

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

2017年10月20日

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（9月22日） W1
2. 变化中的中国城市与未来城市（9月29日） W2
3. 城市大数据类型与典型数据介绍（10月13日） W4（分组完成）
4. **城市大数据的获取（10月20日） W5（分发第一版本的北京老城数据）**
5. 城市大数据的统计与分析（10月27日） W6（本周末日晚上第一次课外沙龙）
6. 城市大数据的可视化（11月3日） W7
7. 学生作业思路汇报与点评（11月10日） W8（预计占用第二大节部分时间）

8. 基于图片大数据的城市空间研究（11月17日） W9
9. 城市大数据挖掘：空间句法（11月24日） W10
10. 城市大数据挖掘：城市网络分析（12月1日） W11
11. 数据增强设计（12月8日） W12
12. 战略、总体及详细规划中的大数据应用（12月15日） W13
13. 城市设计中的大数据应用（12月22日） W14
14. 大模型：跨越城市内与城市间尺度的大数据应用（12月29日） W15
15. 学生作业终期汇报与点评（1月5日） W16（一次或分两次）

上一讲回顾

1. 城市定义：行政、实体、功能地域三个视角
2. 数据、新数据及分类
3. 典型的城市空间大数据
4. 若干思考

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





在线服务 >> 审批结果 >> 建设用地规划许可证

在线服务

- 办事指南 >>
- 表格下载
- 网上咨询
- 状态查询

审批结果 >>

建设项目选址意见书

建设用地规划许可证

建设工程规划许可证

规划核验(验收)

建筑名称核准

建设单位办理项目

在线申报 >>



在线服务 Online Services

建设用地规划许可证

共有 9377 个用地许可证 共 626 页

项目搜索

搜索

证书编号	建设单位	项目名称	建设位置	核发日期	查看
2014规(怀)地字0011号	北京东方美都农业科技有限公司	北京东方美都农业科.....	怀柔区北房镇黄吉营村32号	2014-06-12	查看
2014规(密)地字0012号	北京古北水镇房地产开发有限公司	密云县古北水镇国际.....	密云县古北口司马台	2014-06-06	查看
2014规(大)地字0025号	北京鸿坤伟业房地产开发有限公司	西红门0801-0.....	大兴区西红门镇	2014-06-06	查看
2014规(房)地字0028号	北京京投阳光房地产开发有限公司	房山区长阳镇长阳西.....	房山区长阳镇篱笆房村	2014-06-05	查看
2014规(房)地字0027号	中石化催化剂(北京)有限公司	6000立方米/年.....	房山区房山区城关街道办事处前、后朱各庄村	2014-06-05	查看
2014规(房)地字0026号	中国核电工程有限公司	中国核电工程有限公.....	房山区长阳镇水碾屯村	2014-06-03	查看

在线服务 >> 审批结果 >> 建设用地

在线服务

- 办事指南 >>
- 表格下载
- 网上咨询
- 状态查询
- 审批结果 >>**

建设项目选址意见书

建设用地规划许可证

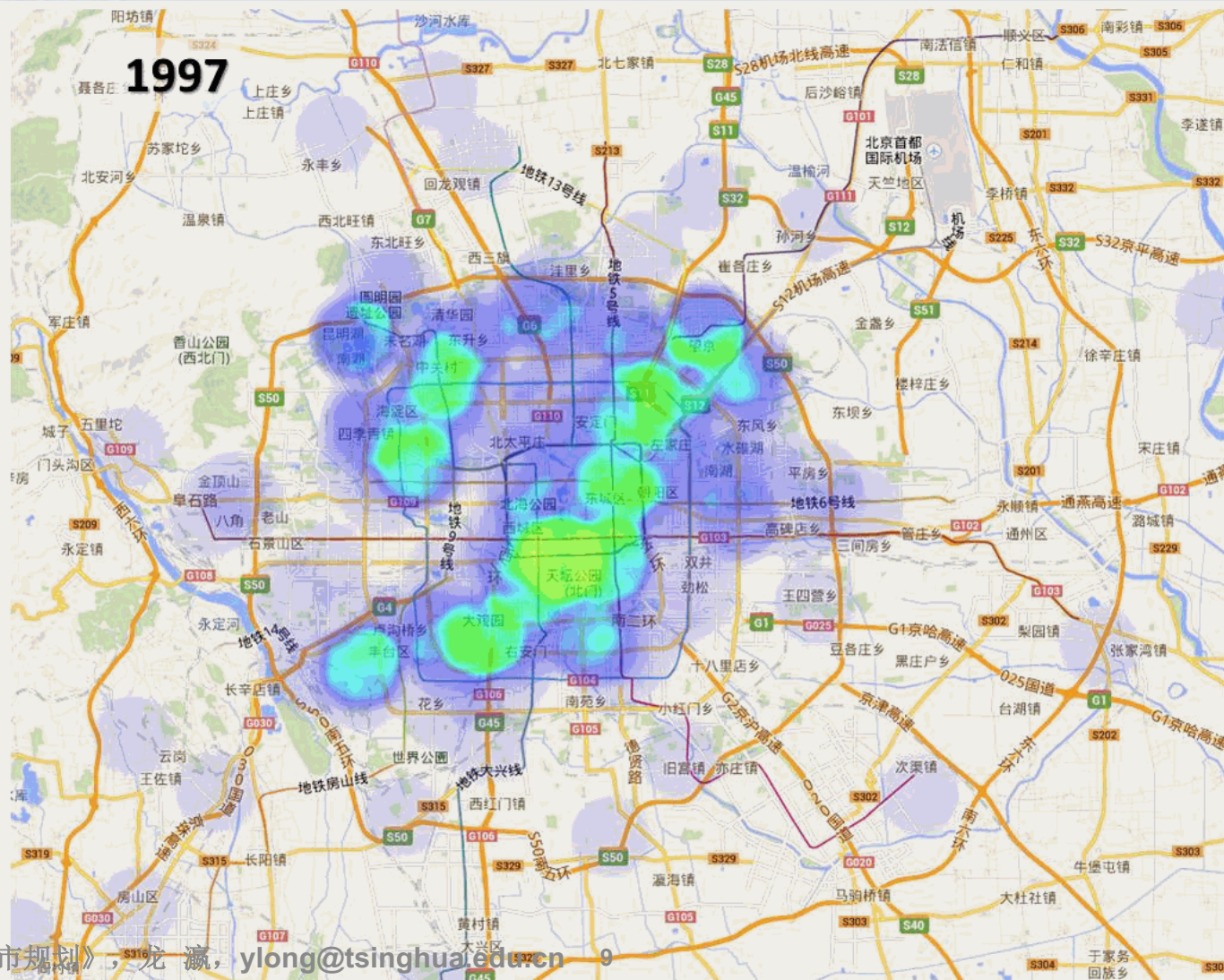
建设工程规划许可证

规划核验（验收）

建筑名称核准

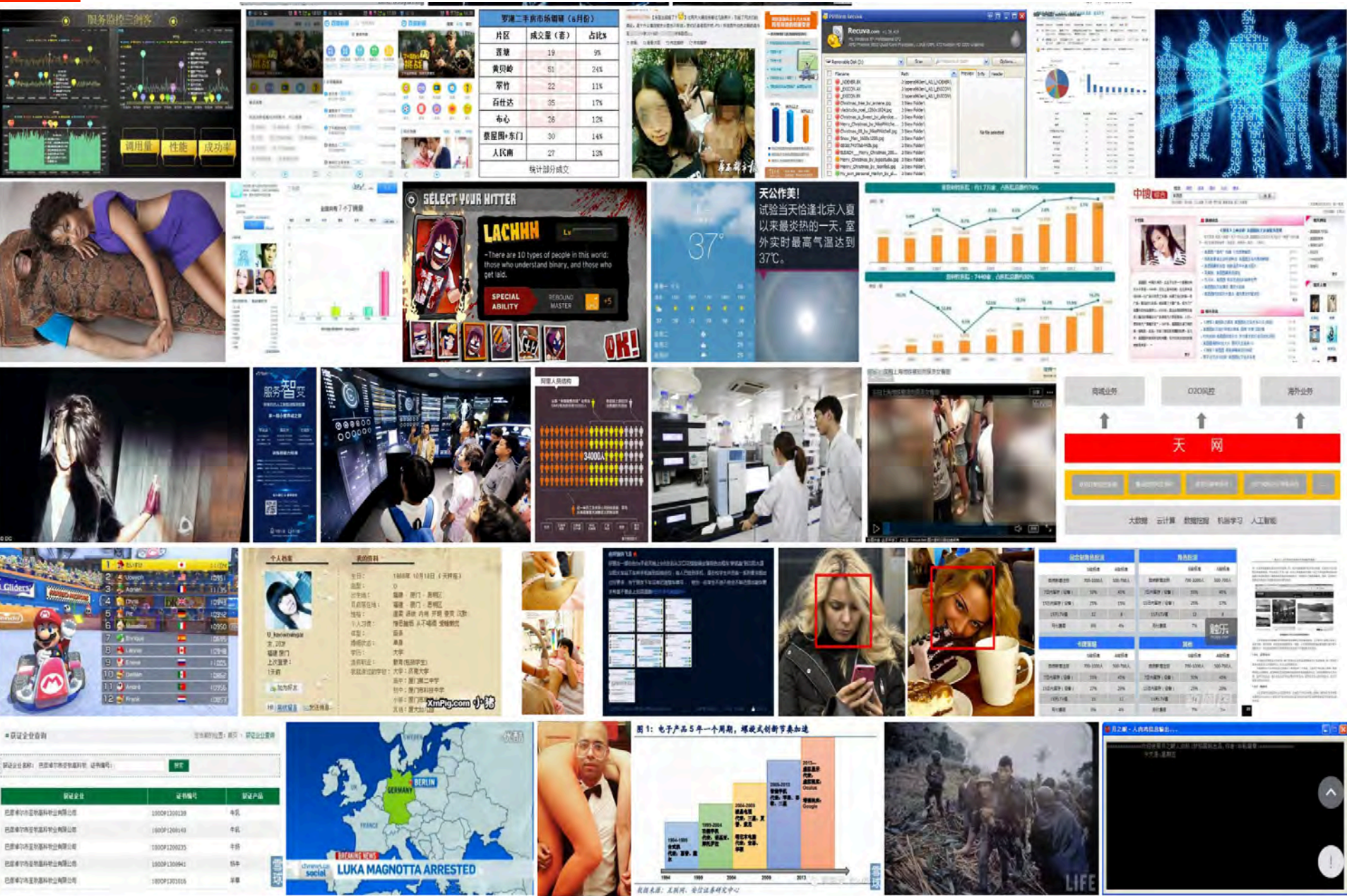
建设单位办理项目

在线申报 >>



城市空间大数据如何获得？

人肉大数据




购买服务 taobao.com

淘宝网 Taobao.com

宝贝 网络数据批量采集 搜索 在结果中排除 请输入要排除的词 确定

包邮 赠送退货运费险 货到付款 海外商品 二手 天猫 正品保障 更多



Python编程从入门到实践

官方正版


赠 视频教程

¥69.80 包邮 95人付款

正版 Python编程从入门到实践 python基础教程书籍python网络爬虫

拓创图书专营店 山东 青岛

掌柜热卖



Python编程从入门到实践

官方正版


赠 视频教程

¥69.80 包邮 0人付款

正版现货 Python编程从入门到实践 python基础教程书籍python网络爬虫

天都图书专营店 北京

掌柜热卖




网站网络数据批量采集、抓取到 EXCEL、批量提取, 定制网络爬虫

¥10.00 17人付款

网站网络数据批量采集、抓取到 EXCEL、批量提取, 定制网络爬虫

haoshuaijiaozhu 上海



数据采集 股票/彩票 网页蜘蛛爬虫

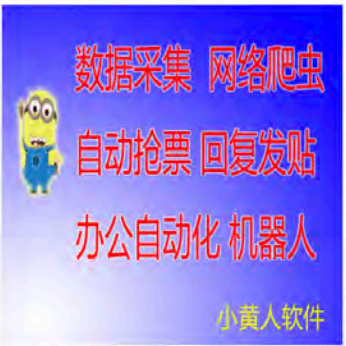
c#、nodejs、.net core跨平台技术

一代码仓库

¥100.00 1人付款

手机app网页数据采集软件定制网络信息蜘蛛爬虫定做批量抓取代做

yunjie721521 江苏 无锡




数据采集 网络爬虫 自动抢票 回复发帖 办公自动化 机器人

小黄人软件

¥20.00 0人付款

批量发帖数据采集提取网站data网络爬虫办公自动化源码软件定制

chenhao0568 北京




专注网页数据采集 指定网站及字段采集数据 多年爬虫经验 按您的格式保存数据 我们是敬业的爬虫

¥20.00 0人付款

python爬虫 网站网页数据网络采集收集文本图片批量抓取提取服务

zhangqijun18 浙江 杭州




大数据CRM深度挖掘系统 数据时代 网络数据采集/抓取/挖掘 THE ERA OF BIG DATA 13581523170(同微信)

¥20.00 包邮 0人付款

网站网络数据批量采集、抓取到 EXCEL、批量提取, 定制网络爬虫

张贵阳 北京



数据采集 网站抓取 采集数据

¥5.00 2人付款

网站网络数据批量采集、抓取到 EXCEL、批量提取, 定制网络爬虫

怀怀伊人 浙江 杭州

急速 省心 保障

1小时内推荐3-5家服务商

网页数据采集
登陆采集
定时采集
提交数据

¥ 100/起 成交数量: 58

数据采集/数据抓取/数据分析/网站采集/网页采集/excel

猪九戒 通2

编程屋

联系我

网站数据采集
一键导入Excel
专业/稳定/开源
登录/查询/筛选
自动更新 高效便捷

¥ 20/起 成交数量: 6

Excel/VBA网抓/采集 网页数据采集/网站采集抓取提取

猪六戒

99网络服务

联系我

数据采集
各行业网站数据采集、分析
资深程序员
为您服务

¥ 500/个 成交数量: 0

各类管理软件数据采集处理软件的定制开发及各种数据库运用开发

猪十戒

速诚软件

联系我

数据采集

¥ 500/起 成交数量: 26

网页数据采集/网站数据抓取/数据采集/数据分析/淘宝数据服务

猪八戒 通1

乾龙科技

联系我

网页数据采集
登陆采集
定时采集
提交数据

¥ 100/起 成交数量: 58

数据采集/数据抓取/数据分析/网站采集/网页采集/excel

编程屋

数据采集
Excel一键提取
专业/高效/稳定/开源
登录/筛选/上传/提交
采集工具定制开发

¥ 500/起 成交数量: 1

Excel/VBA网抓/采集 网页数据采集/软件工具定制

猪六戒

99网络服务

联系我

数据采集
网站数据采集
企业商家
500元
实力派
销量
20000笔/年

¥ 100/套 成交数量: 5

必贝YO数据采集/数据抓取/采集数据/抓取数据/网站采集/网

猪七戒 通1

必贝YO云数据服务平台

联系我

数据采集

¥ 200/起 成交数量: 18

数据采集爬取_数据录入处理_数据收集清洗_文件/数据批量下载

猪十戒

湾风网络

联系我

品牌口碑 SUCCESS
品牌口碑推广
网站产品推广
造就网络大品牌

¥ 968/套 成交数量: 337

餐饮行业品牌推广口碑整合营销事件营销百度公司产品网络营销推广

猪十六戒 通2

成就推广

联系我

网站数据采集
一键导入Excel
专业/稳定/开源
登录/查询/筛选
自动更新 高效便捷

¥ 20/起 成交数量: 6

Excel/VBA网抓/采集 网页数据采集/网站采集抓取提取

全国矢量地图数据

2016年50多图层20多GB



¥50.00 包邮

6人付款

shp/arcgis全国矢量地图数据县市边界
电子地图poi数据2016

lwlucy

浙江 杭州



2017年250家国家风景名胜区位置数据
(xlsx + shp + kmz+jpg)

¥4.99 包邮

0人付款

2017/2015国家风景名胜区分布地图
位置数据poi kmz shp

codeychai

北京



2017年全国百强县市分布地图
(xlsx + shp + kmz+jpg)

¥1.99 包邮

0人付款

2017年全国百强县市分布专题地图地
理位置数据poi shp kmz

codeychai

北京

彩虹软件
RAINBOW

网页数据爬取

微博、微信、地图、新闻
地图 POI, APP 内容、电商数据

¥20.00

8人付款

数据爬虫 数据采集 网页数据 微博
facebook 地图POI 网页监控

doulihongbaichi

北京

百度搜索腾讯地图大众点评网美团
网POI商户商家数据采集爬取获取

¥20.00

2人付款

百度搜索腾讯地图大众点评网美团
网POI商户商家数据采集爬取获取

venior英

福建 厦门

2016 商家POI数据

名称 | 区划 | 地址 | 商圈 | 电话 |
坐标 | 评价 | 照片 | 美食类型 |
营业时间 | 内容等

os服务ios开发 | 三层地理信息平台 | 大数据云计算

诚信 创新 合作 共赢 奉献

QQ:919071615

¥100.00

0人付款

2016 商家POI数据 poi全字段商家数
据 大数据

ourcan

广东 深圳

免费网络资源

- 详见第三讲课件
- www.beijingscitylab.com

相关搜索: 自学成才搞笑图片 自学成才表情包 老教授自学成才 贺成才 自学ps 自学成才的画家牡丹 驾校除名 自学成才 成才 自学网 51自学 ps教程自学网 我要自学网视频教程 自学绘画 自学 我要自学网



[2017大数据培训哪家好?有了解的吗——已回答](#)

光环国际大数据培训/Hadoop技术行业领跑者!一线总监级讲师授课,30人小班教学,22天从入门到精通!

