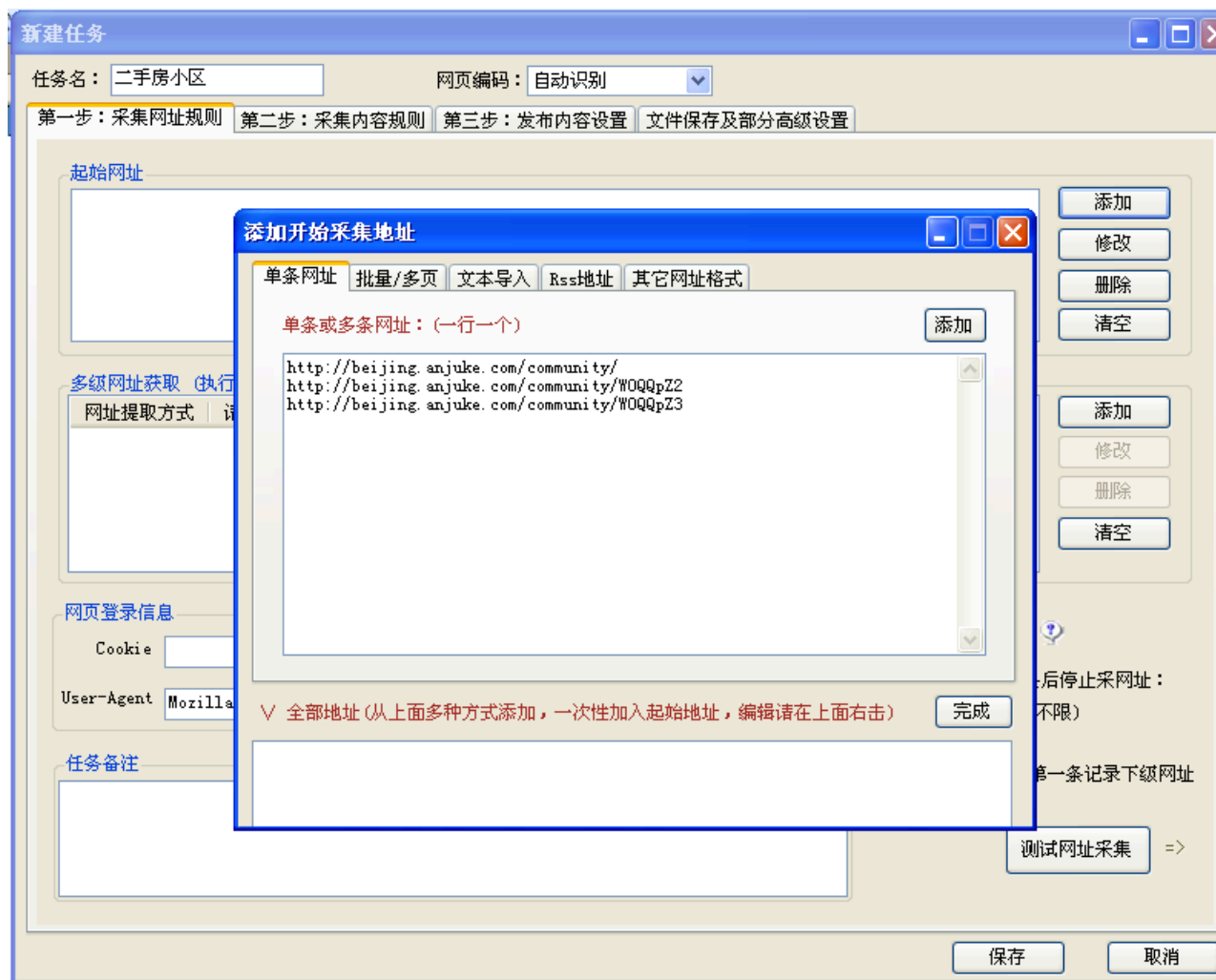
任务:消息a11 ID:586 扩展:火车采集器

运行中:0, 就绪:0, 暂停:0, 停止:1707

欢迎您: ldzgy01

数据的获取与处理

• 火车采集器 1. 添加网址



数据的获取与处理

• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标



数据的获取与处理

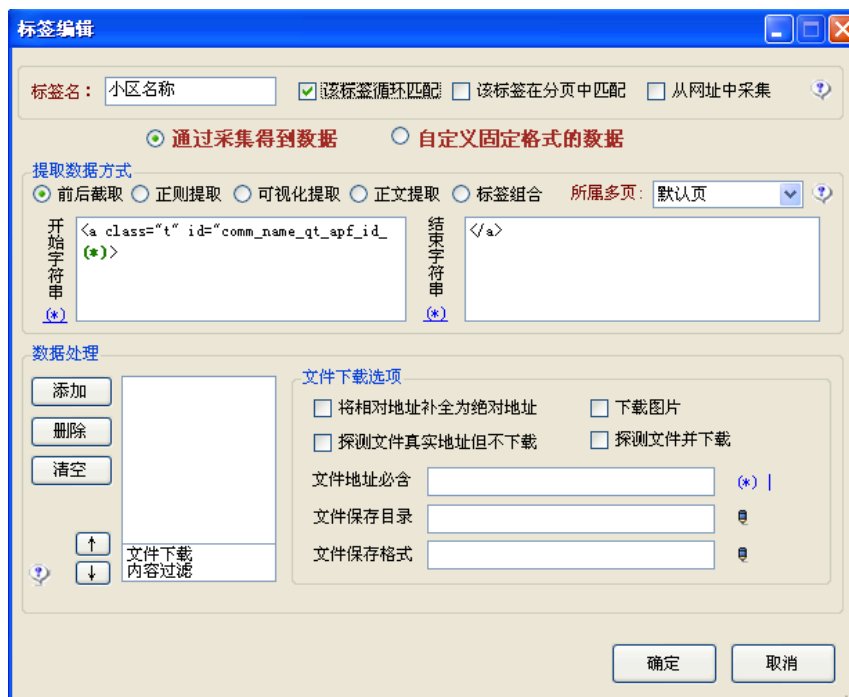
• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

```
...  
<a class="t" id="comm_name_qt_apf_id_11"  
href="http://beijing.anjuke.com/community/view/55  
031" target="_blank">珠江帝景</a>  
...
```



数据的获取与处理

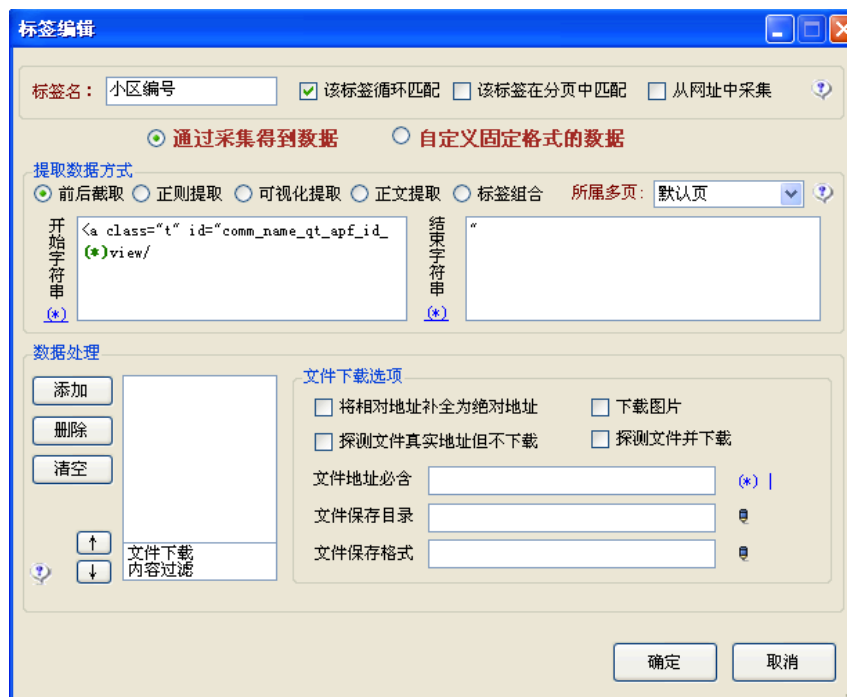
• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

```
...  
<a class="t" id="comm_name_qt_apf_id_11"  
href="http://beijing.anjuke.com/community/view/55031"  
target="_blank">珠江帝景</a>  
...
```



数据的获取与处理

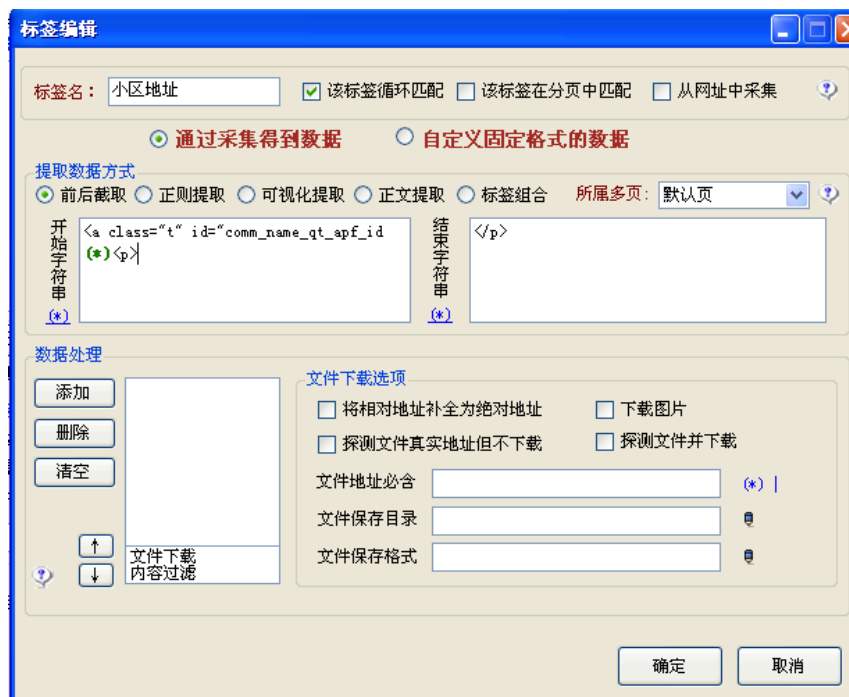
• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

```
...  
<a class="t" id="comm_name_qt_apf_id_11"  
href="http://beijing.anjuke.com/community/view/55031  
" target="_blank">珠江帝景</a></div>  
<p>[朝阳 大望路]西大望路甲23号(广渠路28号)</p>  
...
```



数据的获取与处理

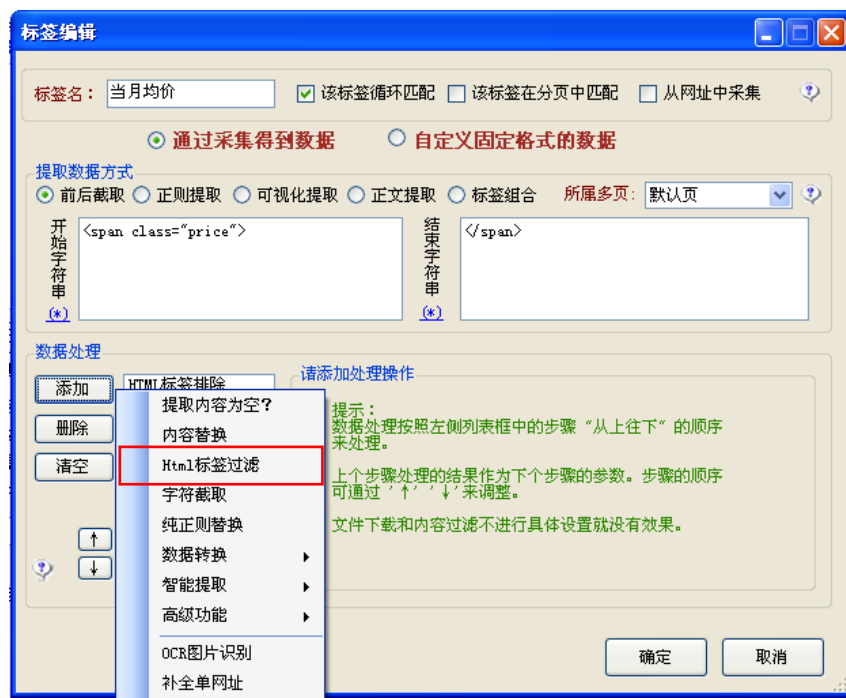
• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

```
...  
<span class="price">  
  <span class="sp1">52759</span>  
  <span class="sp2">元/平米</span><br>  
  <i>↑1.54%</i>  
</span>  
...
```



