# 分析思路

1. 找到数据接口
2. 分析接口参数
3. 构造请求测试接口
4. 反爬（暂不考虑）

# 一级页面

## 接口

<https://apphouse.58.com/api/list/chuzu?localname=xm&os=android&geotype=baidu&signature=679193c2c7058bd340b2b7d37198d823&format=json&params=%7B%22list_extra%22%3A%22geren%22%7D&geoia=24.508839%2C118.181664&v=1&curVer=8.12.3&appId=1&action=getListInfo&location=606%2C610%2C15278&page=1&tabkey=allcity&filterParams=%7B%22biz%22%3A%220%22%7D&ts=1542359451574>

<https://apphouse.58.com/api/list/chuzu?localname=xm&os=android&geotype=baidu&signature=9577db8b188d8df0448ef2c1eaf35aff&format=json&params=%7B%22list_extra%22%3A%22geren%22%7D&geoia=24.521726%2C118.157607&v=1&curVer=8.12.3&appId=1&action=getListInfo&location=606%2C610%2C8638&page=2&tabkey=allcity&filterParams=%7B%22biz%22%3A%220%22%7D&ts=1542373534676>

## 参数分析

|  |  |  |  |
| --- | --- | --- | --- |
| **参数名** | **含义** | **参数值** | **备注** |
| localname | 城市 | 城市拼音缩写  如：xm（厦门），sh（上海） | 必选参数 |
| os | 系统 | android | 必选参数 |
| geotype | 搜索类型 | baidu | 可省略 |
| signature | 签名 | 值不固定，如：  679193c2c7058bd340b2b7d37198d823  9577db8b188d8df0448ef2c1eaf35aff | 可省略 |
| format | 格式 | json | 可省略 |
| params | 参数 | %7B%22list\_extra%22%3A%22geren%22%7D  可能是用于将数据过滤为个人房源  建议不省略 | 建议不省略  省略前后可爬取的数据量不同。  以厦门为例：  省略前：140页  省略后：230页 |
| geoia | 未知 | 24.508839%2C118.181664 | 可省略 |
| v | 未知 | 1 | 可省略 |
| curVer | 未知 | 8.12.3 | 必选参数  如果省略会导致只能看到第一页数据 |
| appId | 未知 | 1 | 可省略 |
| action | 动作 | getListInfo | 必选参数  如果省略会导致查不到数据 |
| location | 位置 | 606%2C610%2C15278  可能是当前手机定位 | 可省略 |
| page | 页码 | 最小页码1  最大页码140（不省略params的情况下） | 必选参数 |
| tabkey | 未知 | allcity | 可省略 |
| filterParams | 过滤参数 | %7B%22biz%22%3A%220%22%7D  同params | 同params |
| ts | 未知 | 1542359451574 | 可省略 |

## 结论

经过对一级页面的参数分析，可将url构造为：

[https://apphouse.58.com/api/list/chuzu?localname={city}&os=android&curVer=8.12.3&action=getListInfo&page={page}&params=%7B%22list\_extra%22%3A%22geren%22%7D](https://apphouse.58.com/api/list/chuzu?localname=%7bcity%7d&os=android&curVer=8.12.3&action=getListInfo&page=%7bpage%7d&params=%7B%22list_extra%22%3A%22geren%22%7D)

city：准备好所有需爬取的城市拼音缩写

page：1-140

根据百姓网的数据格式，我们所需的格式为："标题", "时间", "价格", "手机号", "微信号", "地址", "链接", "城市"，其中微信号在58上没有。

一级页面返回的json数据中，包含标题、时间、价格、地址、城市，没有手机号和二级页面链接。但是有其他关键信息如infoID，infoID可以用来构造获取手机号的接口。而获取二级页面链接需要一个signature参数，该参数目前还没找到（可能与一级接口的signature参数有关）。如果二级链接不重要，是否可以舍弃该字段？

关于手机号的获取，直接请求接口取不到手机号，需要构造请求头来获取。

# 二级页面

## 接口分析

<https://apphouse.58.com/api/detail/zufang/35454811607883?localname=xm&commondata=%7B%22pagesource%22%3A23%2C%22product%22%3A%22topinfo%22%2C%22hasNext%22%3Atrue%2C%22nextObserverIndex%22%3A0%2C%22tracekey%22%3A%22f273f86662925d0532389612d4a26673%22%7D&signature=e82d70721faf9e41a5a6be8052fb17e2&v=1&format=json&version=8.12.3&platform=android&sidDict=%7B%22PGTID%22%3A%22108429343202160137362787315%22%2C%22GTID%22%3A%22113399835202160137349079720%22%7D>

可简化为：

[https://apphouse.58.com/api/detail/zufang/35454811607883?localname=xm&signature=e82d70721faf9e41a5a6be8052fb17e2&v=1&format=json&version=8.12.3&platform=android](https://apphouse.58.com/api/detail/zufang/35454811607883?localname=xm&commondata=%7B%22pagesource%22%3A23%2C%22product%22%3A%22topinfo%22%2C%22hasNext%22%3Atrue%2C%22nextObserverIndex%22%3A0%2C%22tracekey%22%3A%22f273f86662925d0532389612d4a26673%22%7D&signature=e82d70721faf9e41a5a6be8052fb17e2&v=1&format=json&version=8.12.3&platform=android&sidDict=%7B%22PGTID%22%3A%22108429343202160137362787315%22%2C%22GTID%22%3A%22113399835202160137349079720%22%7D)

35454811607883和signature为关键参数，前者应该是infoID，可以在一级页面返回的json中找到，signature暂时还没找到。

暂时不考虑二级接口。

# 手机号获取

## 接口分析

<https://houserent.m.58.com/telsecret/getPhone?infoId=35785361668547&phone=15905875707&pageSource=detail&recomlog=&platform=app&verCode=>