不高薪NO就业!现在报名立减2000元!学大数据优选光环国际!

www.aura.cn 2017-10 - V3 - 广告

[大数据哪里培训好,直达BAT,入学即签订就业保证协议](#)

大数据哪里培训好,AAA教育总监级名师面授,入学签订就业协议,不就业免费从学!大数据哪里培训好,

仅需120天轻松实现高薪就业,“先就业,后付款”,有实力,有保障!

data.mandongzhe.com 2017-10 - V3 - 广告

[北京大数据培训班,名师授课,轻松玩转大数据,做IT界..](#)

北京大数据培训班千锋大数据面授就业班,企业级项目实战,入学签订就业协议.北京大数据培训班<千

锋>大数据培训业内大咖360度全方位指导,零学费入学,就业后再分期付学费!

www.mobiletrain.org 2017-10 - V3 - 广告

[大数据培训是在一线城市好还是二三线城市好?](#)

答:据我所知现在大数据的培训现在只有一线城市有,因为这是一个新兴行业,只有一线城市的公司需要这种先进技术。在北京的话我认为光环大数据还是可以的,都是名师教学,小班面授,保障到学员的学习质量。

2017-06-02 回答者: 嚶瑟的年华 2个回答

[求助,最近去培训,大数据中心在哪里](#)

答:培训的话很多的啊,你去大数据中心干什么去埃现在网上的学习也是很多的埃柠檬学院大数据,就是线上的大数据学习的埃

2017-02-16 回答者: 一步之遥白羊 1个回答

[大数据培训出来好找工作吗?](#)

答:大数据的就业领域是很宽广的,不管是科技领域,还是食品产业,零售业等等,都是需要大数据人才进行大数据的处理,以提供更好的用户体验,以及优化库存,降低成本,预测需求。1.数据分析师
数据分析师是数据师的一种,指的是不同行业中,专门从...

2017-09-11 回答者: 上海西线教育科技有限公司 6个回答



大数据与城市规划

欢迎选课

大数据与城市规划概论
 大数据在城市规划中的应用的研究进展
 城市大数据的获取
 城市大数据的统计与分析
 城市大数据的可视化
 城市大数据挖掘与空间应用
 城市大数据控制与城市设计
 城市大数据与城市研究新方法

参与式规划中的大数据应用
 控制性详细规划中的大数据应用
 战略及总体规划中的大数据应用
 数据增强设计
 大数据与城市规划的交互
 大模型：新数据环境下的城市研究新方法
 学生作业终期汇报与点评



1948
 清华大学建筑学院
 SCHOOL OF ARCHITECTURE
 TSINGHUA UNIVERSITY

沙龙题目：

Data Crawling With Fun 零代码抓数据

龙瀛《大数据与城市规划》课外沙龙系列第二讲

沙龙嘉宾：

李栋 博士
 北京清华同衡规划设计研究院
 技术创新中心常务副主任

主持人：

龙瀛 博士



内容简介：

互联网是当前信息传播的主要途径之一，其中也蕴含了大量城市规划研究中可利用的数据资源。如何快速有效的获取到这些数据、开展针对性的研究和分析，对每一位非计算机背景同学来说都是一个不大不小的挑战。本次沙龙针对上述问题，以零代码为切入点，组合既有的软件工具，介绍了数据资源探查与定位、自动化抓取和清洗等常规方法，并辅以案例介绍和上机实践，帮助同学们尽快掌握互联网数据抓取的相关基本知识和技能。

课程信息

课程号：70000662
 授课教师：龙瀛，清华大学建筑学院
 授课时间：1-16周/周五/第一大节（8:00-9:35）
 授课地点：详见选课系统或联系龙瀛老师
 课时：32
 学分：2

考察方式

中期作业、出勤、大作业
 大作业3~4人一组
选课对象
 城乡规划 建筑学 风景园林 城市地理
 市交通 环境规划与管理 城市经济 公共管理 社会学 计算机等专业的清华大学 **研究生**
 届时将结合国际最新进展和中国城市现实进行授课。

简介：城市大数据（及开放数据）对城市物质和社会空间进行了深入的刻画，为客观认识城市系统并总结其发展规律提供了重要机遇，也是城市规划和研究的重要支撑。本课程将结合中国城市（规划）以及技术发展特点进行讲授，秉承技术方法与城市研究与规划并重的原则，既侧重大数据技术方法的讲解，又重视城市系统和规划设计领域的应用。主要涵盖数据获取、统计、分析、可视化，城市系统分析，各个规划类型的应用，以及最新前沿介绍等内容。



关于龙瀛



已有在线课程



更多信息



二、北京旧城数据介绍

含典型城市大数据介绍

利用新兴数据研究古老的城市

所共享的图层简要说明（完整的数据说明准备中）

• 开发

- 遥感影像：2005、2010、2015image

• 形态

- 分等级路网及交叉口：Road2009、Road2014、Road2016、Junction2009、Junction2014
- 建筑物：Buildings2017
- 土地出让：LandTransaction2013
- 街道形态景观：Streets_Harvey
- 街道可步行性：Streets_WalkScore

• 功能

- 兴趣点（points of interest、POI）：POI2009、POI2011、POI20014
- 公司企业：Firms2010
- 社区尺度三经普：Jobs_economicsurvey3
- 写字楼：Office201609

• 活动

- 街道办事处人口：TownshipPOP
- 位置微博：Weibos2016
- 微博签到：Weibo_POI_Checkins1127.jp
- 大众点评：Dianping2014.jp
- 通勤联系：Commute2016.alpha

• 活力/品质

- 街景图片：SVPs目录
- 微软API深度学习后的街景图片属性：StreetViewPictures_point_MS_API
- 图像分割后的街景图片属性：StreetViewPictures_point_SegNet
- 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）
- 房价：HousingPrice

• 边界与分区

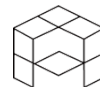
- 街道办事处：Jiedaos
- 居委会：Juweihui
- 利用2014年路网生成的地块：Parcels2011AICP
- 公交分析小区：TAZs
- 产业功能区：FunctionalZones
- 地理编码：GeoGrids

• 其他

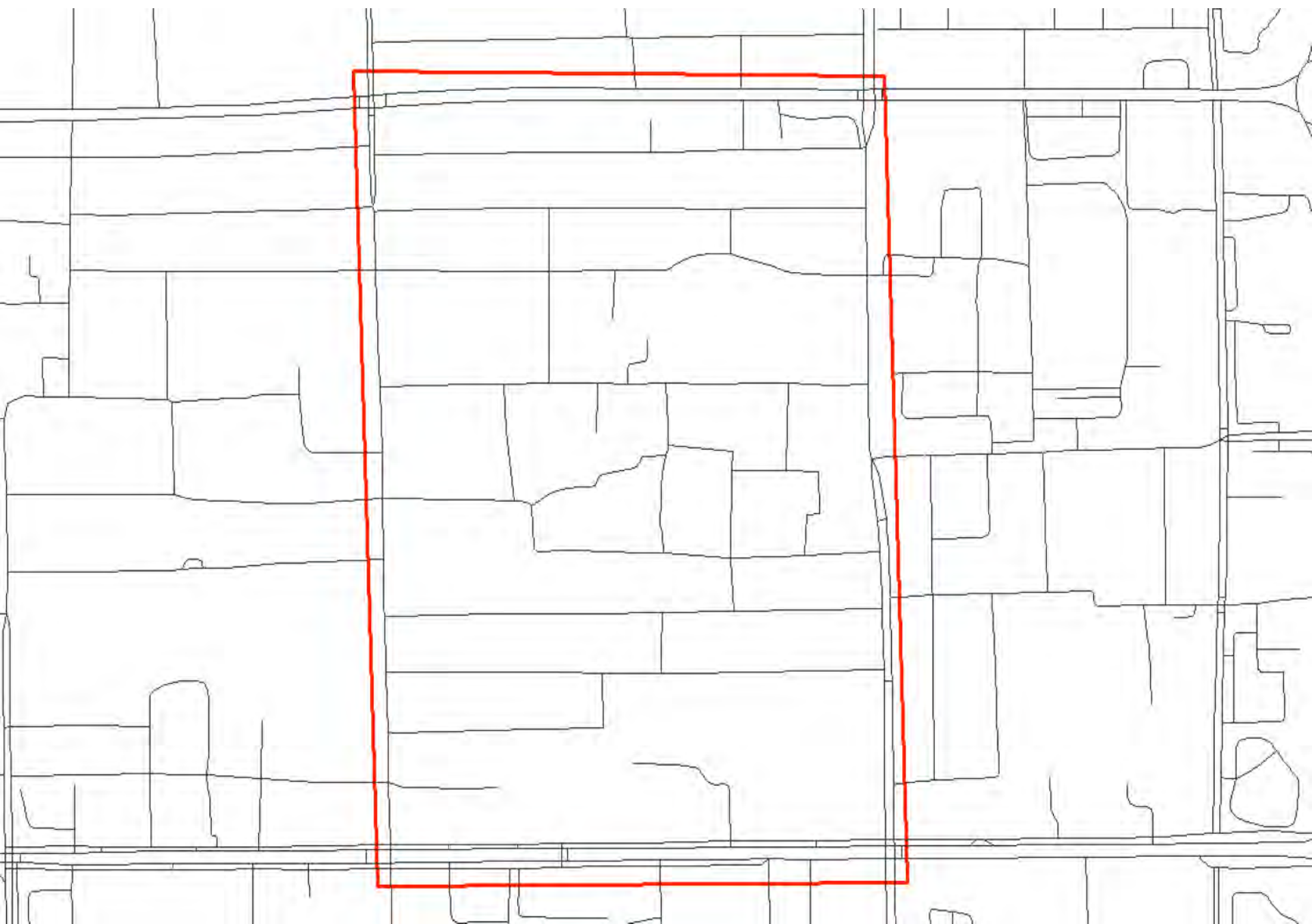
- 天安门广场所在地：Tiananmen
- 公交站点：BusStations201x
- 地铁站点：SubwayStations
- 居住小区：ResidentialCommunity

开发 → 形态 → 功能 → 活动 t

清华大学



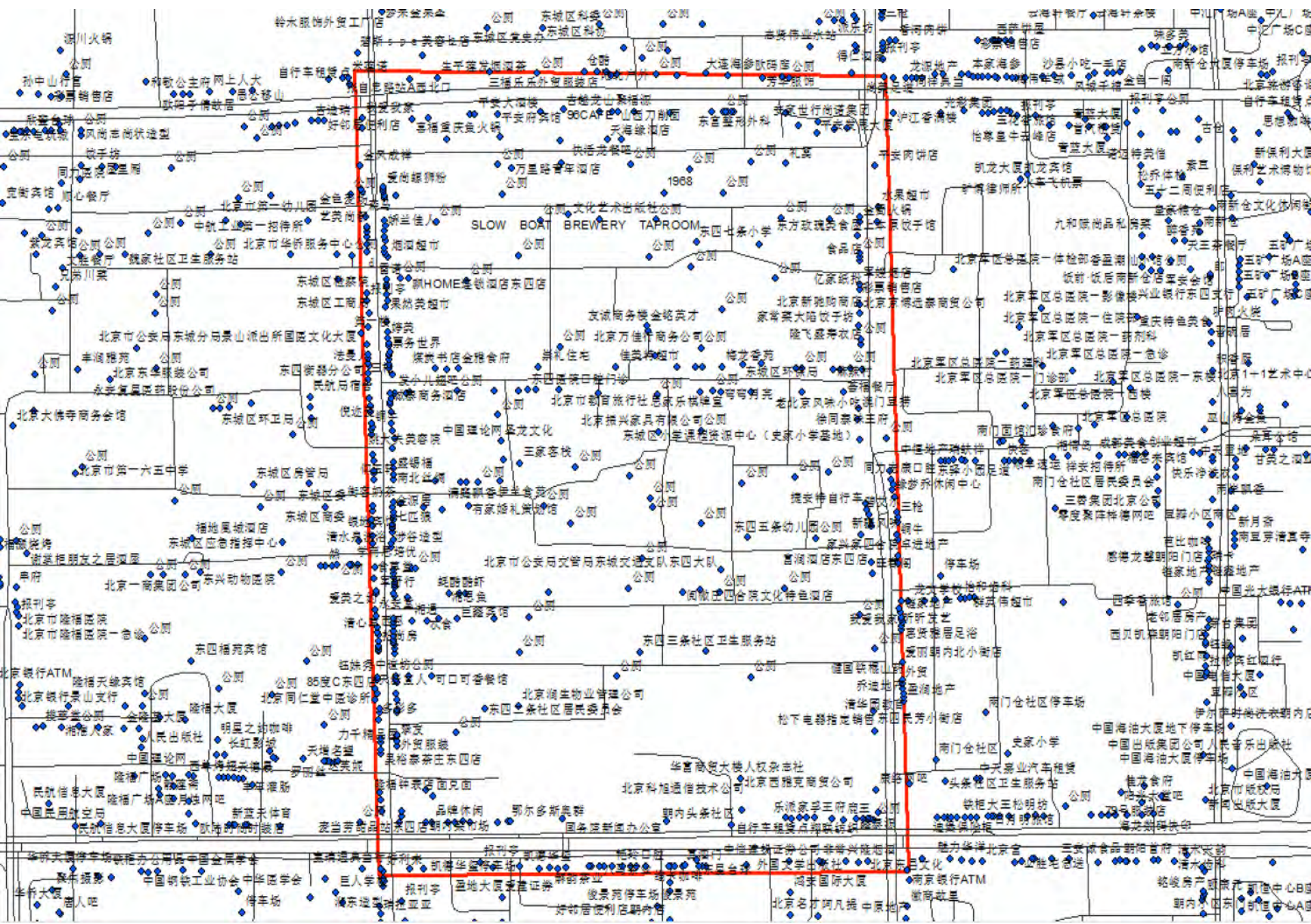
城市形态与设计 (道路网络)