数据的获取与处理

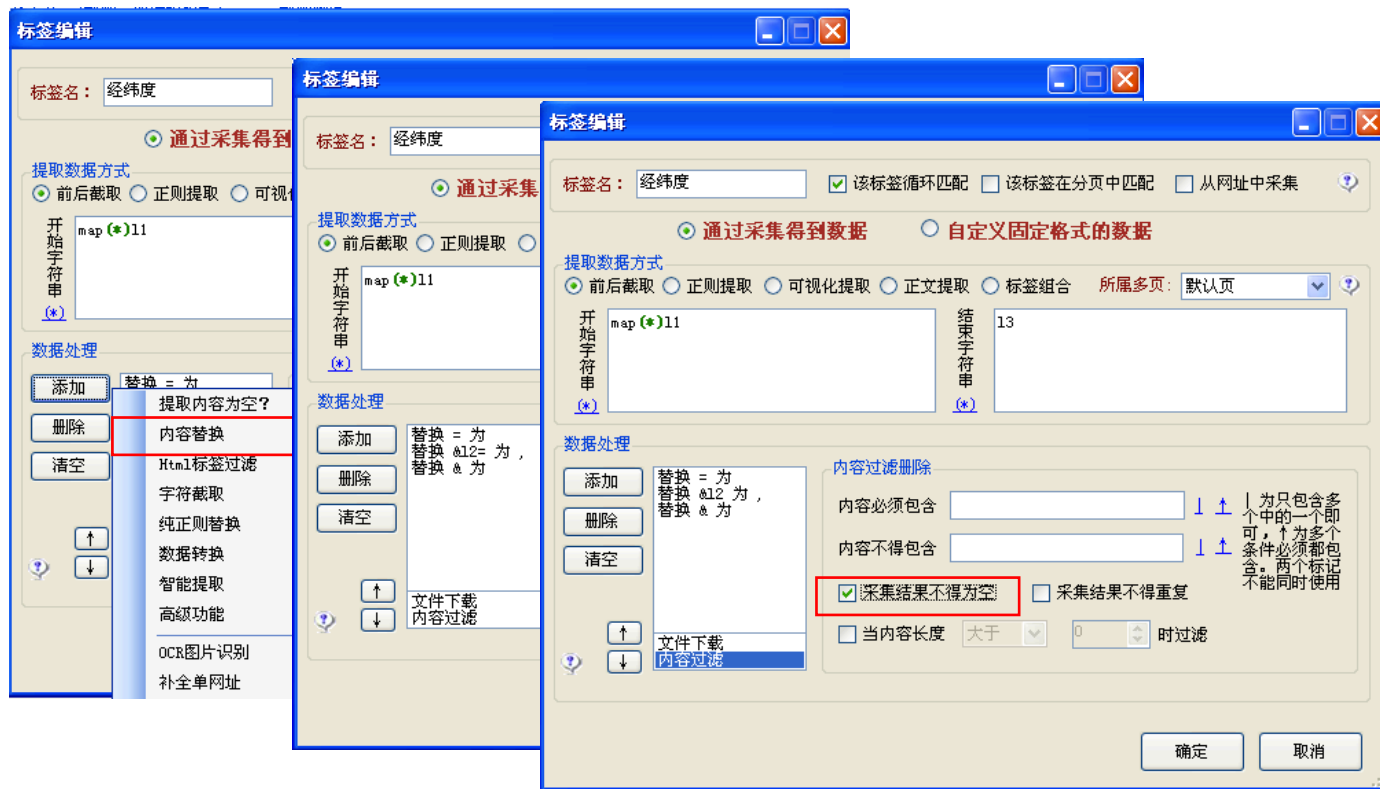
• 火车采集器

1. 添加网址

2. 明确抓取内容

- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

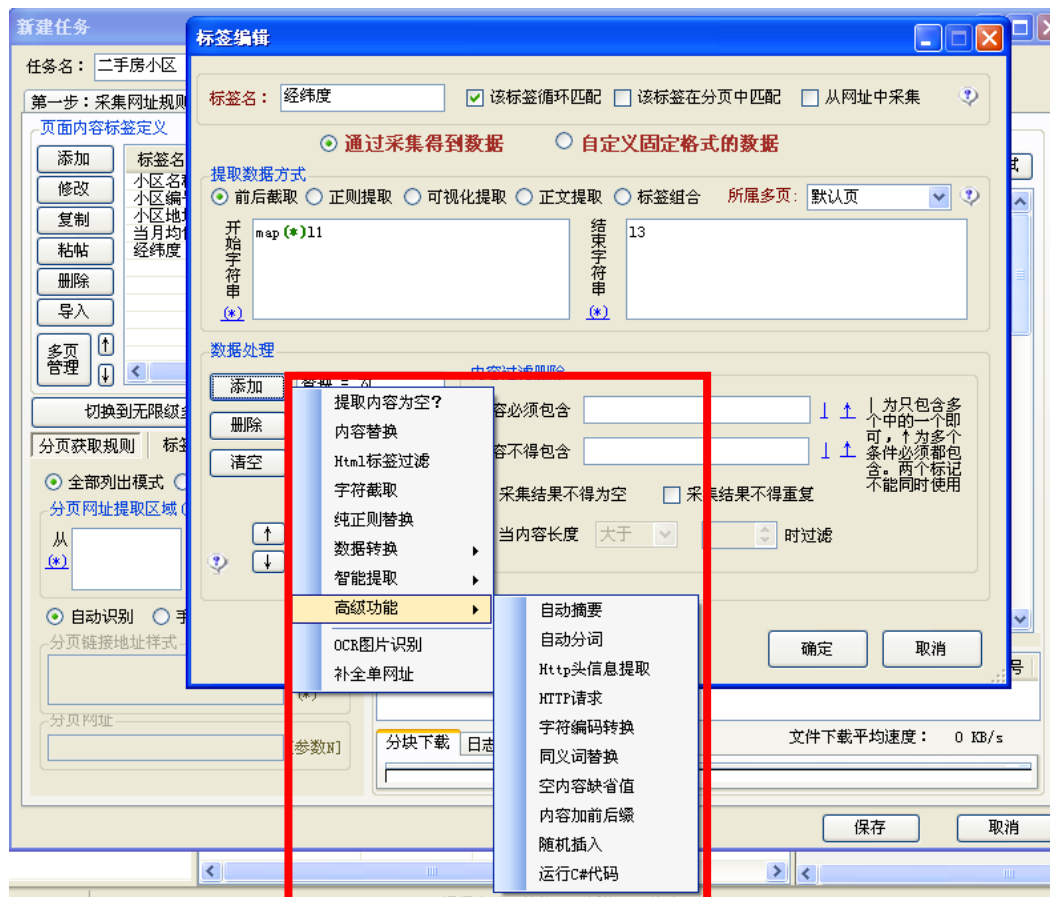
```
...  
<a  
href="http://beijing.anjuke.com/map/sale/#11=39.8956  
069946289&amp;l2=116.486221313477&amp;l3=18&amp;flag  
=1&amp;commname=珠江帝景&amp;commid=55031"  
target="_blank" rel="nofollow">查看地图</a>  
...
```



数据的获取与处理

• 火车采集器

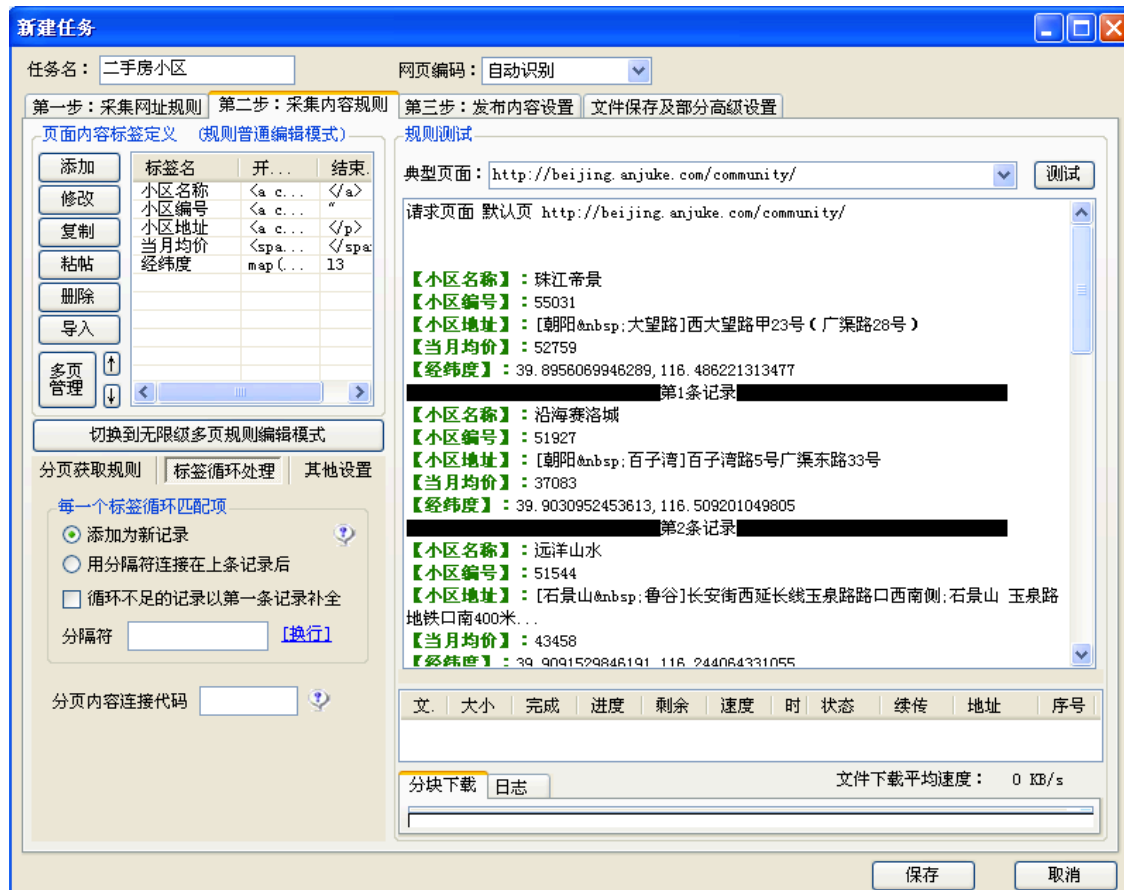
1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试



数据的获取与处理

• 火车采集器

1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试



数据的获取与处理

• 火车采集器

1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试
4. 其他参数

新建任务

任务名: 二手房小区 网页编码: 自动识别

第一步: 采集网址规则 第二步: 采集内容规则 第三步: 发布内容设置 文件保存及部分高级设置

任务运行时线程设置

单任务采集内容线程个数: 3
采集内容间隔时间毫秒数: 400
单任务发布内容线程个数: 3
发布内容间隔时间毫秒数: 400 到 400

Http请求设置

Http代理: 基于Windows身份验证 Http请求

使用IE浏览器代理 不使用代理 使用指定代理

服务器: 端口: 1023
用户名: 密码:

文件下载设置

所有文件保存文件夹: ...
文件链接地址前缀:
文件下载模式: 异步 同步
同时最多文件下载数: 1 单文件下载分块数: 2
下载地址保存为文件: 是 将在“所有文件保存文件夹”中生成“任务id.htm”文件

FTP同步文件上传

FTP服务器: 端口: 21
 启用FTP 用户名: 密码: 匿名 被动模式
文件上传根目录: /FileUpload/ (以/开头和结尾)

上传次序: 先发布数据 先上传文件
 文件上传成功后删除本地文件

测试登录并创建文件夹 测试上传文件Test.zip

任务C#插件: 不使用 插件测试管理
任务PHP插件: 不使用 使用后C#插件将不起作用
任务网址库: 1 网址库批量管理
图片水印: 无 添加 清除

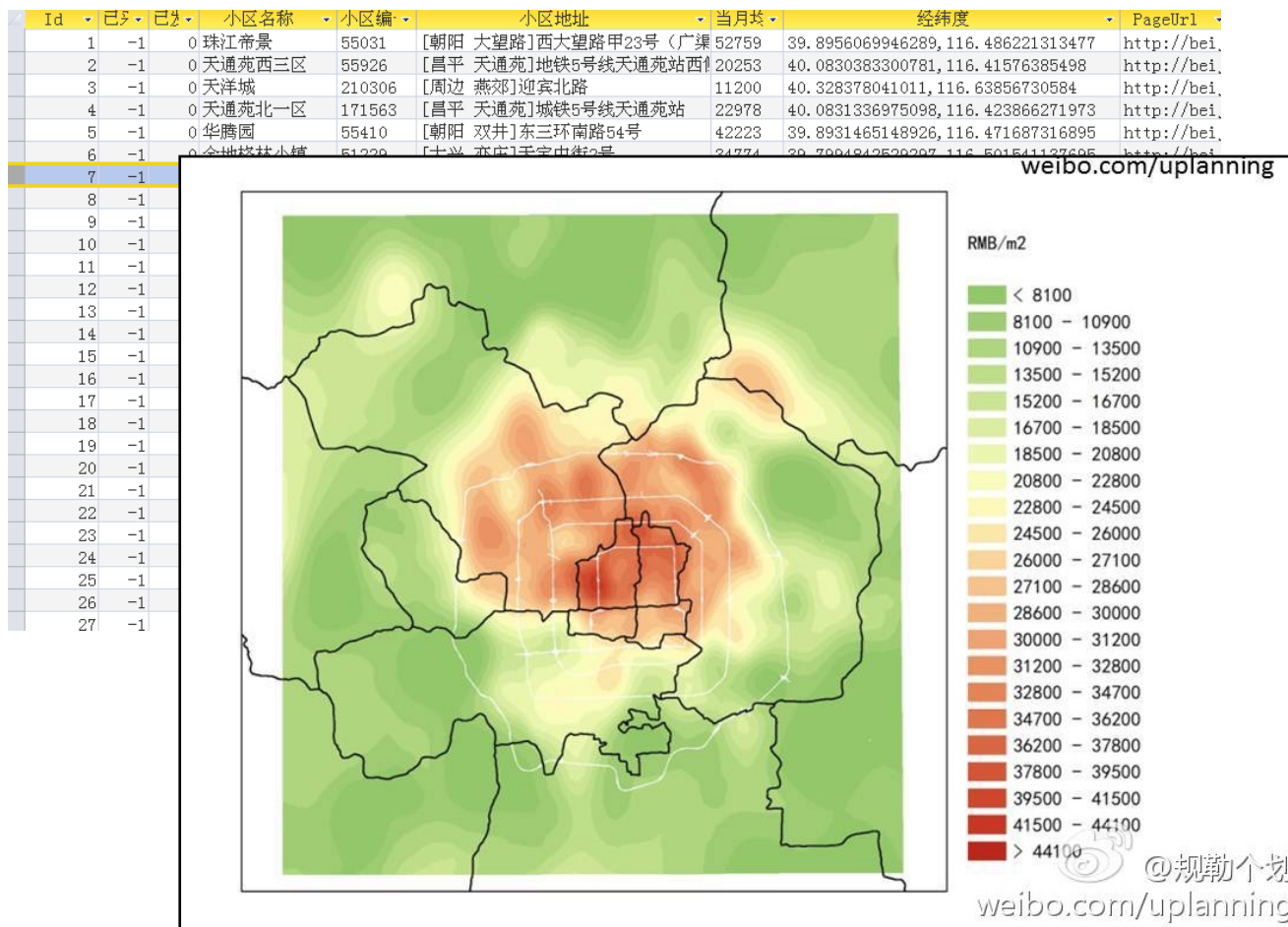
对不符合标签必须包含和不得包含的记录: 删除 标记为未采
 采集数据后直接发布, 不在采集数据库中保存数据, 边采集边发布
 任务运行后进行数据增量导出 (如果设置了发布, 则该项无效)

保存 取消

数据的获取与处理

• 火车采集器

1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试
4. 其他参数
5. Run & pray



数据的获取与处理

• 实际操作：安居客

- 小区列表页
- 小区信息（名称、本月价格、经纬度）
- 翻页
- 小区详细页（网址传递）

安居客北京 > 北京小区

区域: **全部** 朝阳 海淀 丰台 东城 西城 崇文 宣武 石景山 昌平 通州 大兴 顺义 怀柔 房山 门头沟 密

价格: **全部** 8000元以下 8000-1万 1-1.5万 1.5-2万 2-2.5万 2.5-3万 3-4万 4-5万 5万以上

特色: **全部** 绿化优美 别墅 次新房 大型社区 地铁沿线

为您找到 北京 小区 11531 个 排序: 热门度 ▼ 均价 ▼ 涨跌幅 ▼ 竣工时间 ▼

 **西山国际城**
[海淀-田村] 海淀区阜石路甲69号院
竣工日期: 2011年11月
二手房 (1675) | [查看地图](#) | [生活配套](#) | [价格行情](#) | [小区相册](#)
24072元/平米
↑ 1.03%

 **中弘北京像素**
[朝阳-常营] 朝阳北路与草房西路交汇处
竣工日期: 2011年07月
二手房 (4138) | [查看地图](#) | [生活配套](#) | [价格行情](#) | [小区相册](#)
40946元/平米
↑ 1.29%

安居客 北京 - 首页 新房 二手房 租房 商铺写字楼 海外地产 安居头条 房价 问答

首页 > 北京小区 > 海淀小区 > 田村小区 > 西山国际城

西山国际城 + 去对

小区概况 | 二手房 | 租房 | 价格行情 | 房型图 | 实景图 | 生活配套 | 小区问答



西山国际城

受欢迎度: ★★★★★

本月均价: **24072元/m²** 比7月上涨1.03% 贷款

学区房 | 环境优美 | 金牌物业 | 次新小区

生活便利 | 投资推荐 | 住户素质高 | 安全性高

二手房源: **1675套** 租房房源: **2189套**

[查看全部图片](#)

课外了解

抓包工具

数据的获取与处理

- 火车采集器

- 抓包工具：发现隐藏的数据URL

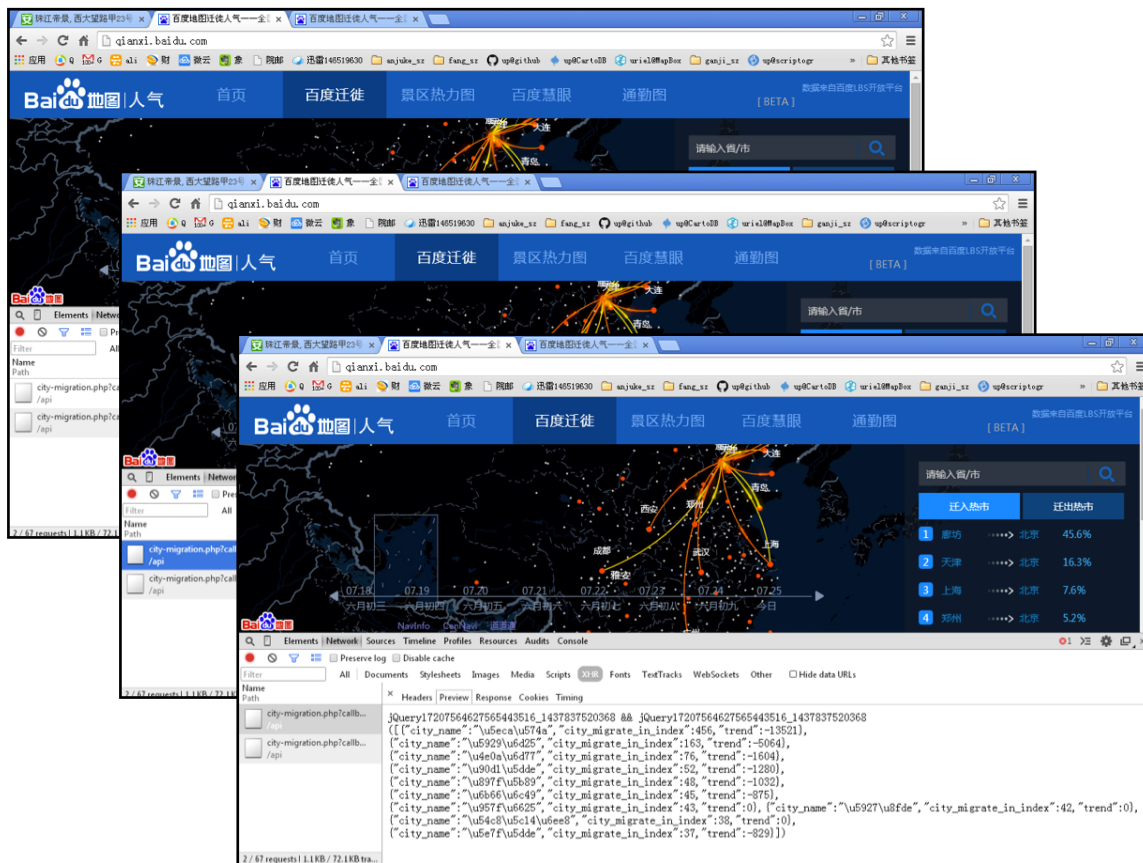


<http://qianxi.baidu.com>

数据的获取与处理

• 火车采集器

- 抓包工具：发现隐藏的数据URL
- XHR = XMLHttpRequest
 - 实现局部更新网页



数据的获取与处理

- 火车采集器

- 抓包工具：发现隐藏的数据URL

```
http://qianxi.baidu.com/api/city-  
migration.php?callback=jQuery1720756  
4627565443516_1437837520368&type=mig  
rate_in&sort_by=low_index&limit=10&c  
ity_name=%E5%8C%97%E4%BA%AC&date_sta  
rt=20150725&date_end=20150725&_  
=1437  
837520779
```