传入构造好的请求头，可直接返回包含手机号的json数据

请求头：

headers = {

'apkbus':'',

'xxzlsid':'8XlHz9-QKN-Lj9-PuM-1q6iIdp4S',

'58mac':'64:A2:F9:1A:F3:07',

'official':'true',

'lon':'118.182277',

'accept-encoding':'gzip,deflate',

'ua':'ONEPLUS A6000',

'uuid':'',

'dirname':'xm',

'rimei':'869897039472336',

'jumpinfo':'',

'productorid':'1',

'platform':'android',

'nop':'22818135553823',

'id58':'98714771502027',

'uid':'57849413414166',

'marketchannelid':'147',

'osv':'8.1.0',

'ppu':'UID=57849413414166&UN=kzuex3zt9&TT=a226cbf8444335632fca03be6614fd18&PBODY=Ri6n6Ql0R6A7Whcg8kuMjhSUbVZ9Cvotd0JnMNVV0UrEuoh3zn\_Q4BicqiSesGkERc-yX\_-s\_VOzeC9sfpmGS2O-uurfY7Yp2mNnR9XpH74NSzy4a2b4v\_vOMuuDDX\_Hhlw7OZR4pGLvQws968fuDEVj0XQ3Pe-cxdDmTv4HdOc&VER=1',

'osarch':'arm64-v8a',

'bundle':'com.wuba',

'brand':'OnePlus',

'uniqueid':'186aefd9a1ade76af392782a254d08b3',

'apn':'WIFI',

'lat':'24.510637',

'owner':'baidu',

'totalsize':'108.2',

'product':'58app',

'os':'android',

'rnsoerror':'0',

'deviceid':'440ac65ec0179a35',

'version':'8.12.3',

'm':'64:A2:F9:1A:F3:07',

'ltext':r'%E6%B9%96%E8%BE%B9%E6%B0%B4%E5%BA%93%E7%89%87%E5%8C%BA',

'r':'2280\_1080',

'currentcid':'606',

'xxzl\_deviceid':'gW4xvI45u60V9+HDtd5m+QVvNeH0JeuY477B9I2JS/pK5j+1hs38QaiQUE/TUPwp',

'xxzl\_smartid':'7e38d7253b06baff5e2da99090d28788',

'58ua':'58app',

'nettype':'wifi',

'bangbangid':'1080866010602960165',

'maptype':'2',

'imei':'869897039472336',

'location':'606,610,15278',

'channelid':'147',

'androidid':'440ac65ec0179a35',

'cid':'606',

'locationstate':'1',

'user-agent':'okhttp/3.4.2'

}

可省略：

'apkbus'

'xxzlsid'

'58mac'

'official'

'lon'

'accept-encoding'

'ua'

'uuid'

'dirname'

'rimei'

'jumpinfo'

'productorid'

'platform'

Nop totalsize

id58

uid

marketchannelid

osv

osarch

bundle

brand

uniqueid

apn

lat

owner

totalsize

product

os

rnsoerror

deviceid

version

m

ltext

r

currentcid

xxzl\_deviceid

xxzl\_smartid

58ua

Nettype

Bangbangid

Maptype

Imei

Location

Channeled

Androidid

cid

locationstate

user-agent 建议使用随机user-agent

经分析后需要传入的请求头参数为ppu和user-agent，ppu需要通过模拟登陆获取，user-gent可提前准备。

# 模拟登陆

## 接口分析

<https://passport.58.com/login/dologin>

请求方式：post

Form\_data：

password ATOMZBB5ggwU4EP9\_G-z3Epo9BTf4RN1n6VzuNc\_qk4-Sfl9284GAklqZe9Q9XIu2-QcAcJNBvhxQTvWVXiueobjHOjZdu96Lf7bJsHWN5sLJMDXcKDtePROpXdPRRWup6otW62DxN-Q5SL3cEBOXBftUwLb2-0bQWesnAQGbOdVc6ED8-BO3KgUr459DApTVMheYYhEMzEhkJW2XSDBbUqfm983KDEQD-KXCsbB0gR9ZjVHEvHjuNtzjM1OeVfa-GKzNEADpxICtIayy8yBNCMcGq2MpVfTVLXw9Y8\_CpISuBbDuL3Iee5aYSrodFRj5lNZVcYvobtxGcZtCllv6cYSAdbdrQ0aILVAo2VdslDz8sE0A\_l4yUyUY1MRuvkLXQJrzZVduQ80O8RPotSMgwmmvFc4qojtiuSbDOhZzRZrDL1t2wu\_1lPnQD3U6M9TY4jvdpwDDEVvZywIclqt8aUCo8qZxONuXrCefXpO02rH8ysL4nbT4DqNIZsXLy3qAQ==

vptype RSA2

rsakeyversion 1

source 58app-android

isremember false

validcodetype 200

username 15905875707

返回的json中，参数ppu为关键参数，用于构造获取手机号的请求头

# 解决方案

## 数据准备

1. 58同城账号密码
2. 待爬取的城市拼音缩写
3. 多个移动端的user-agent

## 爬取思路

1. 通过模拟登陆接口，传入58同城账号密码，获取ppu参数
2. 通过一级页面接口，循环爬取每个城市1-140页的数据，数据结构为标题、时间、价格、地址、手机号、城市、infoID(58同城单条信息的唯一标识符，可用于获取手机号)
3. 针对每行数据，通过获取手机号接口，传入infoID和ppu参数构造的请求头来获取手机号
4. 将数据存入MongoDB，手机号作为唯一ID，同手机号仅保留最新爬取的那行