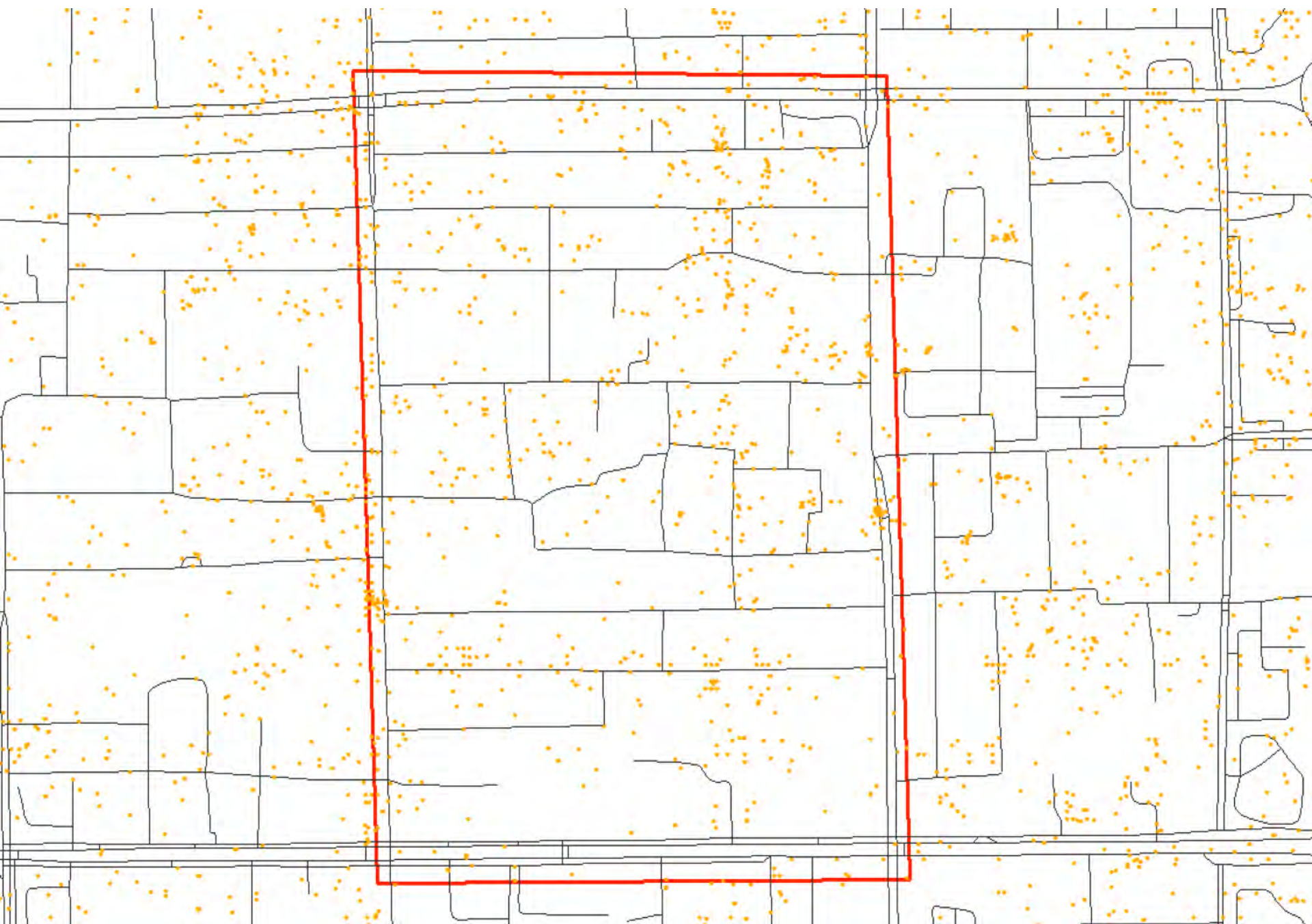
城市形态与设计 (建筑物)



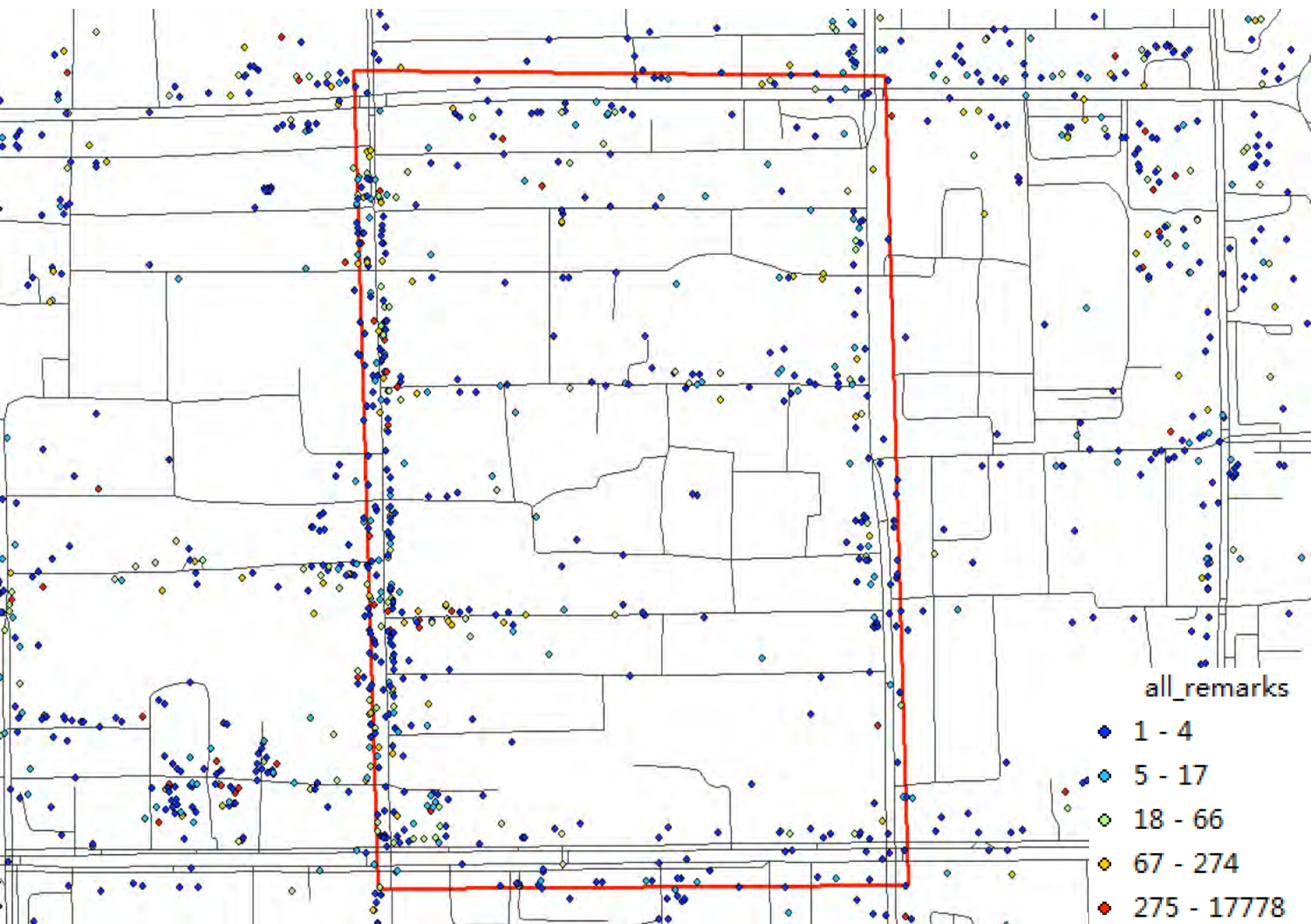
城市功能 (兴趣点)



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

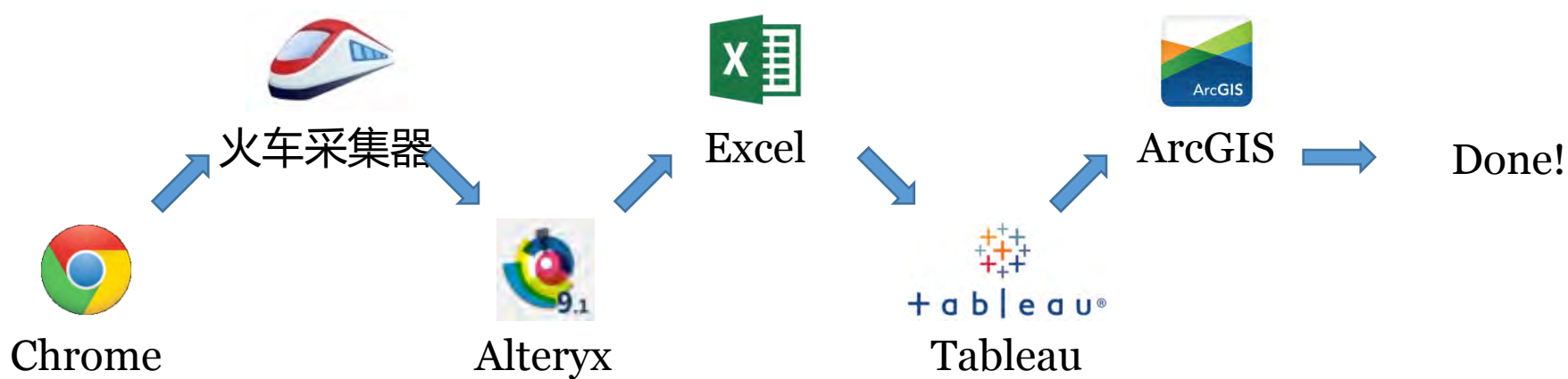
尺度/维度	区域/城市/片区/ 乡镇街道办事处	街区/地块	街区/地块内部	街道	街道内部
开发：遥感解译的土地利用、用地现状图（规划）、土地利用图（国土）	城镇用地面积、建设强度、生态安全格局、适宜开发土地 [城市扩张速度、城市扩张规模]	开发年代、是否适宜开发	肌理变化	角度变化	
形态：分等级路网、道路交叉口、建筑物、土地出让/规划许可、街景	基于道路交叉口的城乡判断、建筑面积、路网密度、交叉口密度、开放空间比例 [再开发比例、扩张比例]	尺度、紧凑度、基于建筑的城市形态类型、建筑密度、容积率、是否为开放空间、开放空间类型、可达性 [再开发与否、扩张与否]	是否有小路、建筑分布规律、是否有内部围墙 [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

感谢清华同衡李栋博士提供的资料支持

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
-

The screenshot shows a Chrome browser window displaying the Baidu homepage. The address bar shows the URL `https://www.baidu.com`. The page content includes the Baidu logo, a search bar, and a music player for 'Stay Here Forever' by Jewel. The Chrome DevTools interface is open at the bottom, with the 'Elements' panel selected, showing the HTML structure of the page. A context menu is open over the page content, and the 'Developer Tools' option is highlighted. The DevTools interface also shows the 'Network', 'Sources', 'Timeline', 'Profiles', 'Resources', 'Audits', and 'Console' panels.

数据的获取与处理

• 火车采集器 LocoySpider

The screenshot displays the LocoySpider software interface, version 8.6. The main window is titled "火车采集器 8.6" and contains several panes:

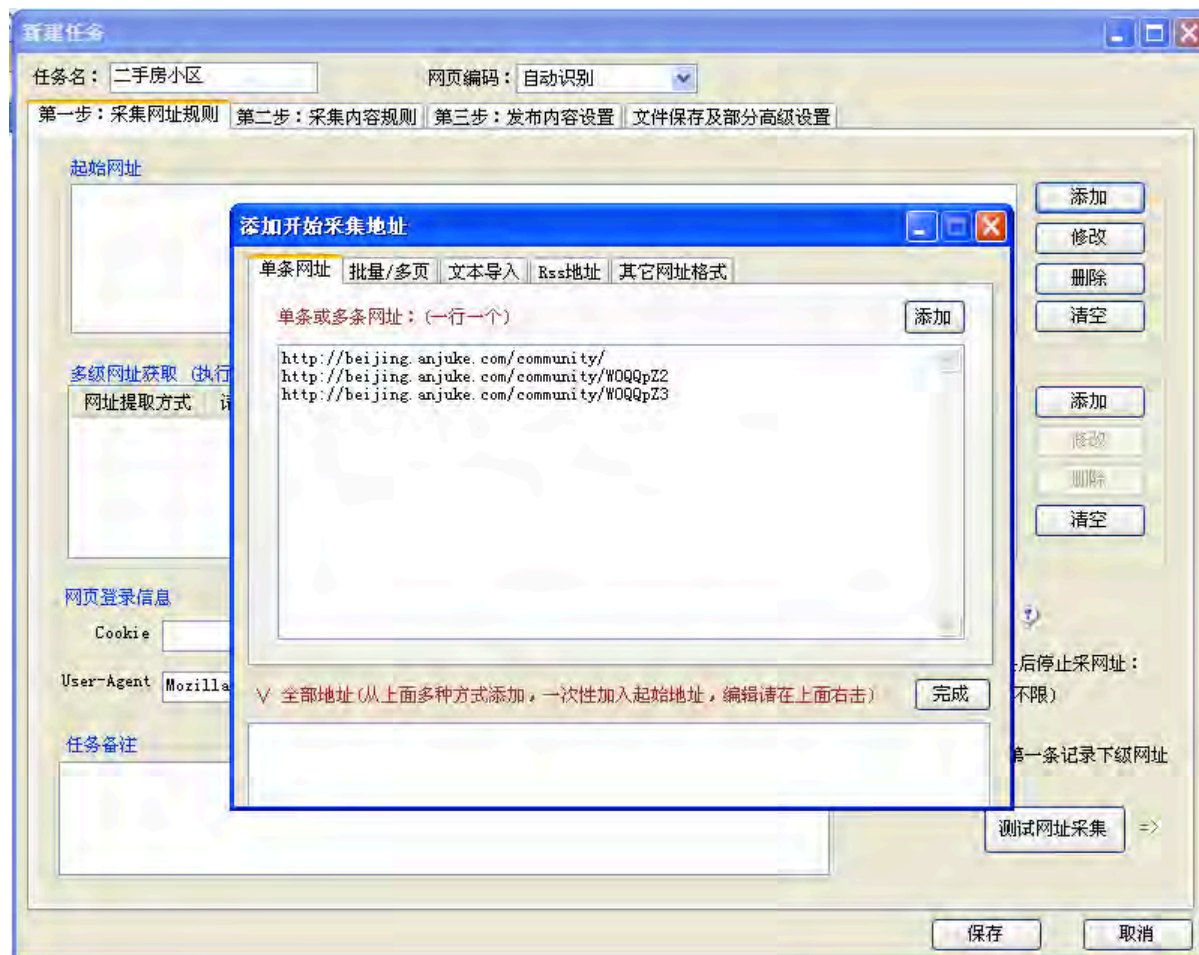
- 任务列表树 (Task List Tree):** A tree view on the left showing various task categories like "lbs weibo", "tango", "定时计划", "百度人气", "点评", "temp", "腾讯易出行 easygo", "20150823", "20150825", "深圳香港", "北京6r", "腾讯易出行 easygo_new", "baidu geo", and "跑步".
- 任务队列 (所有) (Task Queue):** A table listing tasks with columns for "任务名称 [ID]", "网址成...", and "内容成败重".
- 任务运行完成 (Task Execution Complete):** A log window showing detailed execution results for a task, including a red error message: "【totalnum】不符合不得为空设置, 被标记为未采, 记录Id2690, 记录网址 https://api.weibo.com/2/place/nearby timeline.json?sort=0&offset=0 @count=1&page=1&range=5000&starttime=1433520000&endtime=1434729600 @mapiid=CSJ R5k 8026&lat=32.6217780584&long=117.376895206 @access token=2.00x 2idFtidisC14925229adBKel6E&turn2=80".
- 文件列表 (File List):** A table at the bottom showing file details such as "文件", "大小", "完成", "进度", "剩余", "速度", "时间", "状态", "续传", "地址", and "序号".