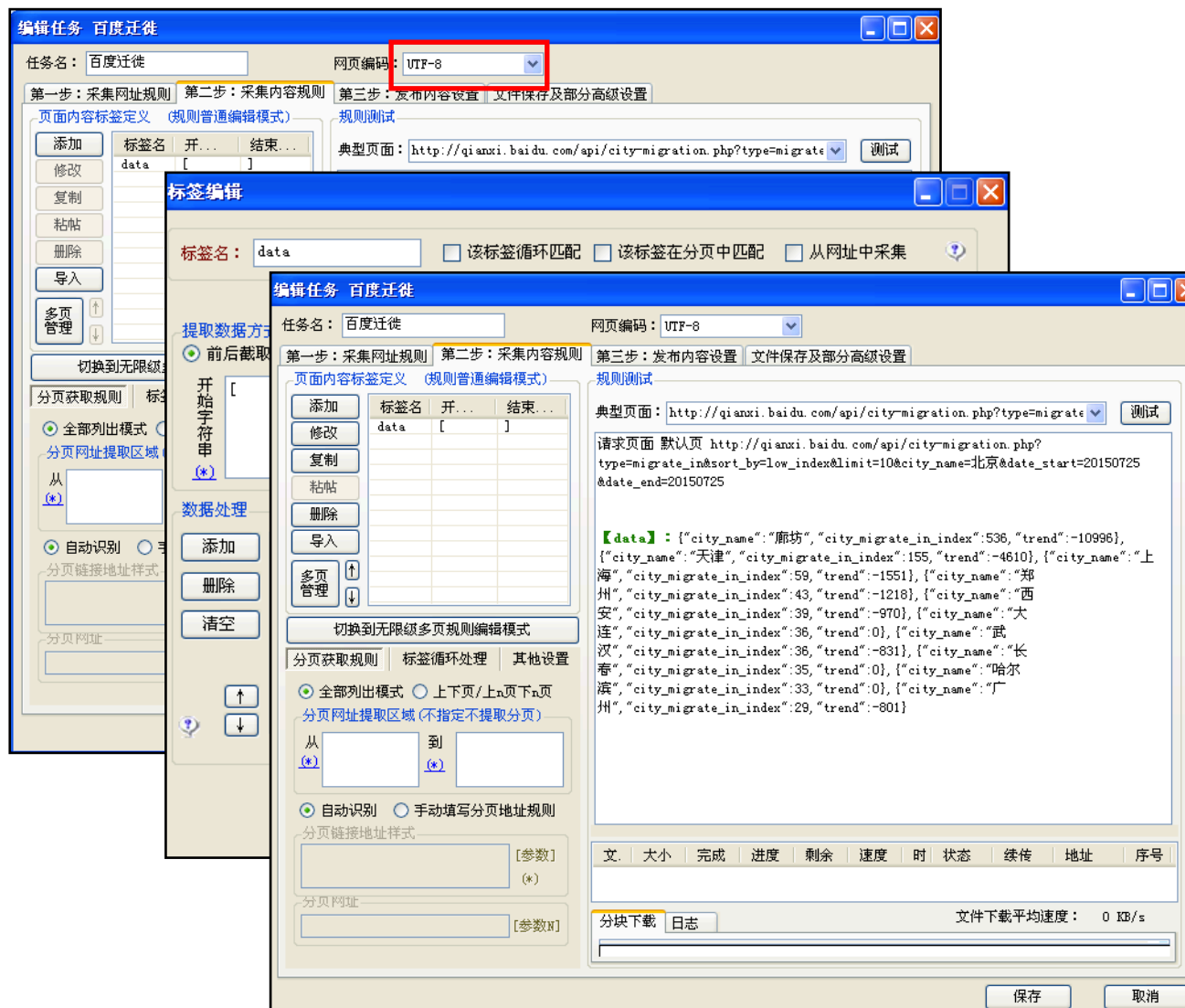
```
http://qianxi.baidu.com/api/city-  
migration.php?type=migrate_in&sort_by=low_index&  
limit=10&city_name=北京  
&date_start=20150725&date_end=20150725
```

```
http://qianxi.baidu.com/api/city-  
migration.php?type=migrate_out&sort_by=low_index  
&limit=10&city_name=北京  
&date_start=20150725&date_end=20150725
```

数据的获取与处理

• 火车采集器

- 抓包工具：发现隐藏的数据URL



数据的获取与处理

• 火车采集器

- 抓包工具：发现隐藏的数据URL
- 按照规则生成URL列表
 - Type X 地名 X 日期

```
http://qianxi.baidu.com/api/city-  
migration.php?type=migrate_in&sort_b  
y=low_index&limit=10&city_name=北京  
&date_start=20150725&date_end=201507  
25
```



Excel



火车采集器

数据的获取与处理

- 实际操作：百度迁徙
 - 查找数据源
 - 尝试参数 postman
 - 抓取结果：完整json、字段化

The screenshot shows the Baidu Migration website interface. The main map displays migration paths from various cities to Beijing. A table on the right lists the top migration sources to Beijing:

Rank	Origin City	Destination City	Percentage
1	廊坊	北京	45.6%
2	天津	北京	16.3%
3	上海	北京	7.6%
4	郑州	北京	5.2%

The browser's developer tools are open, showing the network request for the migration data. The response is a JSON array of objects, each representing a city's migration statistics:

```
jQuery17207564627565443516_1437837520368 && jQuery17207564627565443516_1437837520368
([{"city_name": "\u5eca\u574a", "city_migrate_in_index": 456, "trend": -13521}, {"city_name": "\u5929\u6d25", "city_migrate_in_index": 163, "trend": -5064}, {"city_name": "\u4e0a\u6d77", "city_migrate_in_index": 76, "trend": -1604}, {"city_name": "\u90d1\u5dde", "city_migrate_in_index": 52, "trend": -1280}, {"city_name": "\u897f\u5b89", "city_migrate_in_index": 48, "trend": -1032}, {"city_name": "\u6666\u6c49", "city_migrate_in_index": 45, "trend": -875}, {"city_name": "\u957f\u6625", "city_migrate_in_index": 43, "trend": 0}, {"city_name": "\u5927\u8fde", "city_migrate_in_index": 42, "trend": 0}, {"city_name": "\u54c8\u5c14\u6ee8", "city_migrate_in_index": 38, "trend": 0}, {"city_name": "\u5e7f\u5dde", "city_migrate_in_index": 37, "trend": -829}])
```

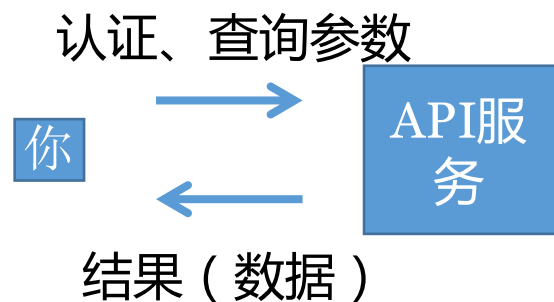
四、基于API的数据采集

感谢清华同衡李栋提供的本部分的资料

数据的获取与处理

• 火车采集器

• API类型数据的获取



API (Application Programming Interface , 应用程序编程接口) 是一些预先定义的函数 , 目的是提供应用程序开发人员基于某软件或硬件得以访问一组程序的能力 , 而又无需访问源码 , 或理解程序内部工作机制的细节。

• 非官方的微博客户端 : Weico、推推.....

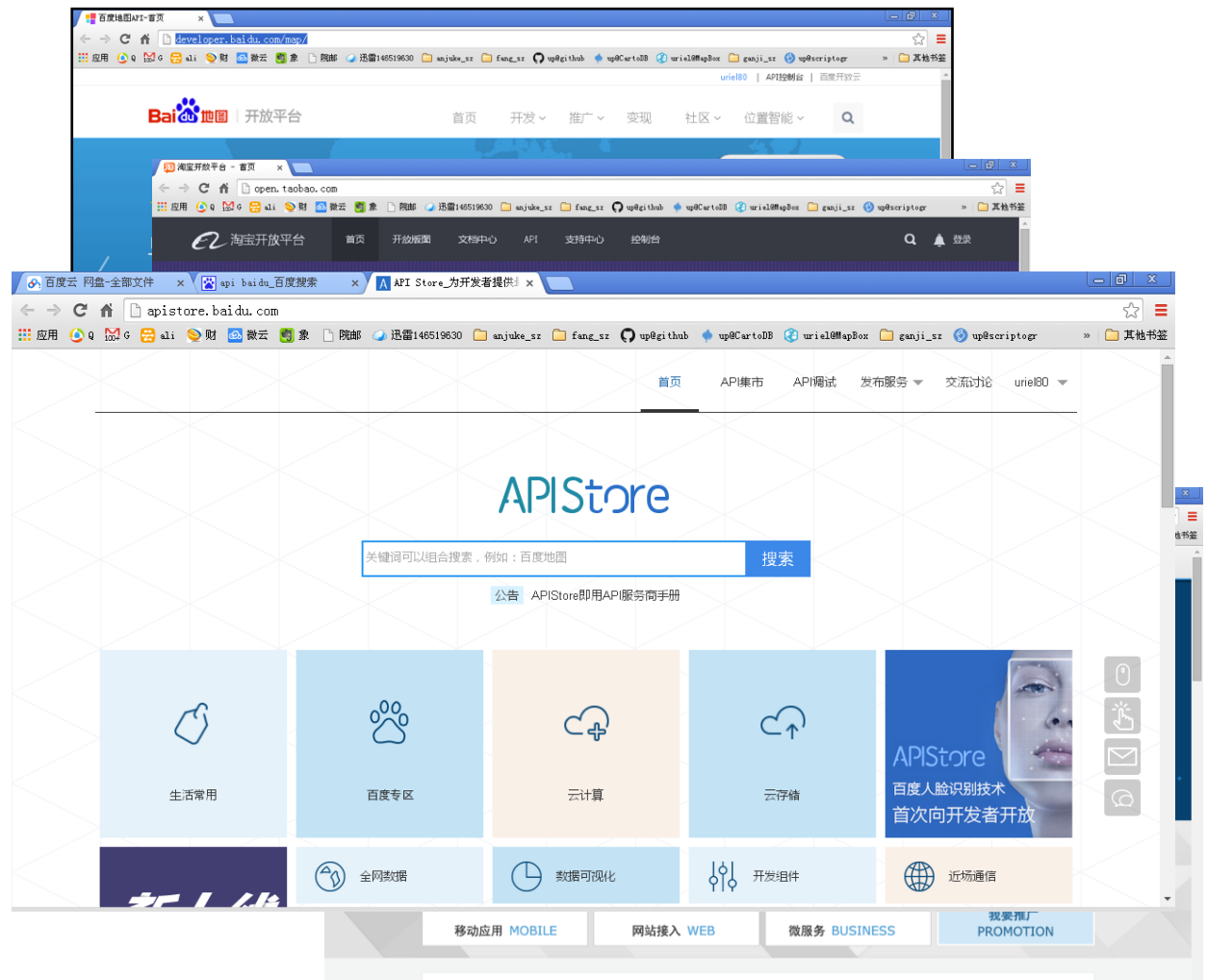
• “XX开放平台”