The status bar at the bottom indicates: "任务:消息all ID:586 扩展:火车采集器" and "运行中:0,就绪:0,暂停:0,停止:1707". The user name "欢迎您: ldrgy01" is also visible in the bottom right corner.

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组二3 [705]	0/0/0	0/27
消息sq07_more组二2 [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	42/575E
数量复查2 [707]	6/0/0	251/554E
数量复查3 [708]	6/0/0	3/5797
数量复查4 [709]	6/0/0	2/579E
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

• 火车采集器

1. 添加网址



• 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

The screenshot shows a web browser displaying a real estate website. The address bar shows the URL: `beijing.anjuke.com/community/?from=navigation`. The page content includes a search bar, a list of properties, and a sidebar with advertisements. A developer tool is open at the bottom, showing the HTML structure of the page. The selected element is a list item with the following HTML code:

```
<li class="list_item over_bor">
  <a href="http://beijing.anjuke.com/community/view/55031" title="珠江帝景52759元/平米" alt="珠江帝景52759元/平米" target="_blank" class="img_content"></a>
  <div class="details">
    <div class="t_b"></div>
    <p>[朝阳 大望路]西大望路甲23号(广渠路28号)</p>
    <p>竣工日期: 2009年10月</p>
  </div>
  <p class="p_links">
    <a href="http://beijing.anjuke.com/map/sale/#11=39.8956069946269812=116.486221313477813=188?flag=1&commname=珠江帝景&commid=55031" target="_blank">查看地图</a>
    <span class="sep_t">|</span>
    <a href="http://beijing.anjuke.com/community/round/55031" target="_blank">生活配套</a>
    <span class="sep_t">|</span>
    <a href="http://beijing.anjuke.com/community/trends/55031" target="_blank">价格行情</a>
  </p>
</li>
```

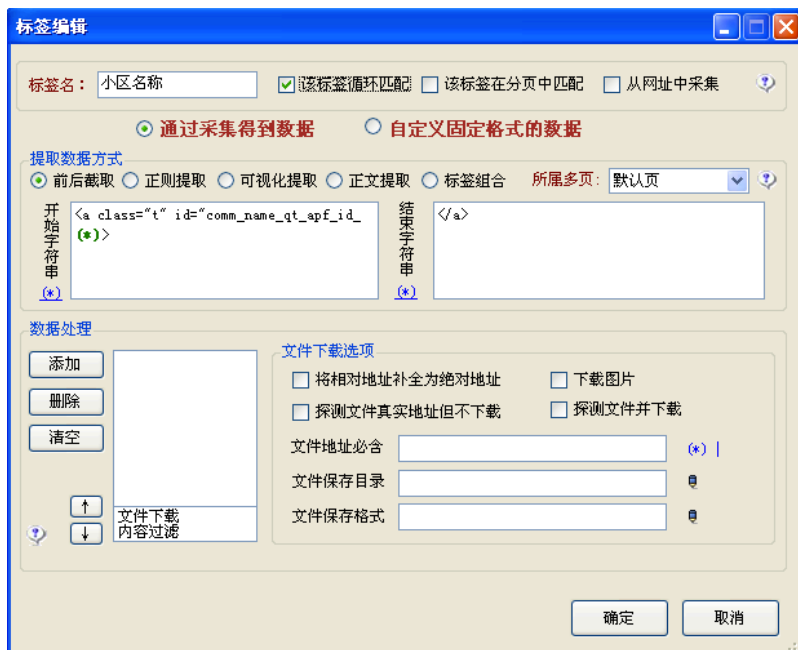
• 火车采集器

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>  
...
```



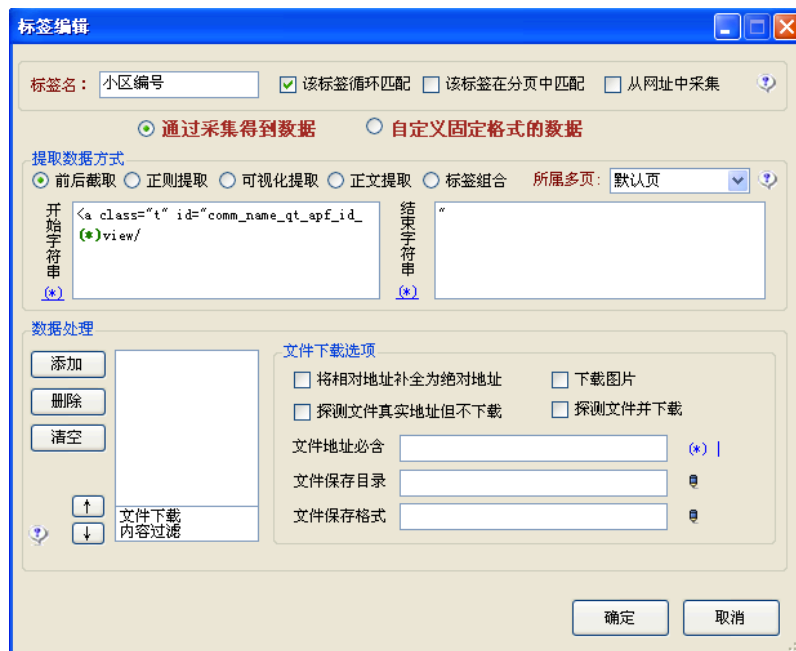
• 火车采集器

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>  
| ...
```



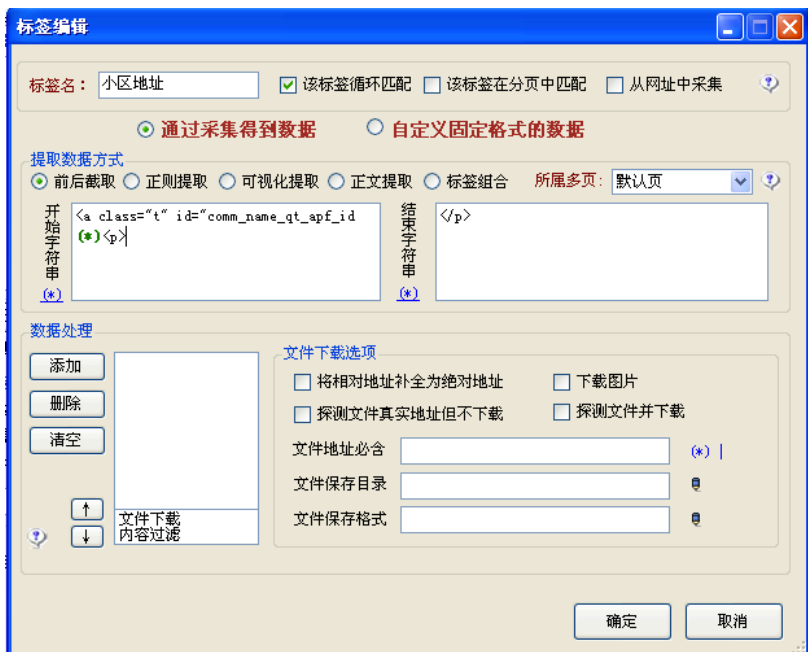
• 火车采集器

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>  
...
```



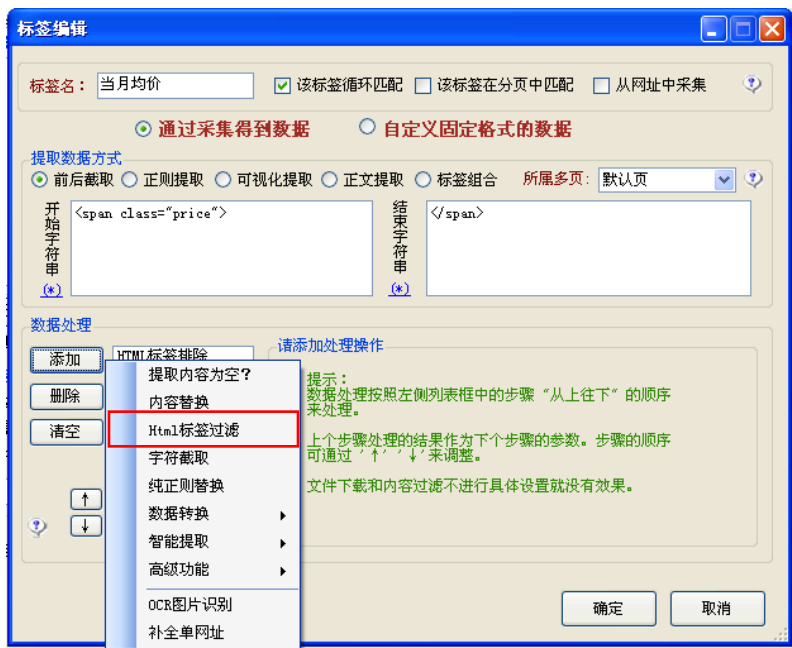
• 火车采集器

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>  
...
```



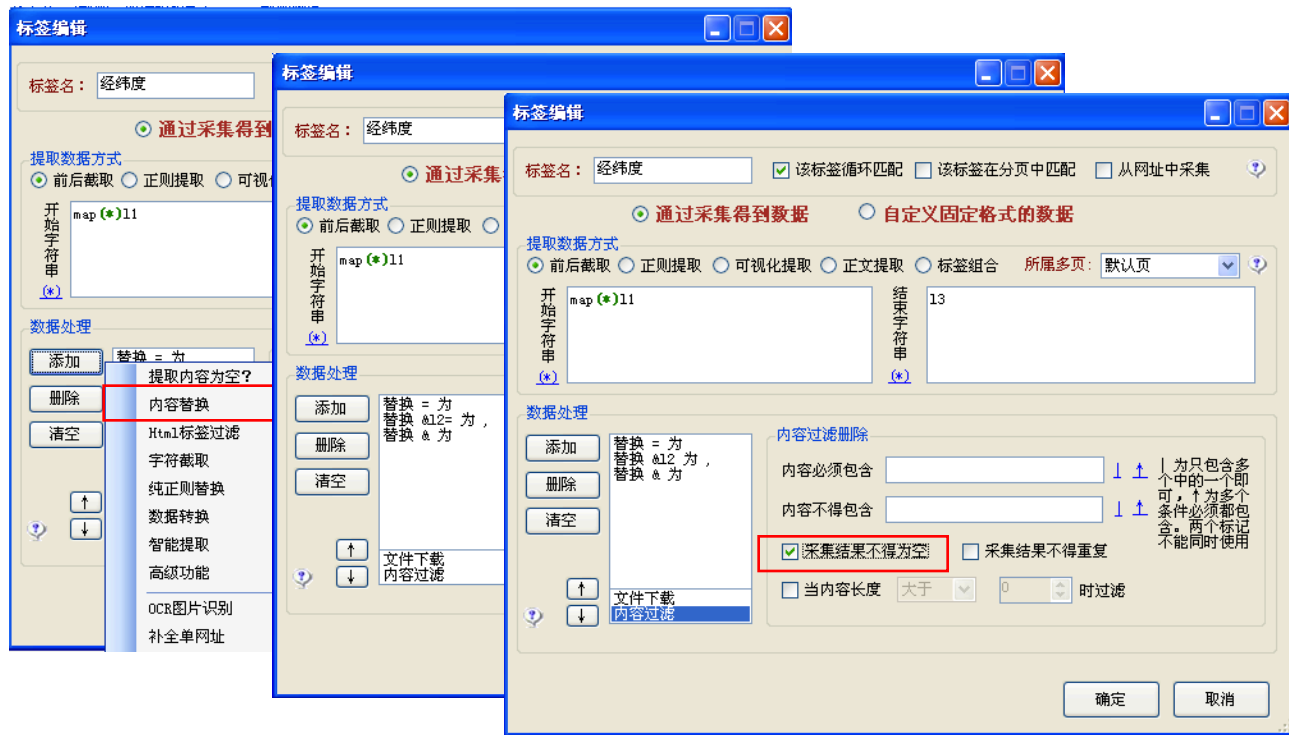
• 火车采集器

1. 添加网址

2. 明确抓取内容

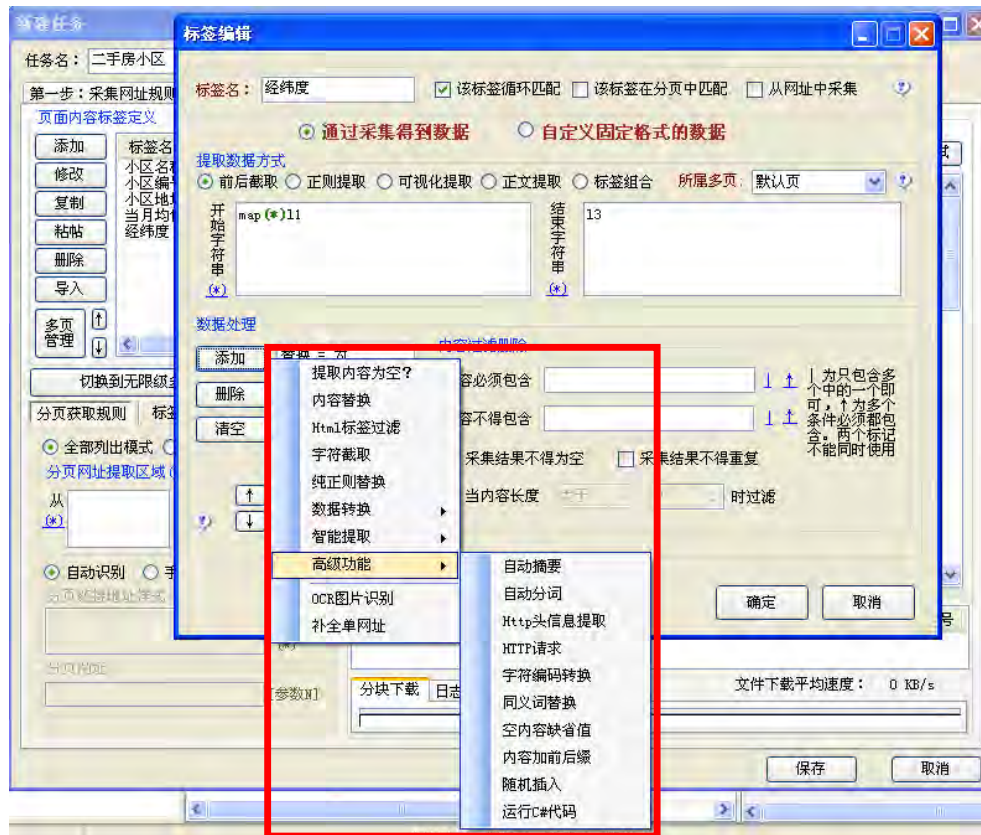
- 小区名称
- 唯一编号
- 小区地址
- 当月价格
- 经纬度坐标

```
...  
<a  
href="http://beijing.anjue.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>  
...
```



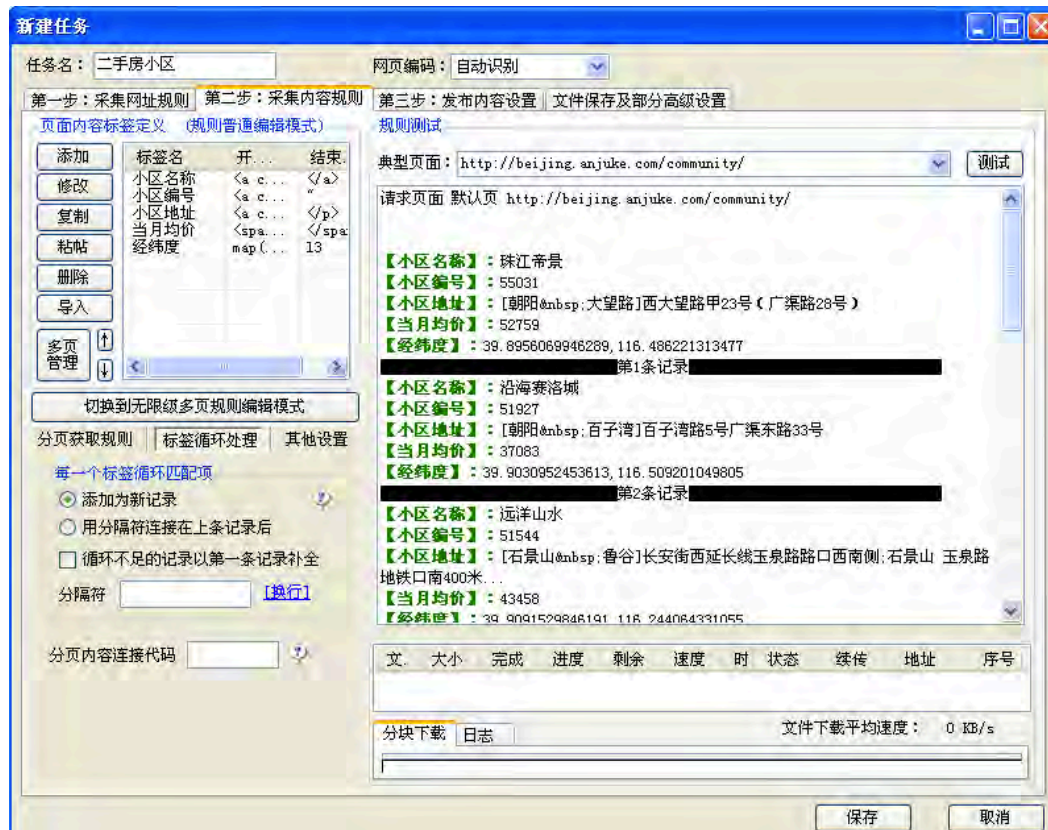
• 火车采集器

1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试



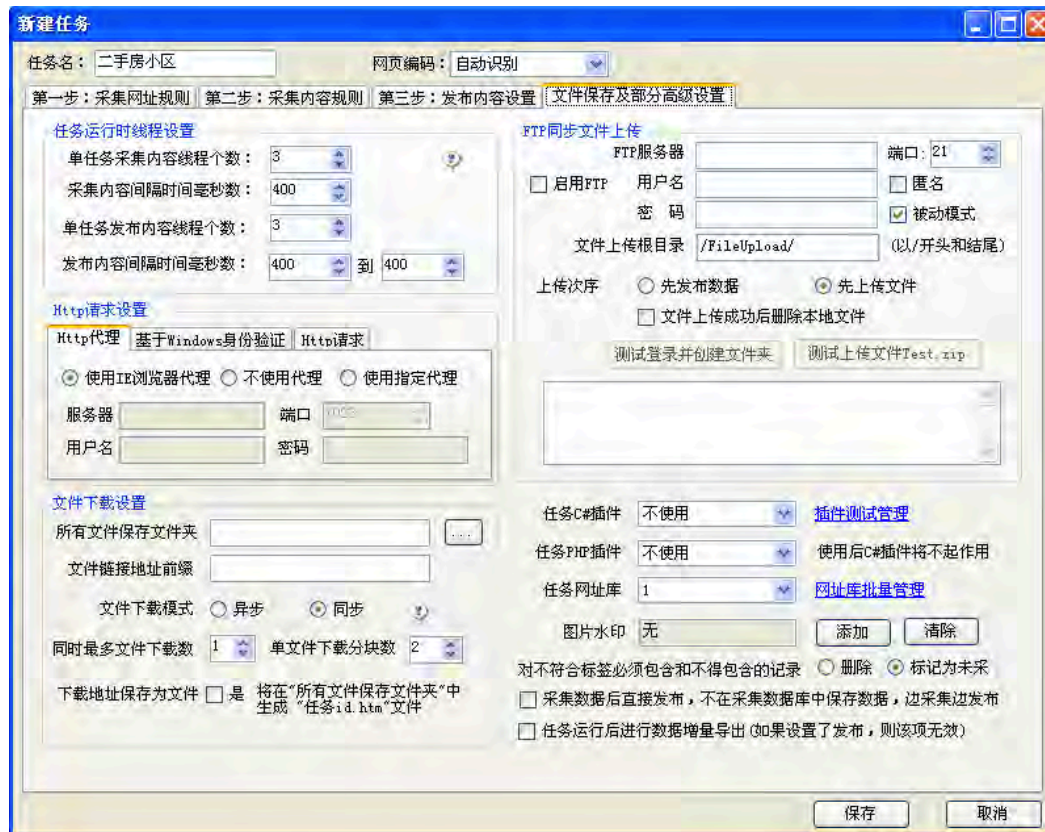
• 火车采集器

1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试



• 火车采集器

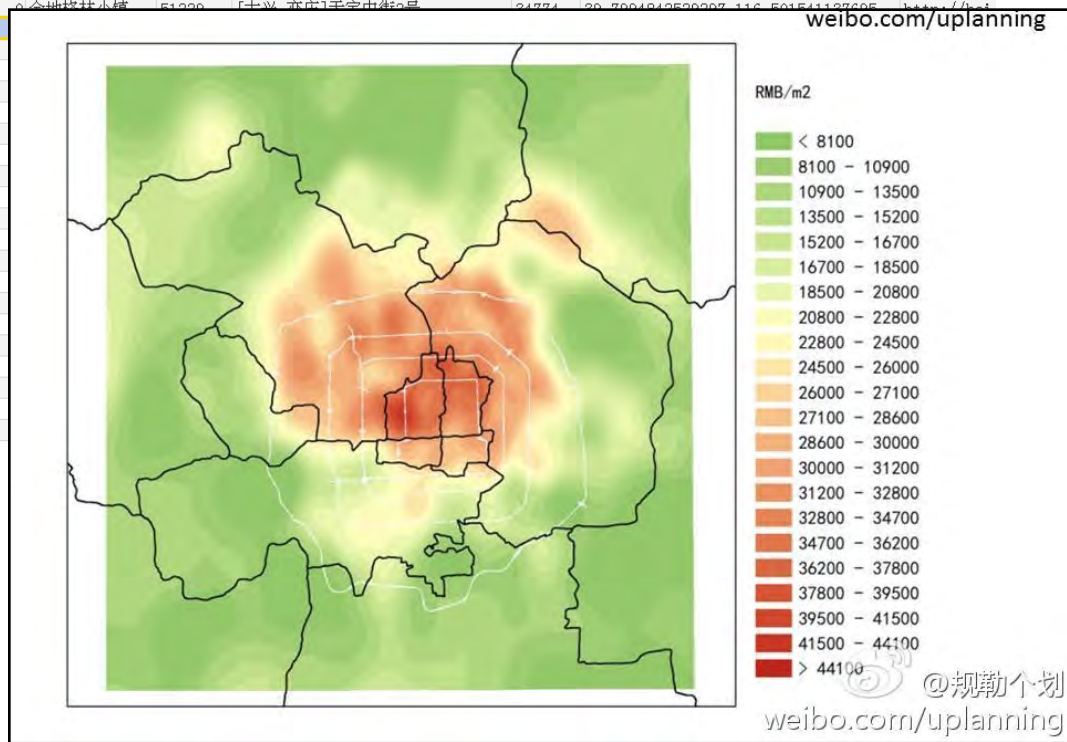
1. 添加网址
2. 明确抓取内容
3. 同步清洗/测试
4. 其他参数



• 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

Id	已发	已读	小区名称	小区编号	小区地址	当月共	经纬度	PageUrl
1	-1	0	珠江帝景	55031	[朝阳 大望路]西大望路甲23号(广渠	52759	39.8956069946289, 116.486221313477	http://bei.
2	-1	0	天通苑西三区	55926	[昌平 天通苑]地铁5号线天通苑西站	20253	40.0830383300781, 116.41576385498	http://bei.
3	-1	0	天洋城	210306	[周边 燕郊]迎宾北路	11200	40.328378041011, 116.63856730584	http://bei.
4	-1	0	天通苑北一区	171563	[昌平 天通苑]城铁5号线天通苑站	22978	40.0831336975098, 116.423866271973	http://bei.
5	-1	0	华腾园	55410	[朝阳 双井]东三环南路54号	42223	39.8931465148926, 116.471687316895	http://bei.
6	-1	0	金地格林小城	54290	[北兴 亦庄]手工业大街9号	24724	39.7994942599997, 116.501541127695	http://bei.
7	-1							
8	-1							
9	-1							
10	-1							
11	-1							
12	-1							
13	-1							
14	-1							
15	-1							
16	-1							
17	-1							
18	-1							
19	-1							
20	-1							
21	-1							
22	-1							
23	-1							
24	-1							
25	-1							
26	-1							
27	-1							



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](#)



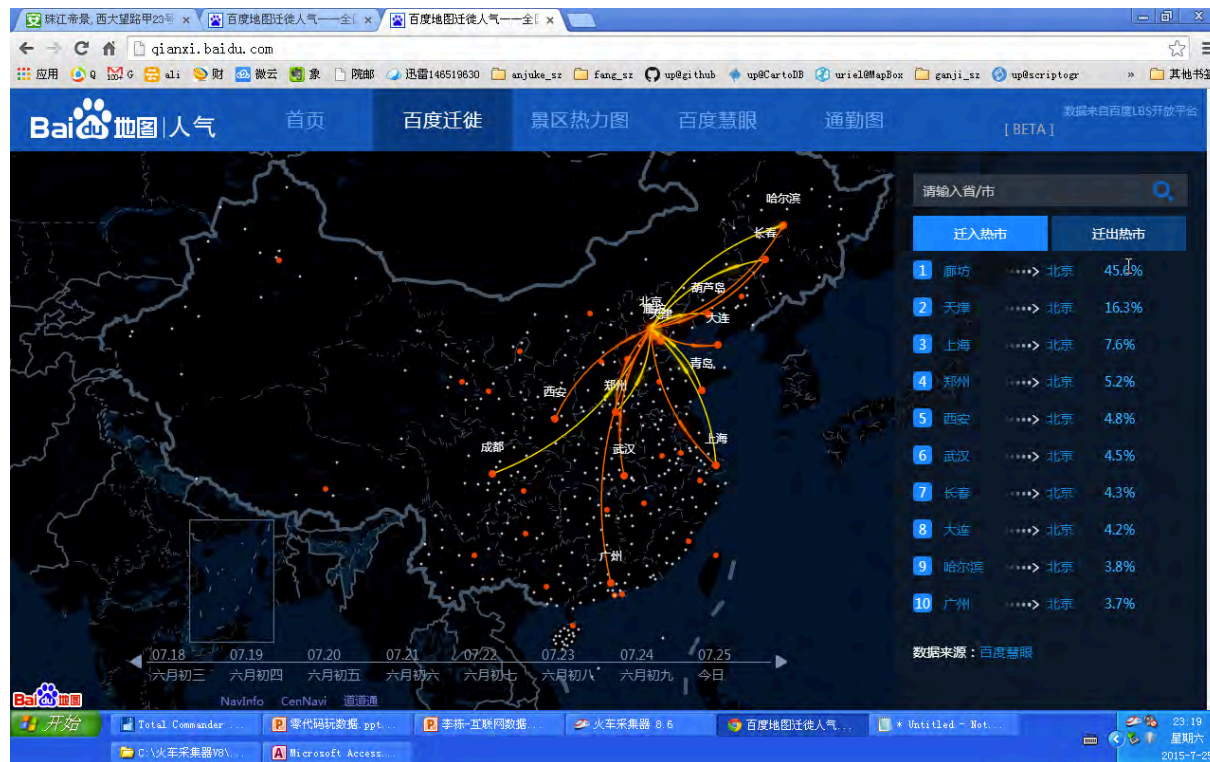
西山国际城
受欢迎度: 
本月均价: **24072元/m²** 比7月上涨1.03% [贷款](#)
[学区房](#) [环境优美](#) [金牌物业](#) [次新小区](#)
[生活便利](#) [投资推荐](#) [住户素质高](#) [安全性高](#)
二手房源: [1675套](#) 租房房源: [2189套](#)

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

The screenshots illustrate the process of capturing and analyzing data from a web application. The top screenshot shows the Baidu Map migration heatmap interface. The middle screenshot shows the 'Elements' panel with the 'city-migration.php/call...' path highlighted. The bottom screenshot shows the 'Network' panel with a JSON response from the 'city-migration.php/call...' endpoint.

```
JSONQuery17207564627565443516_1437837520068 && JSONQuery17207564627565443516_1437837520068  
{  
  "city_name": "\u56e8\u897d", "city_migrate_in_index": 456, "trend": -13921},  
  "city_name": "\u5929\u4e09", "city_migrate_in_index": 163, "trend": -5064},  
  "city_name": "\u4e0a\u6d77", "city_migrate_in_index": 76, "trend": -1604},  
  "city_name": "\u90fd\u5e02", "city_migrate_in_index": 53, "trend": -1280},  
  "city_name": "\u897f\u4e09", "city_migrate_in_index": 48, "trend": -1032},  
  "city_name": "\u6666\u56fd", "city_migrate_in_index": 45, "trend": -875},  
  "city_name": "\u897d\u56fd", "city_migrate_in_index": 43, "trend": 0}, {"city_name": "\u5927\u897d", "city_migrate_in_index": 42, "trend": 0},  
  "city_name": "\u56fd\u56fd", "city_migrate_in_index": 38, "trend": 0},  
  "city_name": "\u897d\u56fd", "city_migrate_in_index": 37, "trend": -929}}}
```