• API服务商同时也在收集用户的信息 : 双向过程

数据的获取与处理

• API类型数据的获取

- 百度地图API
- 微博API
- 腾讯API
- 淘宝API
-



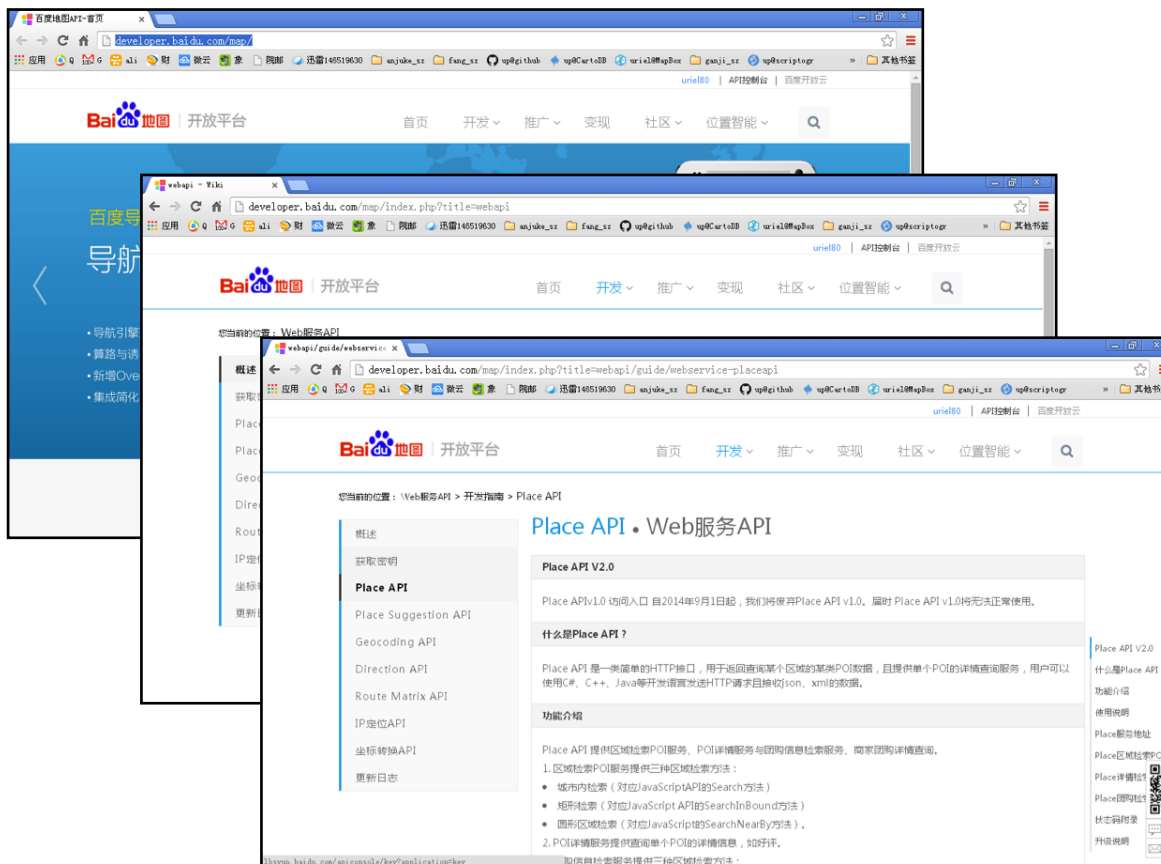
数据的获取与处理

• 火车采集器

- API类型数据的获取

- 百度地图API

- `http://developer.baidu.com/map/index.php?title=webapi`

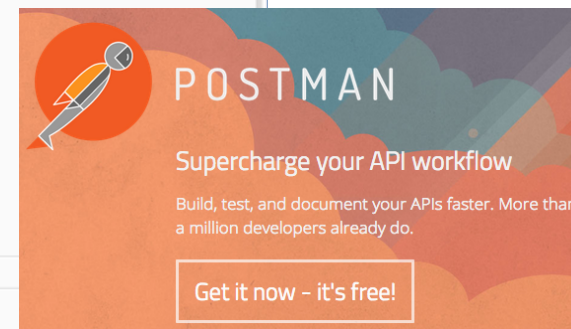
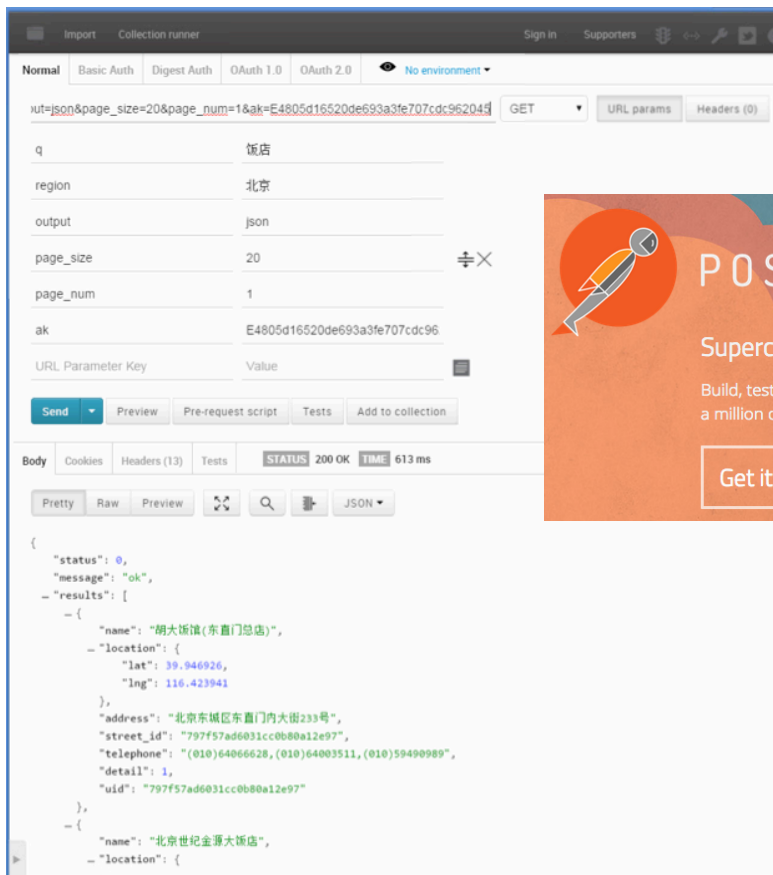


数据的获取与处理

• 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

```
http://api.map.baidu.com/place/v2/search?q=饭店&region=北京&output=json&page_size=20&page_num=1&ak=E4805d16520de693a3fe707cdc962045
```



数据的获取与处理

```
http://api.map.baidu.com/place/v2/search?q=饭店&region=北京  
&output=json&page_size=20&page_num=1&ak=E4805d16520de693a3fe707cdc962045
```

• 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

结果字段（以结果集合里的一条数据为例，灰色表示扩展字段）：

名称	类型	说明
name	string	poi名称
location	object	poi经纬度坐标
lat	float	纬度值
lng	float	经度值
address	string	poi地址信息
telephone	string	poi电话信息
uid	string	poi的唯一标示
detail_info	object	poi的扩展信息，仅当scope=2时，显示该字段，不同的poi类型，显示的detail_info字段不同。
distance	int32	距离中心点的距离
type	string	所属分类，如'hotel'、'cate
tag	string	标签
detail_url	string	poi的详情页
price	string	poi商户的价格
shop_hours	string	营业时间
overall_rating	string	总体评分
taste_rating	string	口味评分
service_rating	string	服务评分

environment_rating	string	环境评分
facility_rating	string	星级（设备）评分
hygiene_rating	string	卫生评分
technology_rating	string	技术评分
image_num	string	图片数
groupon_num	int	团购数
discount_num	int	优惠数
comment_num	string	评论数
favorite_num	string	收藏数
checkin_num	string	签到数

数据的获取与处理

• 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

```
http://api.map.baidu.com/place/v2/search?q=饭店&region=北京&output=json&page_size=20&page_num=1&ak=E4805d16520de693a3fe707cdc962045
```

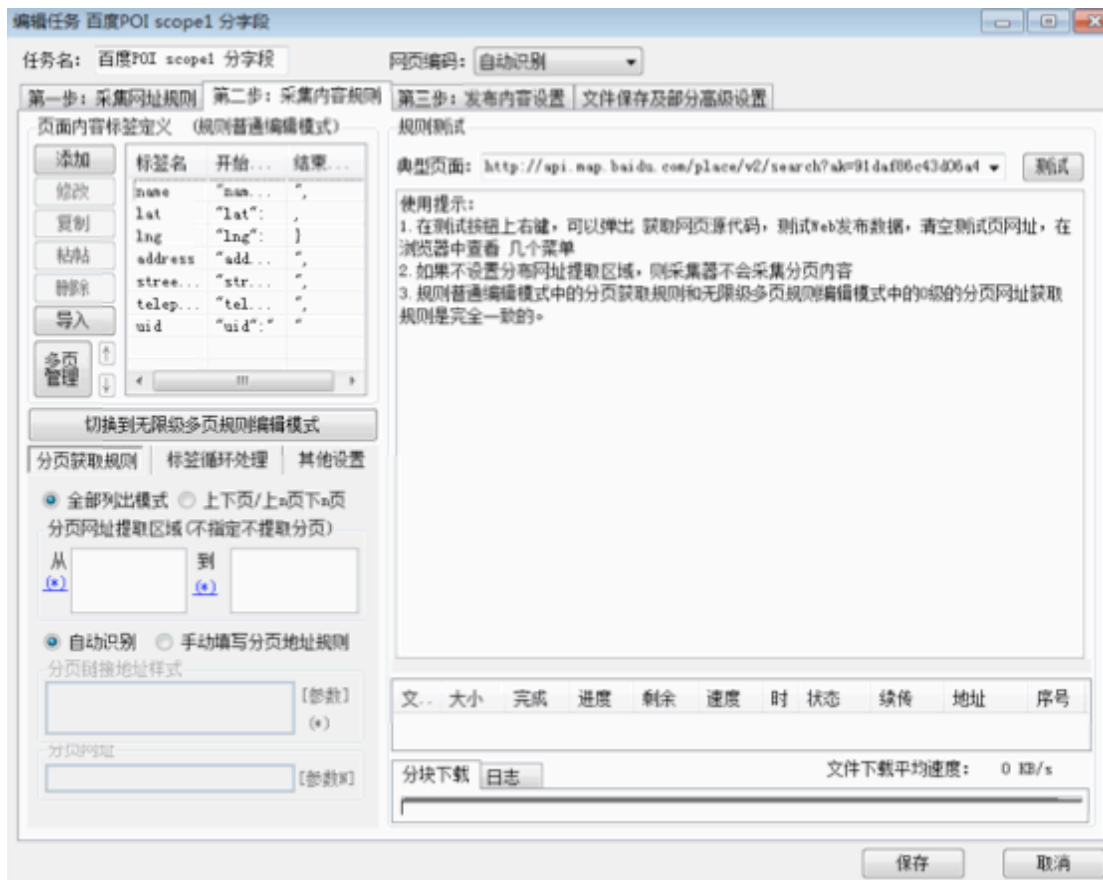
```
{  
  "status": 0,  
  "message": "ok",  
  "total": 15610,  
  "results": [  
    {  
      "name": "北京盘古七星酒店",  
      "location": {  
        "lat": 39.995074,  
        "lng": 116.393207  
      },  
      "address": "中国北京朝阳区北四环中路27号, 盘古大观",  
      "street_id": "925c550f80330fab5a3ac832",  
      "telephone": "13082483603",  
      "detail": 1,  
      "uid": "925c550f80330fab5a3ac832"  
    },  
    {  
      "name": "泰山饭店",  
      "location": {  
        "lat": 40.055374,  
        "lng": 116.335822  
      },  
      "address": "西三旗安宁里北8号(泰山饭店院内南侧)",  
      "street_id": "083a55fa8aece460246e98d1",  
      "telephone": "010-61795939",  
      "detail": 1,  
      "uid": "083a55fa8aece460246e98d1"  
    },  
  ],  
}
```


数据的获取与处理

• 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

```
http://api.map.baidu.com/place/v2/search?q=饭店&region=北京  
&output=json&page_size=20&page_num=1&ak=E4805d16520de693a3fe707cdc962045
```

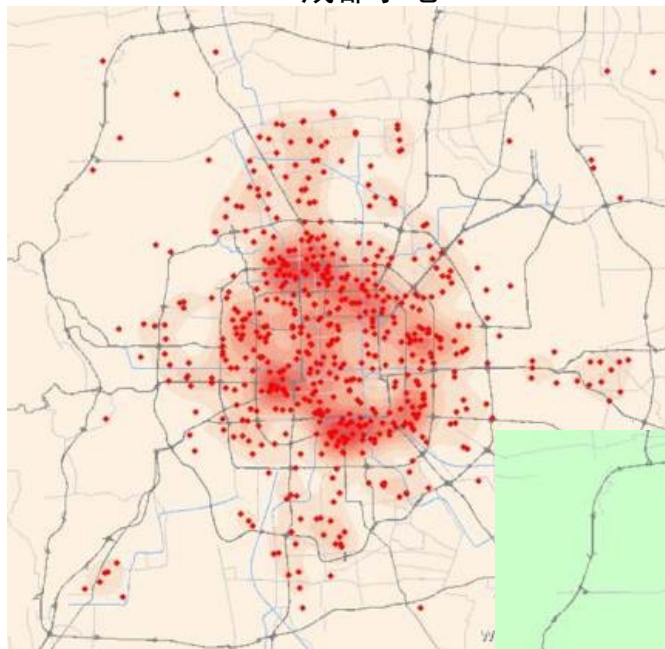


数据的获取与处理

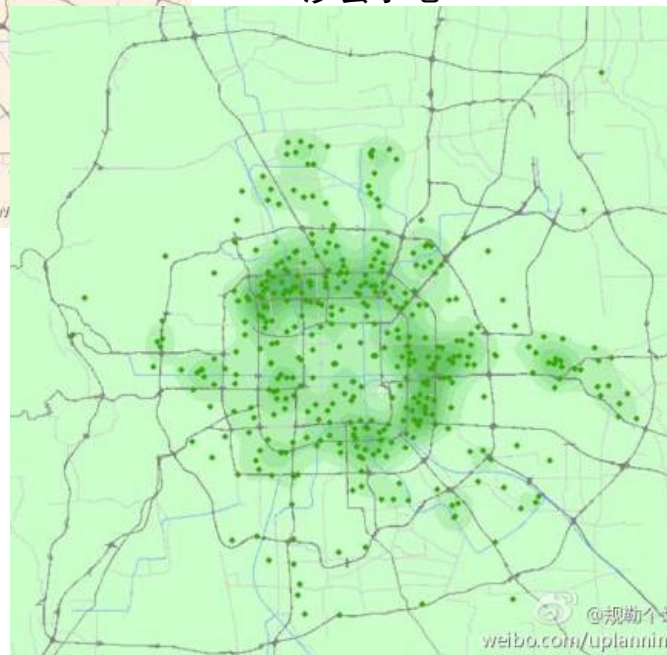
- 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

成都小吃



沙县小吃

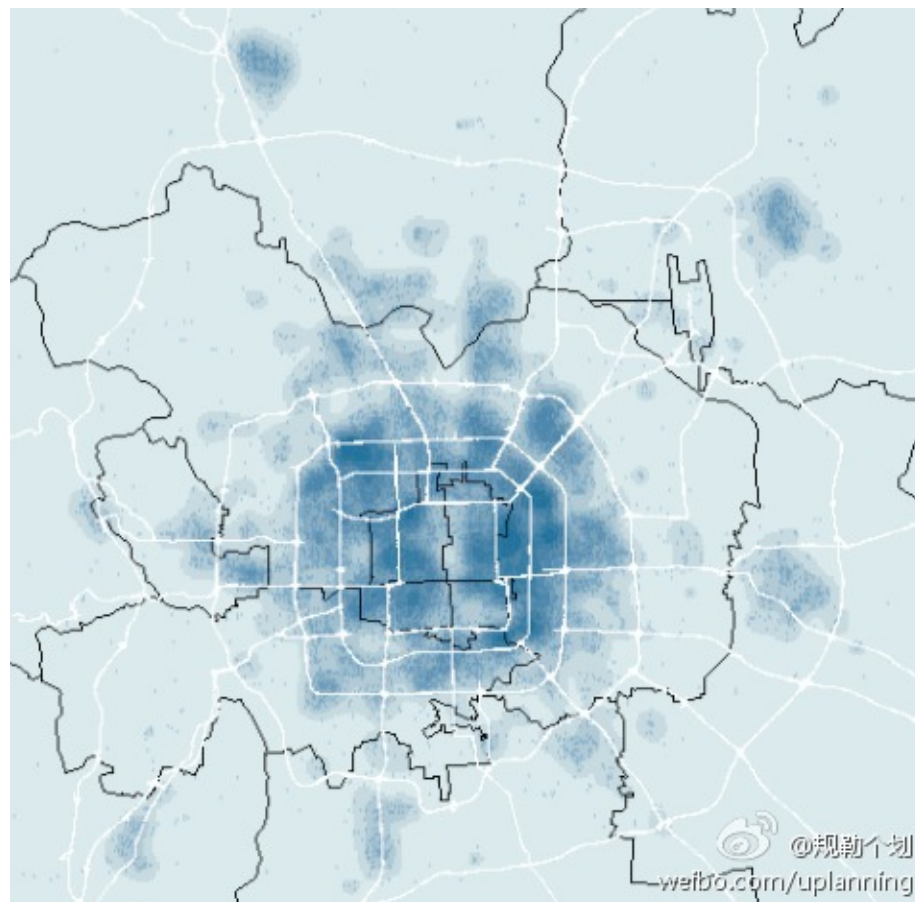


数据的获取与处理

- 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

停车场 ≈20000



数据的获取与处理

- 实际操作：百度地图API
 - 查找接口
 - 尝试参数 postman
 - 抓取结果：完整json、字段化

The screenshot shows a web browser window displaying the Baidu Map API documentation page for Place API V2.0. The page is in Chinese and includes a navigation menu on the left, a main content area, and a right sidebar with a table of contents. The main content area is titled "Place API • Web服务API" and contains the following sections:

- Place API V2.0**: A section with a sub-header "Place API V2.0" and a paragraph stating that Place API v1.0 will be deprecated starting from September 1, 2014.
- 什么是Place API ?**: A section explaining that Place API is a simple HTTP interface for returning POI data for a specific area and providing detailed POI information.
- 功能介绍**: A section titled "功能介绍" (Function Introduction) that lists the services provided by Place API, including:
 - 区域检索POI服务 (Area POI Search Service) with three methods:
 - 城市内检索 (City Search)
 - 矩形检索 (Rectangular Search)
 - 圆形区域检索 (Circular Search)
 - POI详情服务 (POI Detail Service) for providing detailed information for a single POI, such as reviews.

The right sidebar contains a table of contents for the page, including links to "Place API V2.0", "什么是Place API ?", "功能介绍", "使用说明", "Place服务地址", "Place区域检索POI", "Place详情检索", "Place团购检索", "状态码附录", and "升级说明".

数据的获取与处理

• 实际操作：大众点评

- 城市List页
- 分类页
- 分类list页
- 详情页（店名、人均价格、评论数、位置）



数据的获取与处理

• 火车采集器

- API类型数据的获取
- 微博API
 - 位置服务Nearby API

The screenshot shows the Weibo Open Platform API documentation page. The page is titled "开发文档" (Developer Documentation) and "open.weibo.com/wiki". It features a navigation menu with options like "微连接", "微服务", "文档", "支持", "推广", and "我的应用". The main content area is titled "微博API" and lists various API endpoints in a grid format:

- 粉丝服务接口
- 微博接口
- 评论接口
- 用户接口
- 关系接口
- 账号接口
- 收藏接口
- 搜索接口
- 提醒接口
- 短链接接口
- 公共服务接口
- 位置服务接口
- 地理信息接口
- 地图引擎接口
- 支付接口
- OAuth 2.0授权接口

Below the grid, there are two explanatory paragraphs:

标识为 的接口为高级权限接口，需要申请才可以调用，申请方式请见 [高级接口申请](#)

标识为 的接口为需要单独向用户提出SCOPE授权请求的接口，用户单独授权后可以调用，具体使用方法请见 [SCOPE](#)

At the bottom, there is a section titled "粉丝服务 (新手接入指南)" (Fan Service (New User Guide)) with a table of endpoints:

粉丝服务 (新手接入指南)		
读取接口	接收消息	接收用户私信、关注、取消关注、@等消息接口
写入接口	发送消息	向用户回复私信消息接口

数据的获取与处理

• 火车采集器

- API类型数据的获取
- 微博API
 - 位置服务Nearby API

place/nearby_timeline

获取某个位置周边的动态

URL

https://api.weibo.com/2/place/nearby_timeline.json

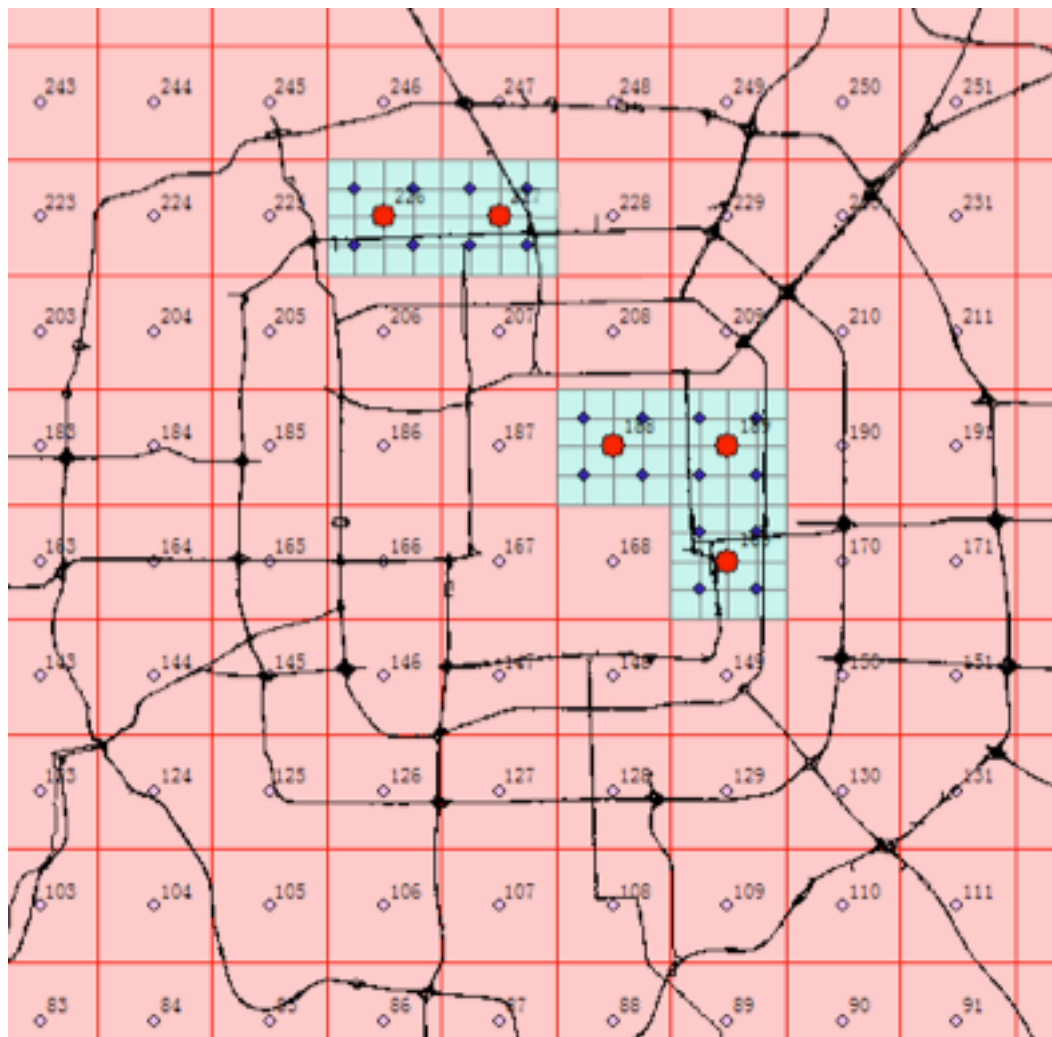
请求参数

	必选	类型及范围	说明
access_token	true	string	采用OAuth授权方式为必填参数，OAuth授权后获得。
lat	true	float	纬度。有效范围：-90.0到+90.0，+表示北纬。
long	true	float	经度。有效范围：-180.0到+180.0，+表示东经。
range	false	int	搜索范围，单位米，默认2000米，最大11132米。
starttime	false	int	开始时间，Unix时间戳。
endtime	false	int	结束时间，Unix时间戳。
sort	false	int	排序方式。默认为0，按时间排序；为1时按与中心点距离进行排序。
count	false	int	单页返回的记录条数，最大为50，默认为20。
page	false	int	返回结果的页码，默认为1。
base_app	false	int	是否只获取当前应用的数据。0为否（所有数据），1为是（仅当前应用），默认为0。
offset	false	int	传入的经纬度是否是纠偏过，0：没纠偏、1：纠偏过，默认为0。