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
```



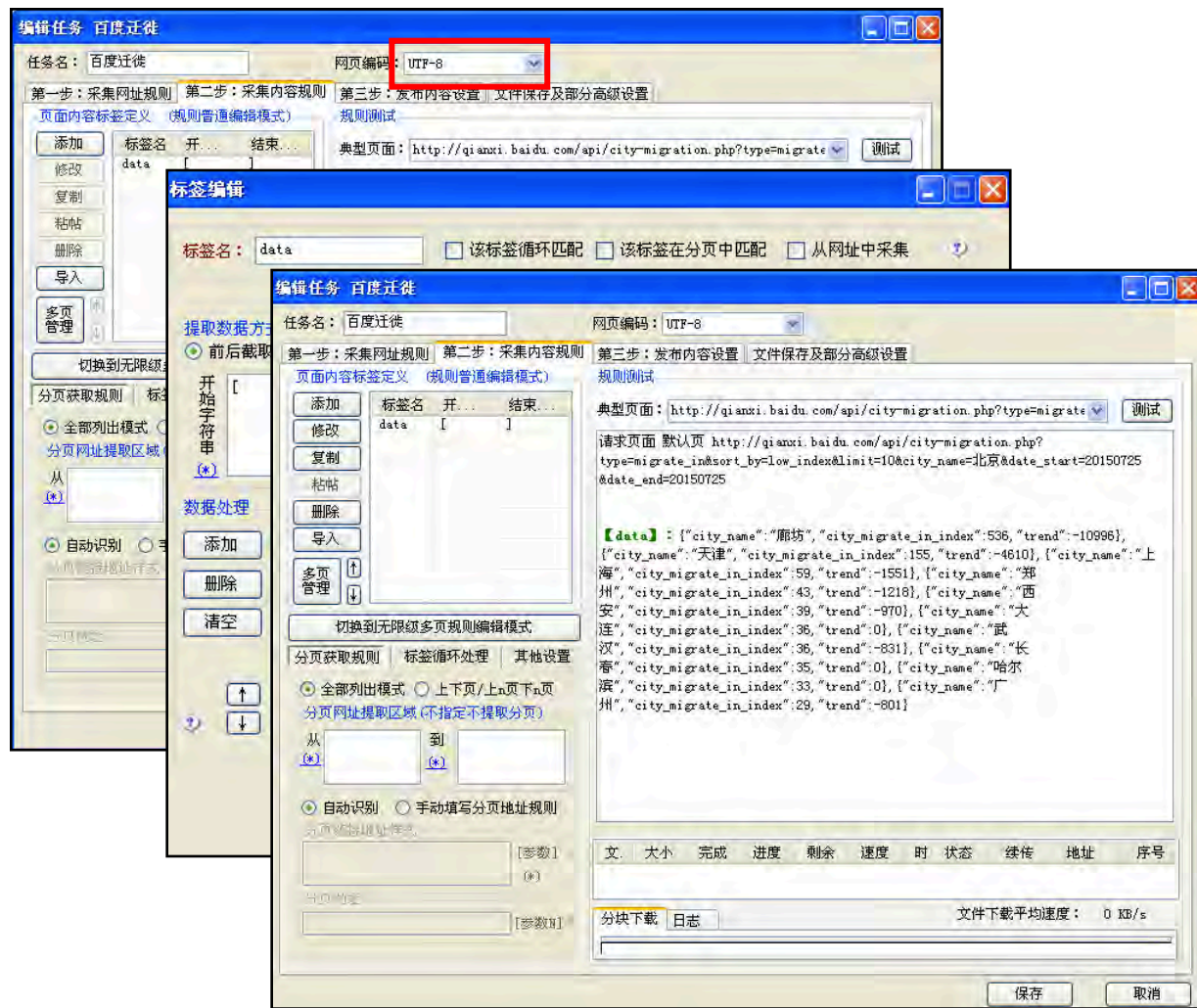
```
http://qianxi.baidu.com/api/city-  
migration.php?type=migrate_in&sort_by=low_index&  
limit=10&city_name=北京  
&date_start=20150725&date_end=20150725
```

```
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 displays the Baidu Migration Heatmap interface. The map shows migration paths from various cities to Beijing. A table on the right lists the top incoming cities 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%

Below the map, the Network DevTools console shows the raw JSON response from the API:

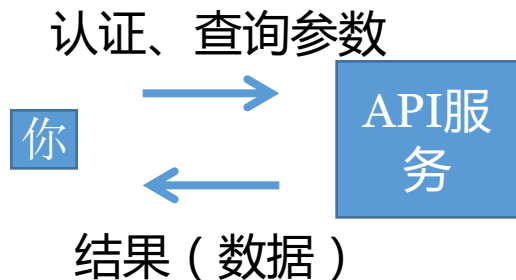
```
jQuery17207564627565443516_1437837520368 && jQuery17207564627565443516_1437837520368
([{"city_name": "\u5eca\u574a", "city_migrate_in_index": 456, "trend": -13521},
{"city_name": "\u5929\u6d25", "city_migrate_in_index": 163, "trend": -5064},
{"city_name": "\u4e0a\u6d77", "city_migrate_in_index": 76, "trend": -1604},
{"city_name": "\u90d\u55de", "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\u66ee", "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

感谢清华同衡李栋博士提供的资料支持

- 火车采集器

- API类型数据的获取



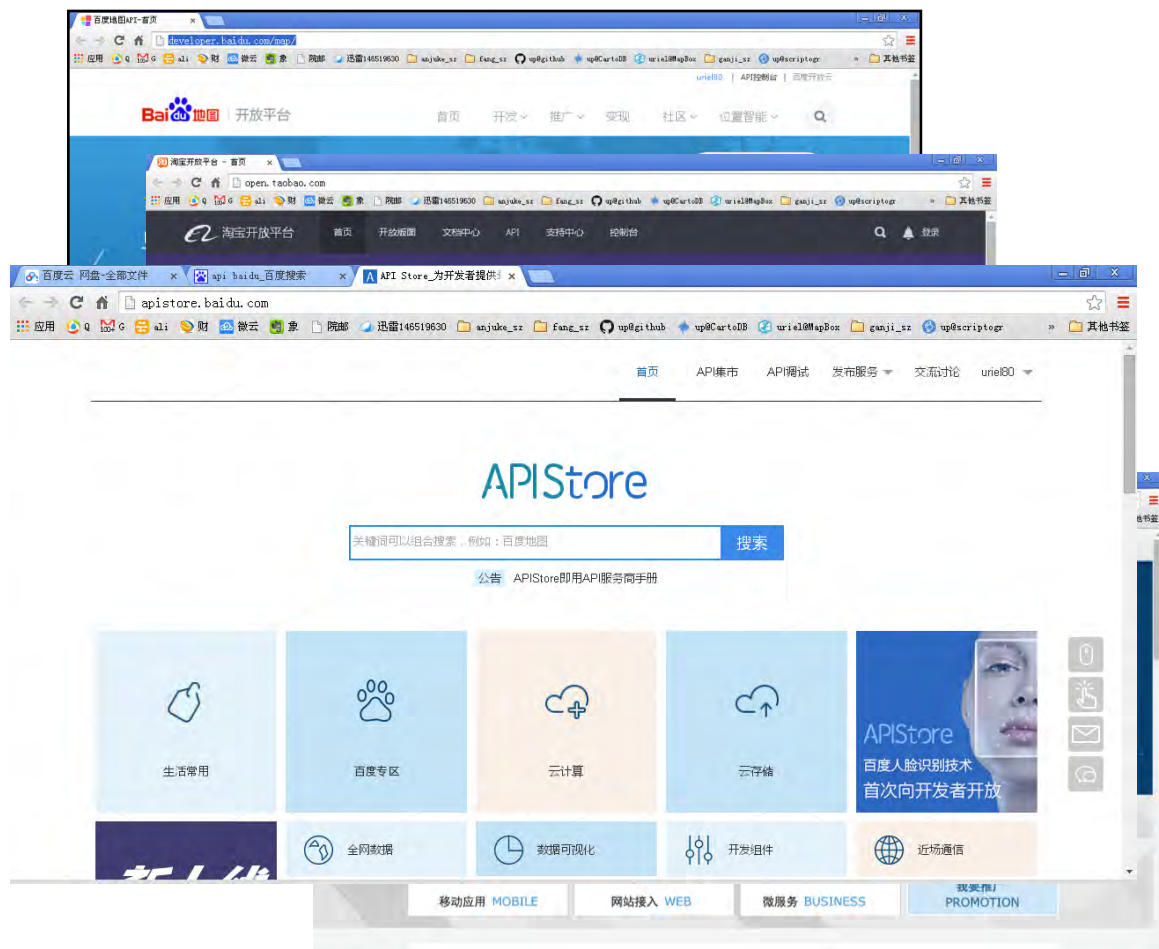
API (Application Programming Interface , 应用程序编程接口) 是一些预先定义的函数 , 目的是提供应用程序开发人员基于某软件或硬件得以访问一组程序的能力 , 而又无需访问源码 , 或理解程序内部工作机制的细节。

- 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
-

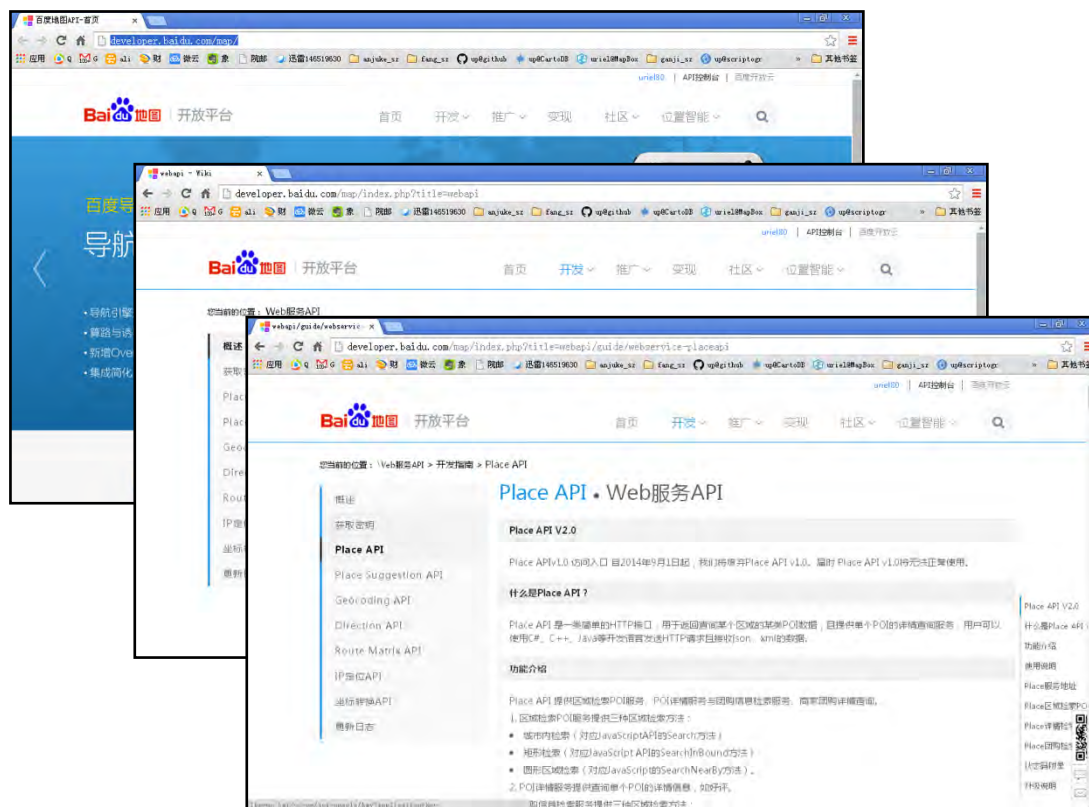


数据的获取与处理

• 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>

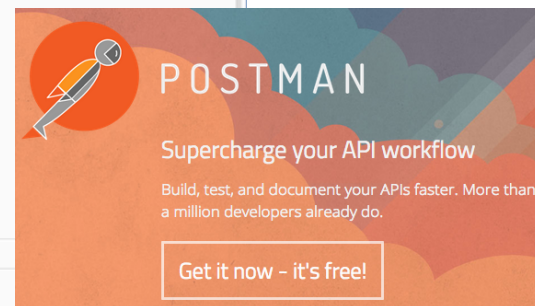
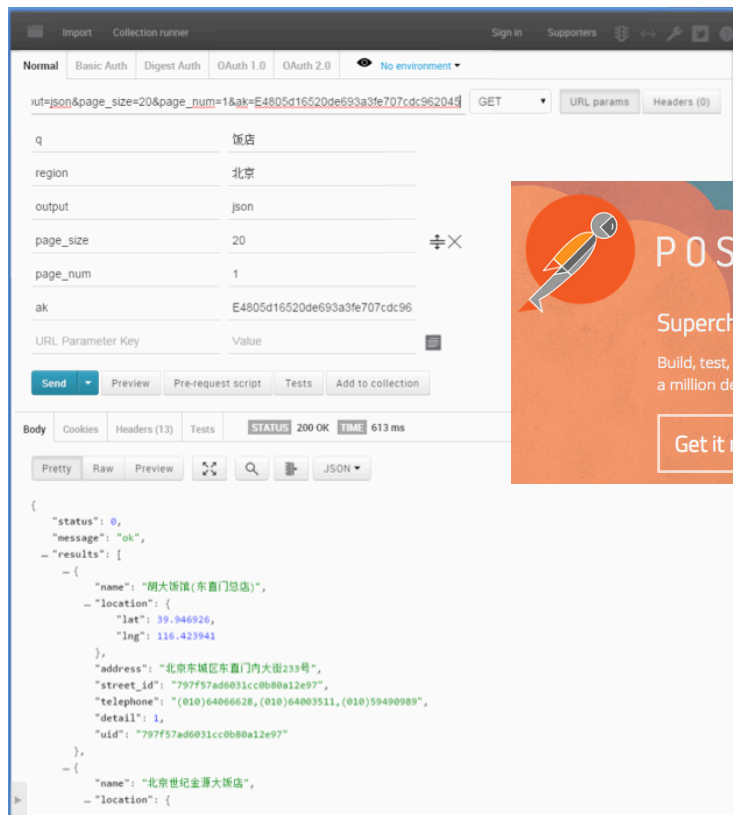


数据的获取与处理

- 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

• 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
```