数据的获取与处理

- 火车采集器

- API类型数据的获取
- 微博API
 - 位置服务Nearby API



数据的分析

```
{
  "created_at": "Sun Apr 19 23:59:52 +0800 2015",
  "id": "3833492062372488",
  "mid": "3833492062372488",
  "idstr": "3833492062372488",
  "text": "逗比 [花心][心]@尤哈Y0j1 http:\n",
  "source_allowclick": 0,
  "source_type": 1,
  "source": "<a href=\"http://app.weib",
  "favorited": false,
  "truncated": false,
  "in_reply_to_status_id": "",
  "in_reply_to_user_id": "",
  "in_reply_to_screen_name": "",
  "pic_ids": [
    "74889468jw1erbbk0k1w0j21w01w0b2d",
    "74889468jw1erbbk58s0kj21w01w0u11",
    "74889468jw1erbbkrr0gkj21w01w0kjp",
    "74889468jw1erbbkv0mej21w01w0hdx",
    "74889468jw1erbb14c7eij21601kwqv5",
    "74889468jw1erbb1722c2j21601kw1ky",
    "74889468jw1erbb1ap5gej21601lux6p",
    "74889468jw1erbbjt1zt4j21601kwe82",
    "74889468jw1erbb1ckft6j21601kwkj1"
  ],
  "thumbnail_pic": "http://ww3.sinaimg",
  "bmiddle_pic": "http://ww3.sinaimg.c",
  "original_pic": "http://ww3.sinaimg.",
  "geo": {
    "type": "Point",
    "coordinates": [
      39.938564,
      116.454475
    ]
  },
  "user": {
    "id": "1955107944",
    "idstr": "1955107944",
    "class": 1,
    "screen_name": "_AimeE_77",
    "name": "_AimeE_77",
    "province": "11",
    "city": "1",
    "location": "北京 东城区",
    "description": "丢了的东西再好，也是",
    "description": "丢了的东西再好，也是垃圾",
    "url": "",
    "profile_image_url": "http://tp1.sinaimg.cn/1955107944/50x50",
    "cover_image_phone": "http://ww4.sinaimg.cn/profile/1955107944",
    "domain": "",
    "weihao": "",
    "gender": "f",
    "followers_count": 743,
    "friends_count": 241,
    "pagefriends_count": 4,
    "statuses_count": 62,
    "favourites_count": 552,
    "created_at": "Sat May 21 21:37:32 +0800 2015",
    "following": false,
    "allow_all_act_msg": false,
    "geo_enabled": true,
    "verified": false,
    "verified_type": -1,
    "remark": "",
    "ptype": 0,
    "allow_all_comment": true,
    "avatar_large": "http://tp1.sinaimg.cn/1955107944/100x100",
    "avatar_hd": "http://ww3.sinaimg.cn/profile/1955107944",
    "verified_reason": "",
    "verified_trade": "",
    "verified_reason_url": "",
    "verified_source": "",
    "verified_source_url": "",
    "follow_me": false,
    "online_status": 0,
    "bi_followers_count": 71,
    "lang": "zh-cn",
    "star": 0,
    "mbtype": 2,
    "mbrank": 1,
    "block_word": 0,
    "block_app": 0,
    "level": 1,
    "type": 1,
    "ulevel": 0,
    "badge": {
      "uc_domain": 0,
      "enterprise": 0,
      "anniversary": 0,
      "taobao": 0,
      "travel2013": 0,
      "gongyi": 0,
      "gongyi_level": 0,
      "bind_taobao": 1,
      "hongbao_2014": 1,
      "suishoupai_2014": 0,
      "dailv": 0,
      "zongyiji": 0,
      "yiqipao_2015": 0
    },
    "badge_top": "",
    "has_ability_tag": 0,
    "extend": {
      "privacy": {
        "mobile": 1
      },
      "mbprivilege": "000000000000000000000000",
      "credit_score": 80,
      "urank": 18
    },
    "annotations": [
      {
        "place": {
          "poiid": "B2094750D66DA0F8429F",
          "lon": 116.454475,
          "title": "Let's Burger Plas",
          "type": "checkin",
          "lat": 39.938564
        },
        "client_mblogid": "iPhone-5AF85807-1"
      }
    ],
    "reposts_count": 0,
    "comments_count": 3,
    "attitudes_count": 5,
    "mlevel": 0,
    "visible": {
      "type": 0,
      "list_id": 0
    }
  },
  "url_objects": [
    {
      "url_ori": "http://t.cn/z85psnr",
      "object_id": "1022:100101B2094750D66DA0F8429F",
      "info": {
        "url_short": "http://t.cn/z85psnr",
        "url_long": "http://weibo.com/100101B2094750D66DA0F8429F",
        "type": 36,
        "result": true,
        "title": "",
        "description": "",
        "last_modified": 1434031652,
        "transcode": 0
      },
      "object": {
        "object_id": "1022:100101B2094750D66DA0F8429F",
        "containerid": "100101B2094750D66DA0F8429F",
        "object_domain_id": "1022",
        "object_type": "place",
        "safe_status": 1,
        "show_status": "10",
        "act_status": "00",
        "last_modified": "Thu Jun 11 22:14:34 2015",
        "timestamp": 143403165613,
        "uuid": 3524408906907855,
        "activate_status": "0",
        "object": {
          "summary": "朝阳区三里屯路11号三里屯Village北区B1楼NLG-461",
          "position": " 39.938564 116.454475",
          "object_type": "place",
          "id": "1022:100101B2094750D66DA0F8429F",
          "display_name": "Let's Burger Plas",
          "address": {
            "region": "北京市",
            "formatted": "北京市朝阳区",
            "fax": "餐饮美食",
            "email": "",
            "postal_code": "",
            "locality": "北京市",
            "street_address": "朝阳区三里屯路11号三里屯Village北区B1楼NLG-461",
            "telephone": "",
            "country": "中国"
          },
          "keyword": "Let's Burger Plas",
          "action": [
            {
              "place_review": {
                "place_review": {
                  "image": {
                    "height": "88",
                    "width": "88",
                    "url": "http://ww3.sinaimg.cn/large/4e704b16jw100101B2094750D66DA0F8429F"
                  },
                  "target_url": "http://weibo.cn/pages/100101B2094750D66DA0F8429F",
                  "biz": {
                    "biz_id": "100101",
                    "containerid": "100101B2094750D66DA0F8429F"
                  },
                  "url": "http://weibo.cn/pages/100101B2094750D66DA0F8429F",
                  "mobile": {
                    "page_id": "100101B2094750D66DA0F8429F",
                    "card": {
                      "scheme": "sinaweibo://pageinfo?pageid=100101",
                      "status": 1,
                      "contents": [
                        {
                          "text": "Let's Burger Plas",
                          "朝阳区三里屯路11号三里屯Village北区B1楼NLG-461"
                        }
                      ],
                      "is_async": 0,
                      "pic": "http://ww3.sinaimg.cn/large/4e704b16jw100101B2094750D66DA0F8429F",
                      "type": 0
                    }
                  },
                  "url": {
                    "scheme": "sinaweibo://pageinfo?pageid=100101",
                    "status": 1,
                    "name": "Let's Burger Plas"
                  }
                }
              }
            }
          ],
          "actions": [
            {
              "like_count": 52,
              "follower_count": 0,
              "asso_like_count": 0,
              "card_info_un_integrity": false
            }
          ],
          "darwin_tags": [
            {
              "rid": "0_0_100101_2789638820038132338",
              "distance": 5000
            }
          ]
        }
      }
    }
  ]
}
```

对比

- 软件 vs. 命令行/脚本
 - 运行速度和效率
 - 软硬件成本
 - 软件学习的成本
 - 软件：入门相对容易
 - 命令行/脚本：复用性更高
 - 功能
 - 功能的局限：火车采集器只支持get的方式，不支持post
 - 命令行/脚本：开源、库（功能）丰富

五、大作业安排

大作业要求

1. 每个人提交一份课程感想，800字以上，包括建议
2. 利用所提供的北京二环的数据，开展城市研究、规划设计或城市管理支持工作
3. 鼓励以小组为单位共同开展现场调研
4. 如果与其他课程共用课程作业，则建议自觉提高作业难度
5. 10月9日前，将分组情况报给课代表（包括是否与其他课程共用作业，需要提供课程名称）
 - 每组请有一名目前了解ArcGIS的同学
 - 鼓励跨院系跨专业

关于场地调研



睡睡梦想家 ★

南锣鼓巷真的连锁店越来越多了，哈根达斯，鲜果时间之类的都来凑热闹，已经没有只有在这里才能猜到的好吃的感觉了，等啥时候文字也没了，我就再也不会来了!!!

北京·南锣...

2015-8-12 13:54 来自 小米Note

收藏

转发

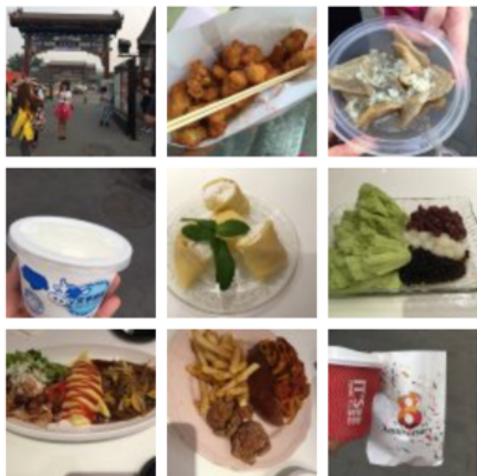
评论 4

👍 2



在他乡在远方 ✓

#南锣鼓巷#抱着绝对不吃连锁店的心里，看着人多就跑去试一试，结果证明都很难吃，最后身心疲惫找餐厅，蛋包饭很不错，份量还很大。



2015-5-2 13:45 来自 iPhone 6 Plus



小恶魔呆喜 ★

前几天再去南锣鼓巷，已经不是几年前我第一次的时候那个样子了，满眼都是花花绿绿的男男女女，店铺也都越来越充满了商业气息，开的很多店都是连锁性质的了，没有了过去那种不可替代独一无二的特性~特别伤心~



小飞侠不叫阿童木

真心觉得南锣鼓巷好令我失望，里面的礼品店又不是太有特色，小吃店又好像很多都是连锁的，至少是快餐式的。最绝的是走到南锣鼓巷到北锣鼓巷交接等人的时候，毛家臭豆腐毛家臭豆腐毛家臭豆腐的喇叭一直在响，循环循环。有臭豆腐卖的地方，该怎么说好呢。相比之下还是没什么人的五道营给我感觉好。



2013-9-30 23:01 来自 搜狗高速浏览器

美食



首页

团购

找优惠

订座

会员卡

同城活动

社区

北京餐厅 > 东城区 > 东四 > 西餐 > 其他西餐 > Slow Boat Brewery(东四八条店)

Slow Boat Brewery(东四八条店)

手机买单 积分抵现

其它1家分店

★★★★★ 483条评论 人均: 95元 | 口味: 9.2 | 环境: 8.6 | 服务: 8.9

地址: 东城区 东四八条56-2号

电话: 010-65385537

更多信息

写点评



1175

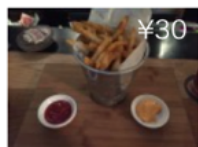
添加照片

推荐菜 环境 价目表 官方相册 品牌故事

炸薯条 (118) 啤酒 (112) 经典汉堡 (97) 酸黄瓜 (62) 薯条汉堡 (39) 蓝色奶酪汉堡 (32) 汉堡 (23)

薯条 (22) 猴拳 (18) 香草黑啤 (14) IPA (14) 黑啤 (11) 烤香肠 (11) 咖啡黑啤 (8)

更多



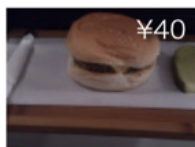
¥30

炸薯条



¥60

啤酒



¥40

经典汉堡



酸黄瓜



¥50

薯条汉堡



蓝色奶酪汉堡



你可能会喜欢

推广



法兰客(Franc...)