结果字段（以结果集里的一条数据为例，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的唯一标识
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
```

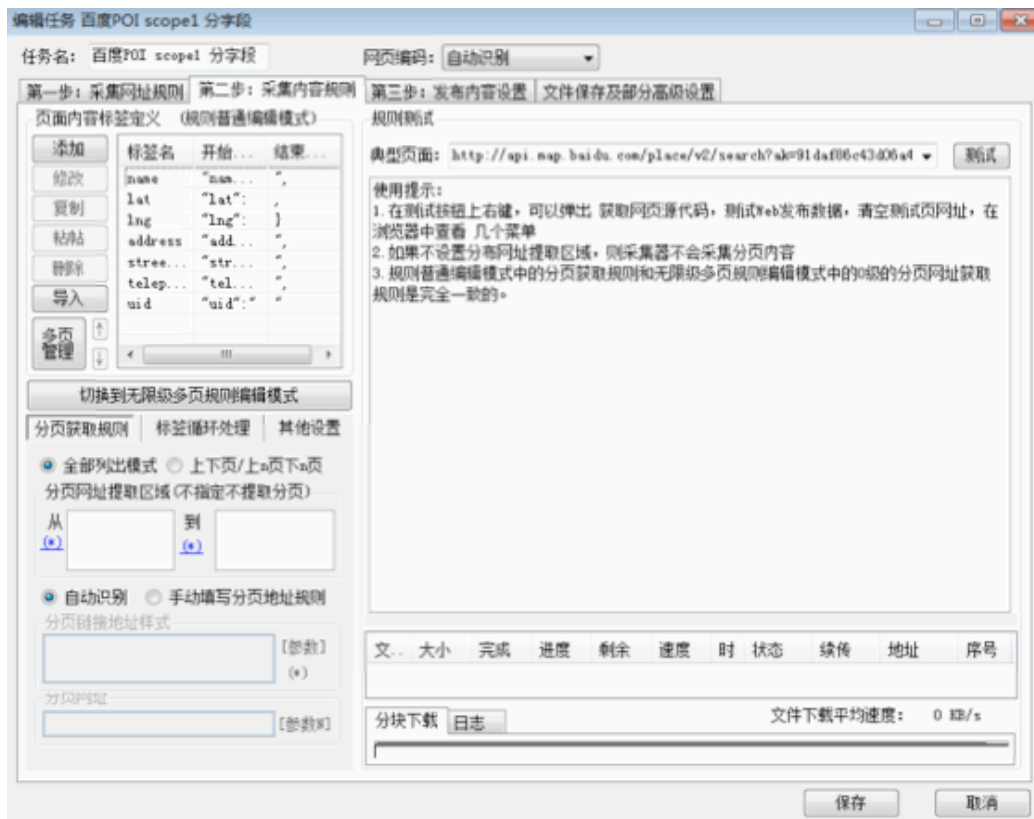
```
{  
  "status": 0,  
  "message": "ok",  
  "total": 15610,  
  "results": [  
    {  
      "name": "北京盘古七星酒店",  
      "location": {  
        "lat": 39.995074,  
        "lng": 116.393207  
      },  
      "address": "中国北京朝阳区北四环中路27号, 盘古大观",  
      "street_id": "925c550f80330fab5a3ac832",  
      "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
```

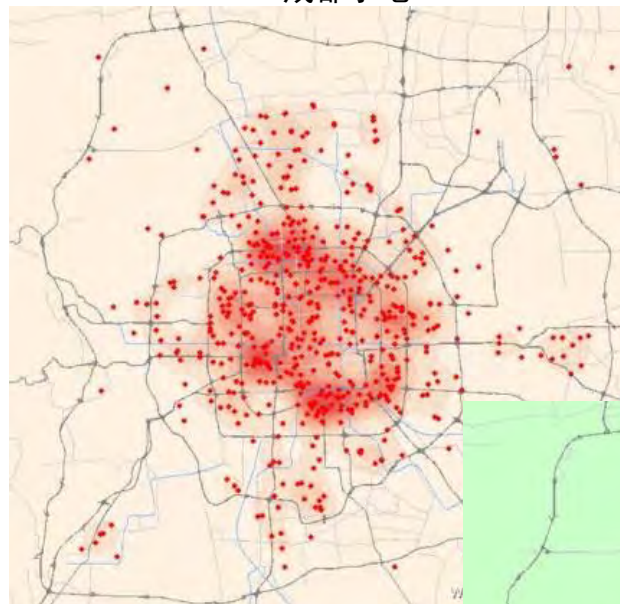


数据的获取与处理

- 火车采集器

- API类型数据的获取
- 百度地图API
 - Place API

成都小吃

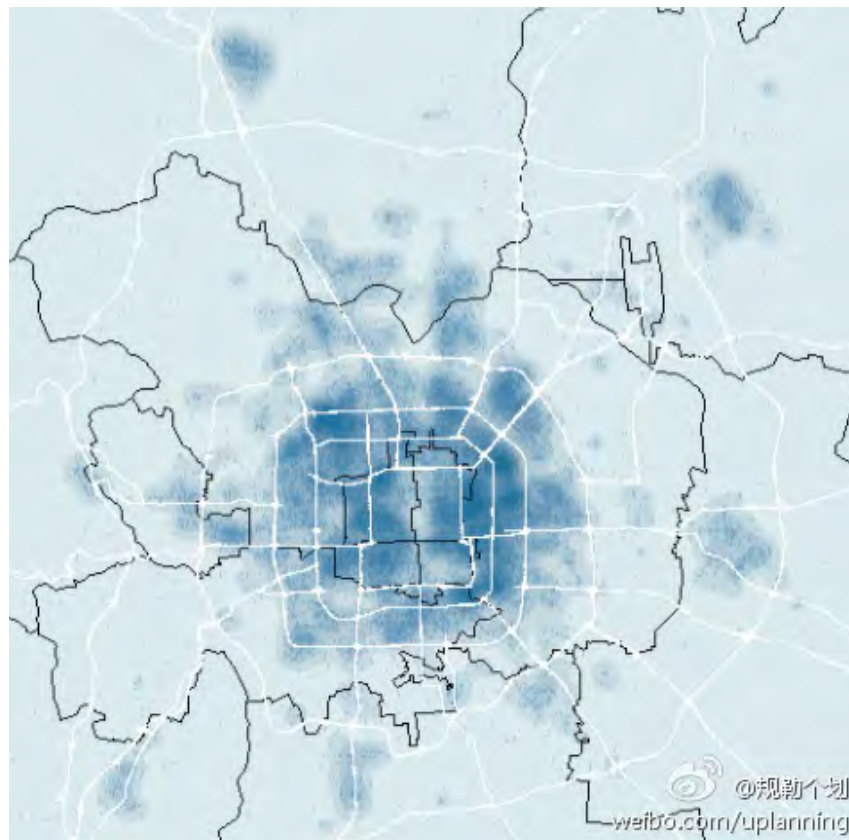


沙县小吃



- 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 Maps API documentation page for Place API V2.0. The page is in Chinese and provides information about the API, including its purpose, usage, and features.

Place API • Web服务API

Place API V2.0

Place API v1.0 访问入口 自2014年9月1日起,我们将废弃Place API v1.0,届时 Place API v1.0将无法正常使用。

什么是Place API ?

Place API 是一类简单的HTTP接口,用于返回查询某个区域的某类POI数据,且提供单个POI的详情查询服务。用户可以使用C#, C++, Java等开发语言发送HTTP请求且接收json、xml的数据。

功能介绍

Place API 提供区域检索POI服务, POI详情服务与团购信息检索服务。商家团购详情查询。

1. 区域检索POI服务提供三种区域检索方法:
 - 城市内检索 (对应JavaScript API的Search方法)
 - 矩形检索 (对应JavaScript API的SearchInBound方法)
 - 圆形区域检索 (对应JavaScript的SearchNearBy方法)。
2. POI详情服务提供查询单个POI的详情信息,如好评。

团购信息检索服务提供三种区域检索方法:

右侧边栏包含以下链接：[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](#), [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）



- 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 header includes the Weibo logo and navigation links: 微连接, 微服务, 文档, 支持, 推广, 我的应用. The main content area is titled '开发文档' (Developer Documentation) and features a search bar and a sidebar with navigation links: 首页, 平台公告, 平台概述, 政策规范, 视频直播, 头条文章, 微服务, 轻应用, 商业接口, 微博支付, 粉丝服务平台. The main content area is titled '微博API' and contains a grid of API interface buttons: 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, and OAuth 2.0授权接口. Below the grid, there are two explanatory paragraphs: '标识为 [高级权限图标] 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' and '标识为 [SCOPE图标] 的接口为需要单独向用户提出SCOPE授权请求的接口, 用户单独授权后才可以调用, 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)' with a table of API 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 [高级权限图标]
写入接口	发送消息	向用户回复私信消息接口 [SCOPE图标]

• 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

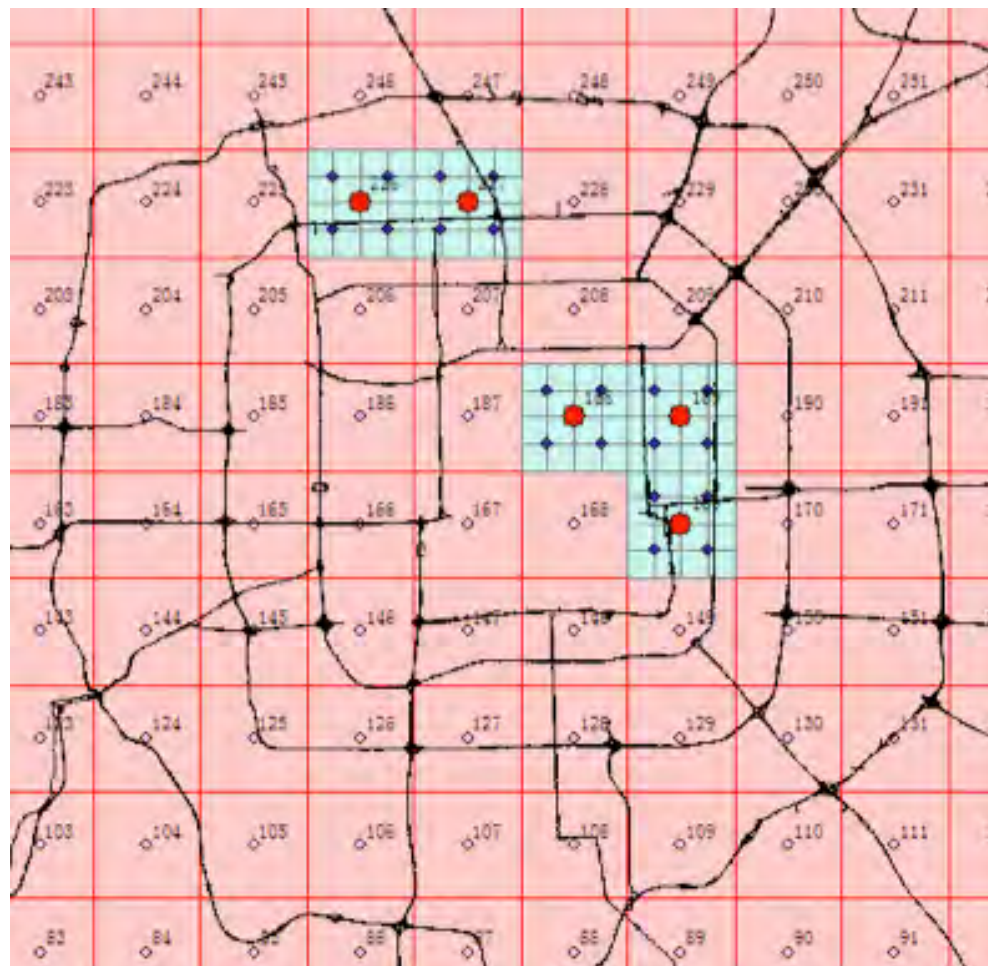
URL

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。

- 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

• 初步分析

• 从LBS微博分析黄金周人口流动

```
{
  "created_at": "Sun Apr 19 23:59:52 +0800 2015",
  "id": "3833492062372488",
  "mid": "3833492062372488",
  "idstr": "3833492062372488",
  "text": "谁偷[花心][心]@允皓Y0ji http://",
  "source_allowclick": 0,
  "source_type": 1,
  "source": "<a href='\"http://app.weibo.com/\">http://app.weibo.com/\",
  "favorited": false,
  "truncated": false,
  "in_reply_to_status_id": "",
  "in_reply_to_user_id": "",
  "in_reply_to_screen_name": "",
  "pic_ids": [
    "74889468jw1erbbk0k10wj21w01w0b2d",
    "74889468jw1erbbk58s0kj21w01w0u11",
    "74889468jw1erbbkrr0gkj21w01w0kjp",
    "74889468jw1erbbkzv0mej21w01w0hdx",
    "74889468jw1erbb14c7aij216o1kwqv5",
    "74889468jw1erbb172c2j216o1kw1ky",
    "74889468jw1erbb1ap5gej216o1lux6p",
    "74889468jw1erbbj1rt4j216o1kwe82",
    "74889468jw1erbb1ckft6j216o1kwkjl"
  ],
  "thumbnail_pic": "http://ww3.sinaimg.cn/large/4e704b16jw1erbb14c7aij216o1kwqv5.jpg",
  "bmiddle_pic": "http://ww3.sinaimg.cn/large/4e704b16jw1erbb14c7aij216o1kwqv5.jpg",
  "original_pic": "http://ww3.sinaimg.cn/large/4e704b16jw1erbb14c7aij216o1kwqv5.jpg",
  "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，也是垃圾",
    "description": "丢了的东西再好，也是垃圾",
    "url": "",
    "profile_image_url": "http://tp1.sinaimg.cn/1955107944/50x50x0/0",
    "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 2014",
    "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/1955107944/100x100x0/0",
    "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,
    "mctype": 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,
      "followers_count": 743,
      "gongyi_level": 0,
      "bind_taobao": 1,
      "hongbao_2014": 1,
      "suishoupai_2014": 0,
      "daily": 0,
      "zongyiji": 0,
      "yiqipao_2015": 0
    },
    "badge_top": "",
    "has_ability_tag": 0,
    "extend": {
      "privacy": {
        "ptype": 0,
        "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-5AFB5807-11"
      }
    ],
    "reposts_count": 0,
    "comments_count": 3,
    "attitudes_count": 5,
    "mlevel": 0,
    "visible": {
      "type": 0,
      "list_id": 0
    }
  },
  "badge": {
    "uc_domain": 0,
    "enterprise": 0,
    "anniversary": 0,
    "taobao": 0,
    "travel2013": 0,
    "gongyi": 0,
    "followers_count": 743,
    "gongyi_level": 0,
    "bind_taobao": 1,
    "hongbao_2014": 1,
    "suishoupai_2014": 0,
    "daily": 0,
    "zongyiji": 0,
    "yiqipao_2015": 0
  },
  "badge_top": "",
  "has_ability_tag": 0,
  "extend": {
    "privacy": {
      "ptype": 0,
      "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-5AFB5807-11"
    }
  ],
  "reposts_count": 0,
  "comments_count": 3,
  "attitudes_count": 5,
  "mlevel": 0,
  "visible": {
    "type": 0,
    "list_id": 0
  },
  "uri_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 22:00:00 2015",
        "timestamp": 1434031650613,
        "uuid": 3524408906907855,
        "activate_status": "0",
        "summary": "朝阳区三里屯路11号",
        "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",
          "telephone": "",
          "country": "中国"
        },
        "keyword": "Let's Burger Plas",
        "action": [
          {
            "place_review": {
              "place_review": {
                "image": {
                  "height": "80",
                  "width": "88",
                  "url": "http://ww3.sinaimg.cn/large/4e704b16jw1erbb14c7aij216o1kwqv5.jpg"
                },
                "target_url": "http://weibo.cn/pages/100101B2094750D66DA0F8429F",
                "biz": {
                  "biz_id": "100101",
                  "containerid": "100101B2094750D66DA0F8429F"
                },
                "url": "http://weibo.cn/pages/100101B2094750D66DA0F8429F",
                "mobile": {
                  "card": {
                    "scheme": "sinaweibo://pageinfo?pageid=100101100101B2094750D66DA0F8429F",
                    "status": 1,
                    "contents": [
                      {
                        "text": "Let's Burger Plas",
                        "is_async": 0,
                        "pic": "http://ww3.sinaimg.cn/large/4e704b16jw1erbb14c7aij216o1kwqv5.jpg",
                        "type": 0
                      }
                    ],
                    "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
      }
    }
  ]
}
```

- 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. 每组利用所提供的北京老城的数据，开展城市研究、规划设计或城市管理支持工作，形成一篇研究论文
 - 中期（开题）和最终汇报以PPT形式为主，鼓励在线可视化
2. 必须开展现场调研
 - 纸上得来终觉浅
3. 如果与其他课程共用课程作业，则建议自觉提高作业难度
 - 如果共用，则请把信息请反馈给助教

此外，期末每个人提交一份课程感想，800字以上，包括建议

清华大学《大数据与城市规划》课程结课汇报 (2016)

TSINGHUA UNIVERSITY BIG DATA AND URBAN PLANNING: FINAL PRESENTATION

主题：城市大数据应用研究——以北京二环内为例

07:50—09:50 上半场汇报 (建筑学院南115; 每组发言15分钟, 问答5分钟)

07:50—08:10 街区要素对城市活力的影响

温祺 清华大学水利水电工程系; 许达, 吴承霖, 张雅敬 清华大学建筑学院

08:10—08:30 历史文化街区街道声音景观探究

周旭影, 刘雪源, 刘为煜, 金程 中国城市规划设计研究院

08:30—08:50 人本尺度城市道路交叉口品质评价

徐特, 徐智伟 清华大学环境学院; 张玉环, 罗圣西 清华大学土木工程系

08:50—09:10 不同空间尺度下情绪研究方法实证对比及其与空间品质关系探究

许宁婧, 王川小雨, 于睿智, 李晶 清华大学建筑学院

09:10—09:30 基于房租数据模型的存量规划研究

沙焯星, 聂聪, 姚宇, 李明玺 清华大学建筑学院

09:30—09:50 基于空间视角的北京二环内影响街道安全性因素与评价

万耀璘 清华大学土木工程系; 冯杰 清华大学电子工程系; 邓铸, 王玲玉 清华地球系统科学研究中心

09:50—10:00 休息及转场

10:00—12:40 下半场汇报 (建筑学院中203; 每组发言15分钟, 问答5分钟)

10:00—10:20 城市物质空间与社会空间国际化匹配程度探讨

周宜笑, 高浩歌, 周宏宇 清华大学建筑学院; 李政圆 清华大学土木工程系

10:20—10:40 空气质量对旅游景点活力的影响调研

詹旭强, 唐博, 唐波哈 清华大学建筑学院; 吴子扬 清华大学环境学院

10:40—11:00 二环内新建轨道站点周边商业活力影响研究

郭璋, 许阳, 肖钧航 中国城市规划设计研究院; 华高英 中国建筑科学研究院

11:00—11:20 新数据环境下大规模城市旧城街道“街道画像”

谢湘雅, 杨东 清华大学建筑学院

11:20—11:40 北京旧城居住小区的开放性评估与资源共享对策研究

秦岭, 刘钊启, 吕回 清华大学建筑学院; 吴承恩 清华大学土木工程系

11:40—12:00 基于空间数据的北京二环内夜间街道活力与影响因素分析

裴昱, 唐义琴, 吴濯杭 清华大学建筑学院; 李婷婷 清华大学土木工程系

12:00—12:20 北京旧城城市气味景观探究

封蓉, 刘璐, 马頔翔, 程情儀 清华大学建筑学院

12:20—12:40 北京旧城建成环境对步行行为的影响研究

曹哲静, 李文越, 张章, 徐高峰 清华大学建筑学院

授课教师：龙瀛，清华大学建筑学院，ylong@tsinghua.edu.cn

点评专家：李栋、刘岱宗、茅明睿、吴康、张纯（按照姓氏拼音排列）

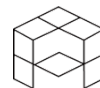
时间：2016年12月30日（周五上午）7:50-12:40

地点：上半场：清华大学建筑学院 南115；下半场：清华大学建筑学院 中203



(更多信息)

清华大学



- 感官数据 | 北京旧城还能闻到哪些“京味儿”
 - http://www.thepaper.cn/newsDetail_forward_1638120
- 感官数据 | 在老北京城还能听到“磨剪子来”的声音吗
 - http://www.thepaper.cn/newsDetail_forward_1638230
- 感官数据 | 什么样的街道，夜间更有活力
 - http://www.thepaper.cn/newsDetail_forward_1639878
- 感官数据 | 空气质量影响了北京旧城旅游景点的活力吗
 - http://www.thepaper.cn/newsDetail_forward_1640765
- 感官数据 | 什么样的街道，夜间更有活力
 - http://www.thepaper.cn/newsDetail_forward_1639878

注：还有多篇作业发表在国际、国内会议上，以及国内期刊投稿中。

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

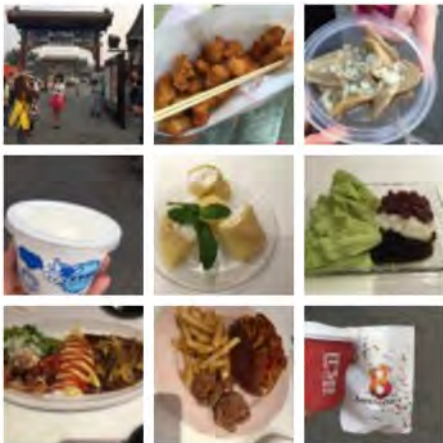
评论 4

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

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



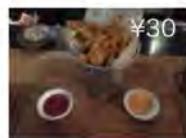
1175

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



炸薯条



啤酒



经典汉堡



酸黄瓜



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



西餐达人

Petrichor1106

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 蓝色奶酪漢堡



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

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.....

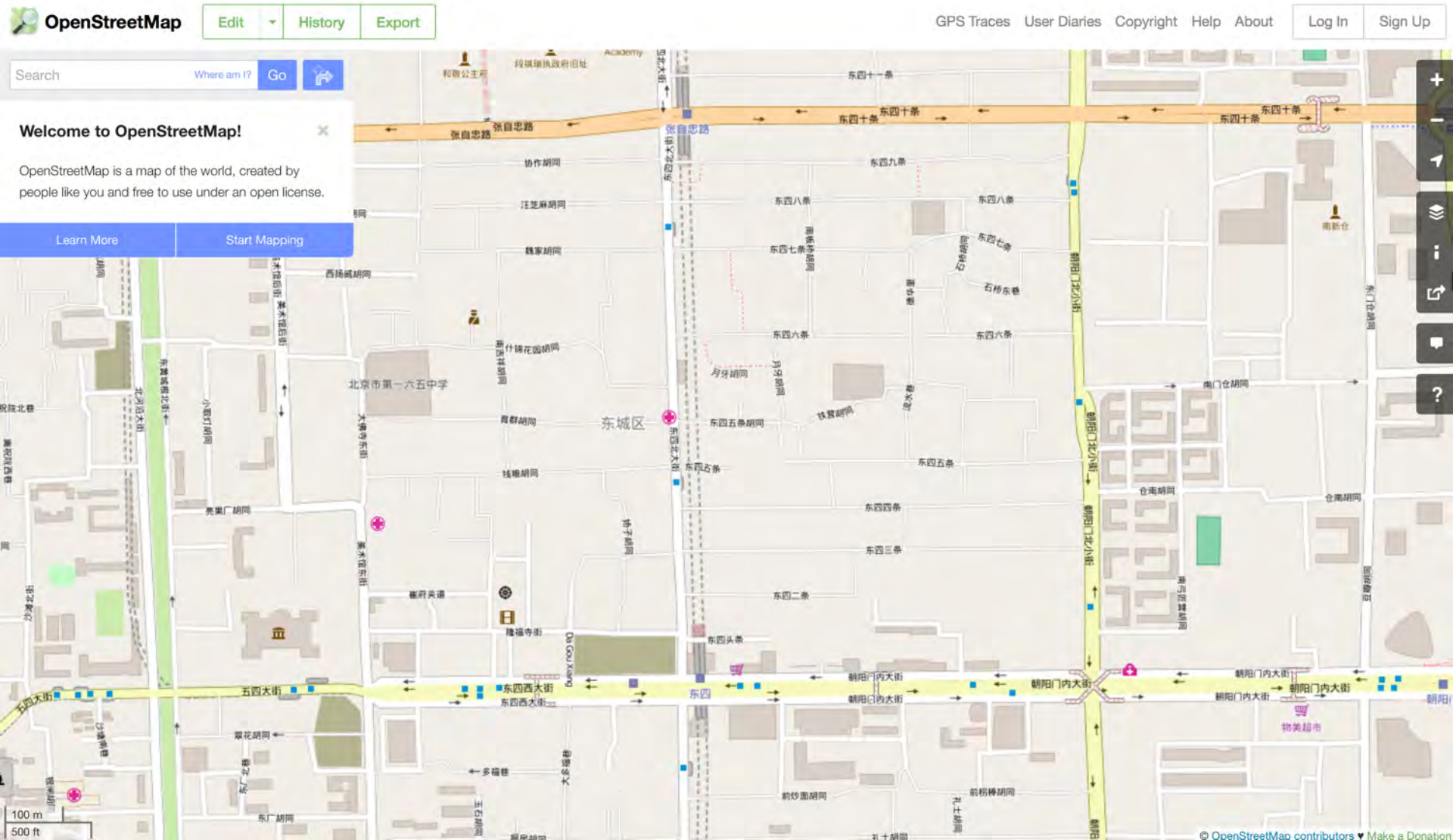
更多

喜欢的菜: 经典汉堡 香草黑啤 西海岸IPA



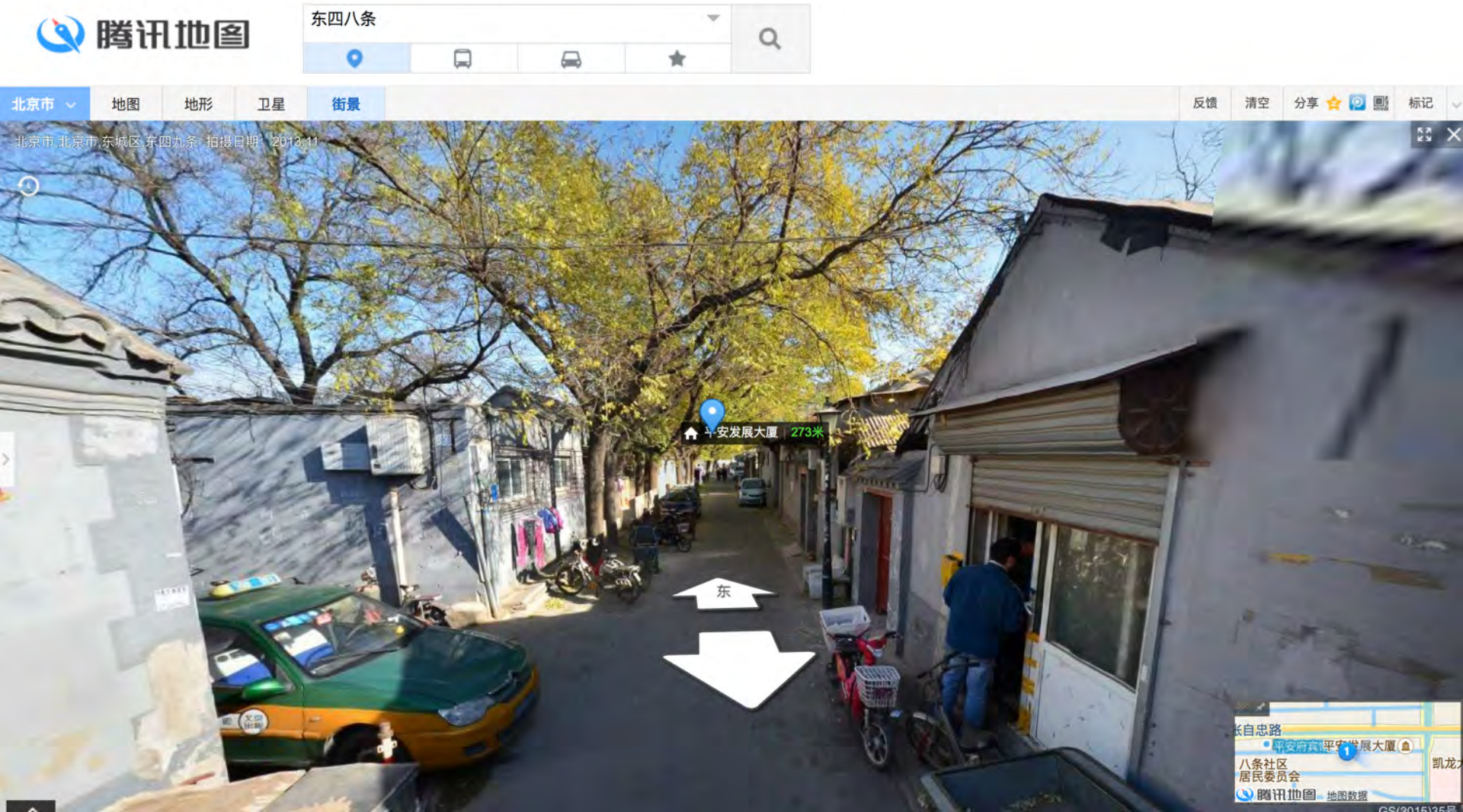
05-22 Slow Boat Brewery

赞(1) 回应(3) 收藏 举报



- www.openstreetmap.org
- 可以开放下载各个城市的数据: <http://download.bbbike.org/osm/bbbike/>

街景地图



- <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



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

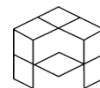
在线考试

表单

投票

360测评

清华大学



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

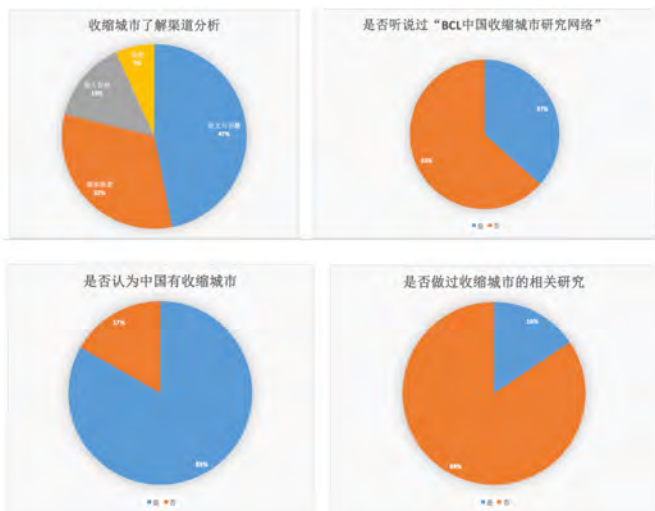
样本总数: 474 份

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:

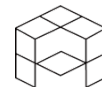
本报告包含样本数量: 474份



- 鄂尔多斯成为被提及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。



清华大学



BCL
Beijing City Lab

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

关于做研究的小窍门

- 第一讲/第二讲：参考文献的重要性
 - 外国人的姓名写法、认真与否、文献等级
- 第三讲：
 - 论文与报告的区别（是否有科学问题）
 - 问题： Problem vs Question
- 第四讲：两类论文
 - 方法：证明方法优于已有的其他方法（效率/科学性、规划师/公众/同行评价？）
 - 实证：证明发现，与其他人发现的异同，对理论的贡献

课后安排

- 每组了解分发数据的基本情况，初步进行选题（欢迎与老师交流想法）
 - 北京老城，围绕开发、形态、功能、活动、活力等维度
- 下周邀请清华同衡李栋博士介绍“无代码抓数据”
- 本课件待放到课程网站
 - <https://www.beijingcitylab.com/bdup2017/>
- 阅读材料：
 - 城市规划与研究中的开放数据（大数据）获取方法及应用分析技术_郑晓伟
 - API与LBS类型数据的获取方法以及城市规划与研究中的应用_郑晓伟
 - 数据获取分析与表达_董磊
- 参考资料
 - 火车头软件使用说明
 - 城市规划系本科生城乡社会综合调研报告
- 答疑
 - ylong@tsinghua.edu.cn
 - 建筑学院新501办公室（默认每周下课后-11AM、Open Office Hour）
 - 建议提前两天通过info系统进行预约
 - 助教：徐婉庭，wantingsu1031@qq.com



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

清华大学