★★★★★

东城区

人均 ¥125

大众点评



西餐达人

Petrichor1106 VIP

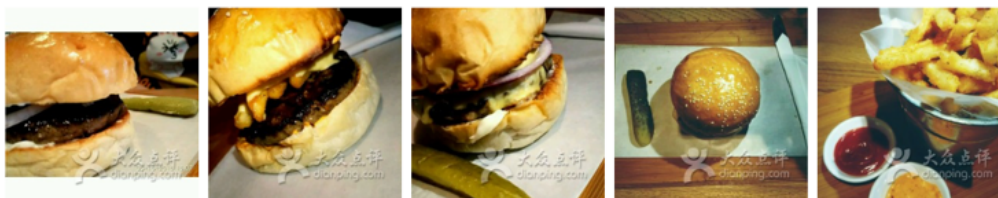
口味: 4 环境: 4 服务: 4 人均: 50

很隐蔽的店, 路过n多次都没发现它的存在, 只闻到了阵阵麦香鱼的香气。某天偶然发现, 今日终于掀开了它神秘的面纱。

整体氛围感觉很热闹, 歪果仁很多, 空位很少。服务员有说.....

更多

喜欢的菜: 炸薯条 经典汉堡 薯条汉堡 蓝色奶酪汉堡



05-16 更新于16-05-16 20:11 Slow Boat Brewery 签到点评 赞(2) 回应(1) 收藏 举报



christylu VIP

口味: 4 环境: 4 服务: 4 人均: 105

竟然是我美国来的同事带我来的, 让我这个北京人有点.....他是在一篇国外的介绍北京各大brewery 文章上看到的。他自己昨晚来过, 所以今晚我们先去试了文章中提到的另一家.....

更多

喜欢的菜: 经典汉堡 香草黑啤 西海岸IPA



05-22 Slow Boat Brewery 赞(1) 回应(3) 收藏 举报

街景地图

腾讯地图

东四八条



北京市 地图 地形 卫星 街景

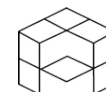
反馈 清空 分享 标记

北京市,北京市,东城区,东四九条 拍摄日期: 2013.11



- <http://map.qq.com>
- 非现场调查、街道界面测量、公共空间评价
- 平面地图也是一个重要的数据源

清华大学



BCL
Beijing City Lab

在线调查 <http://www.sojump.com>



注册

登录

QQ登录

首页

应用场景

企业版服务

样本服务

问卷中心

since 2006

全球最大的免费问卷调查平台

马上和我们873万发布者、4.48亿填写者一起使用问卷星吧!

免费使用



- 不限使用时间
- 不限答卷数
- 免费下载结果



问卷调查

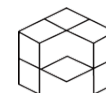
在线考试

表单

投票

360测评

清华大学



BCL
Beijing City Lab

中国收缩城市研究基本态势调查



问卷星™

专业的在线问卷调查平台 www.sojump.com

中国收缩城市研究基本态势调查问卷—自定义查询

作者：bicorn 时间：2016年4月17日

调查背景：[参考资料](#)

调查方法：

开始时间：2016-4-6 结束时间：2016-4-17

样本总数：474 份

原始数据来源：<http://www.sojump.com/report/7655145.aspx?qc=>

本报告分析内容：自定义查询

本报告样本筛选规则：

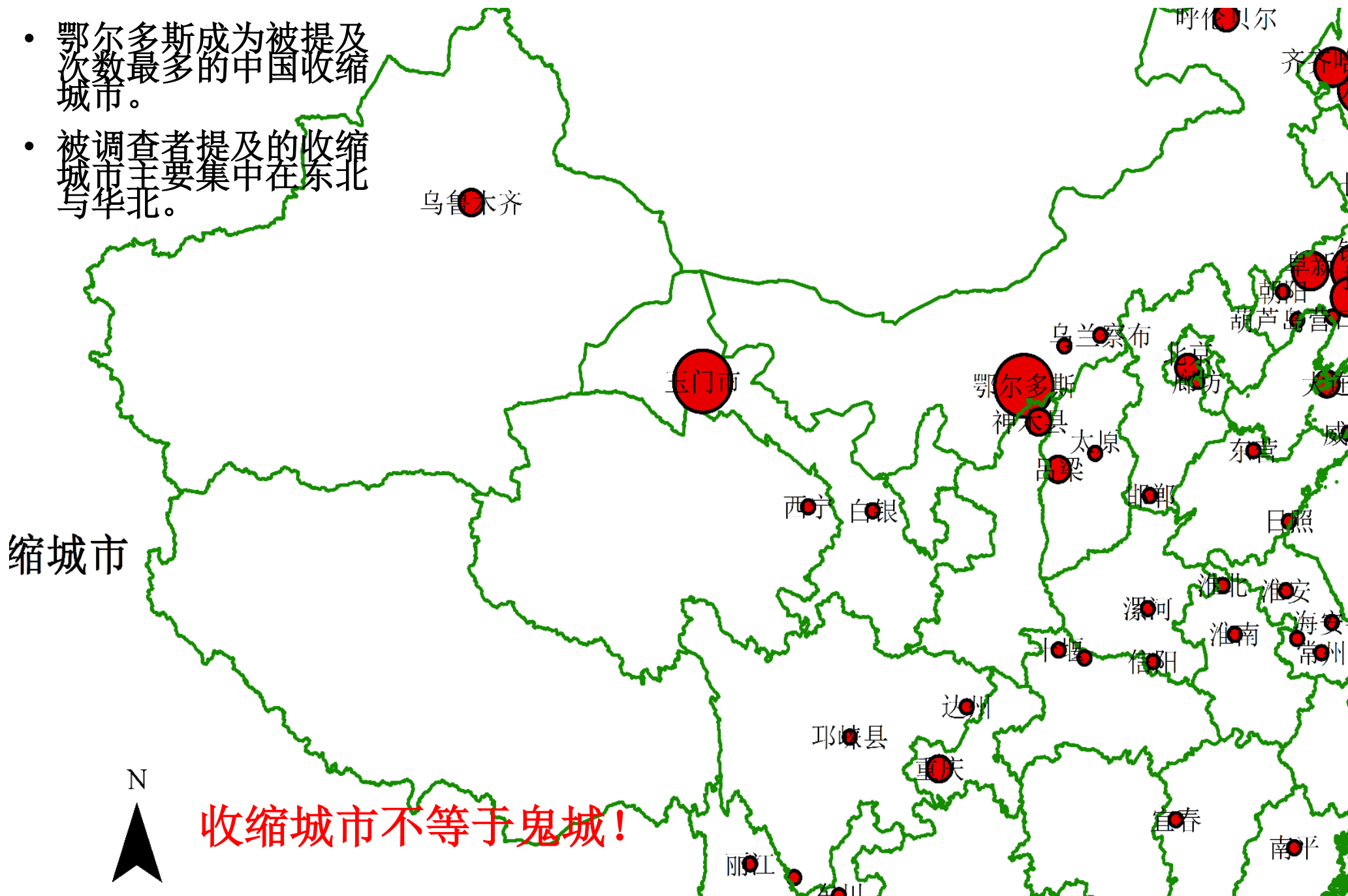
本报告包含样本数量：474份

清华大学



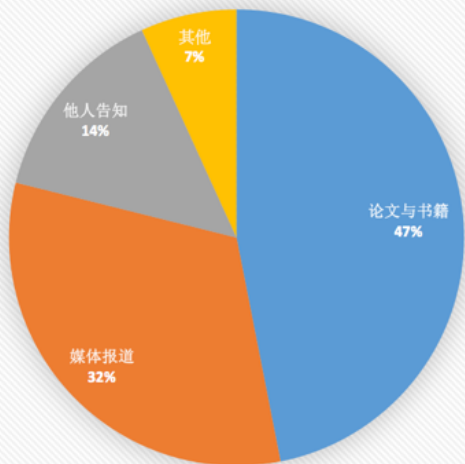
例：中国收缩城市研究基本态势调查

- 鄂尔多斯成为被提及收缩次数最多的中国收缩城市。
- 被调查者提及的收缩城市主要集中在东北与华北。

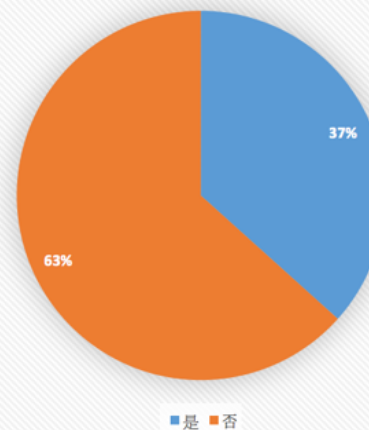


例：中国收缩城市研究基本态势调查

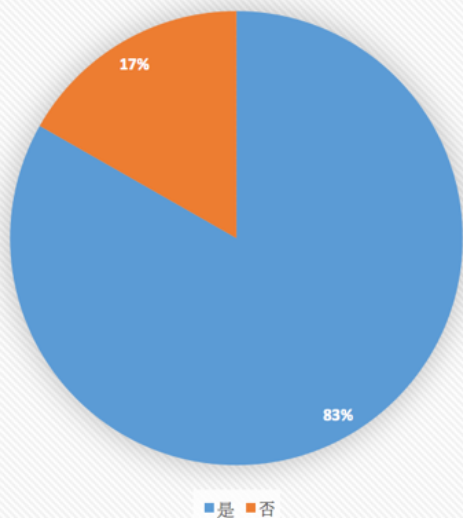
收缩城市了解渠道分析



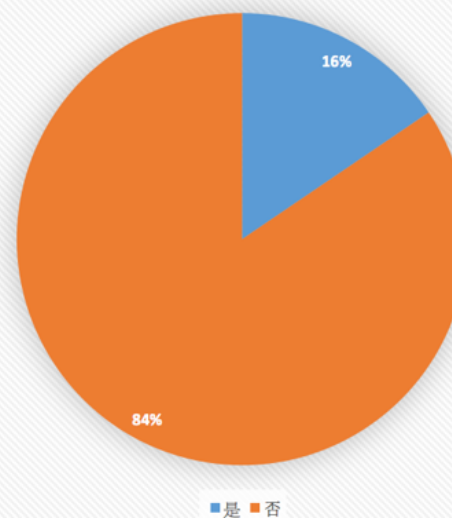
是否听说过“BCL中国收缩城市研究网络”



是否认为中国有收缩城市



是否做过收缩城市的相关研究



基于互联网的调研 vs 现场调查

- （上述开放数据以及街景图片）
- 不受天气影响（午后的雷暴）
- 不受地理距离影响（总体城市设计课）
- 同等样本下，人力和财力成本更低

- 样本有偏（盲人摸象、充分利用有偏性）
- 纸上得来终觉浅，现场调研多少次都不为过

现场调研与互联网调研的结合

本科生的课程作业展示

- 东四历史街区街道界面及商业化发展调研
- 如果没有时间，则将在课后分发给大家

课后安排

- 初步分组，4人一组（自我组织），设定选题，北京旧城，围绕开发、形态、功能、活动、活力等维度（下一节课具体安排作业、分发北京旧城数据和介绍数据获取方法）
- 国庆后计划邀请外界人士介绍数据抓取的操作
- 阅读材料待放到课程网站（火车采集器入门材料）
 - <http://www.beijingcitylab.com/big-data-and-urban-planning/>
- 答疑
 - ylong@tsinghua.edu.cn
 - 建筑学院新501办公室（默认每周五上午10:00-11:30）



龙瀛, ylong@tsinghua.edu.cn, 新建筑馆501, 13661386623



北京城市实验室
Beijing City Lab

<http://www.beijingcitylab.com>



新浪微博: 龙瀛a1_b2 北京城市实验室BCL

微信公众号: [beijingcitylab](https://www.beijingcitylab.com)

清华大学

