

# 基于 FME 求算最佳租房方案

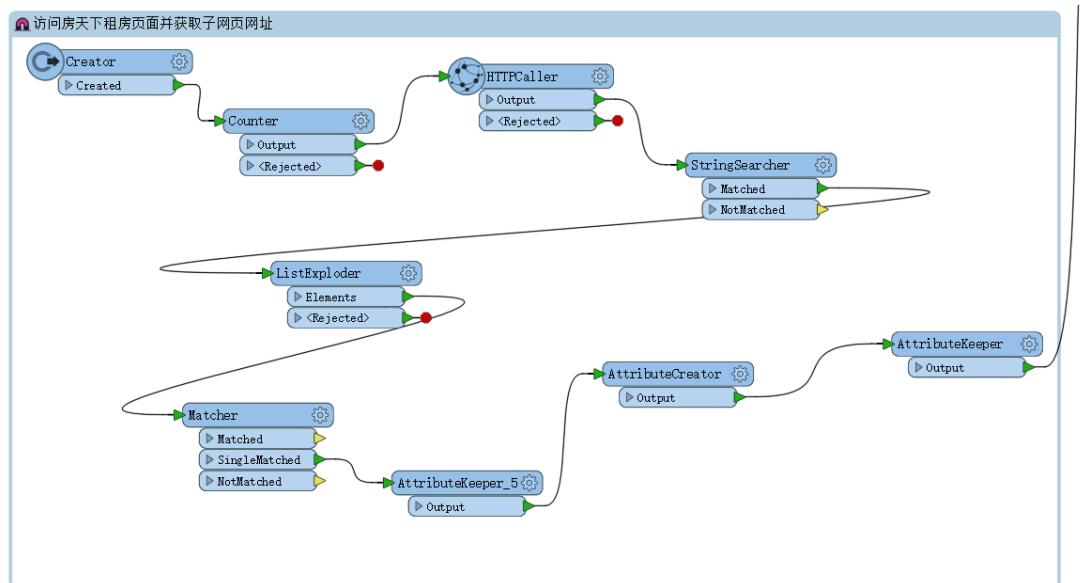
## 摘要

本文基于 FME 实现了最佳租房的求算。首先抓取了租房网站上房源的信息，再根据房源所在的公寓的名称或详细地址利用百度地图 API 地理解码功能获取房源的坐标信息。将房源坐标由经纬度转为米坐标后，计算各房源到上班地点的距离，再结合房源价格、附近有无地铁、合租户数、配套设施给各房源计算一个分数。

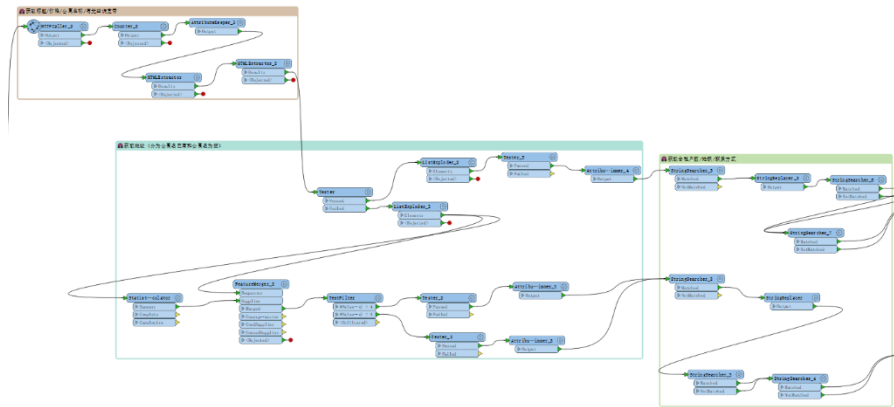
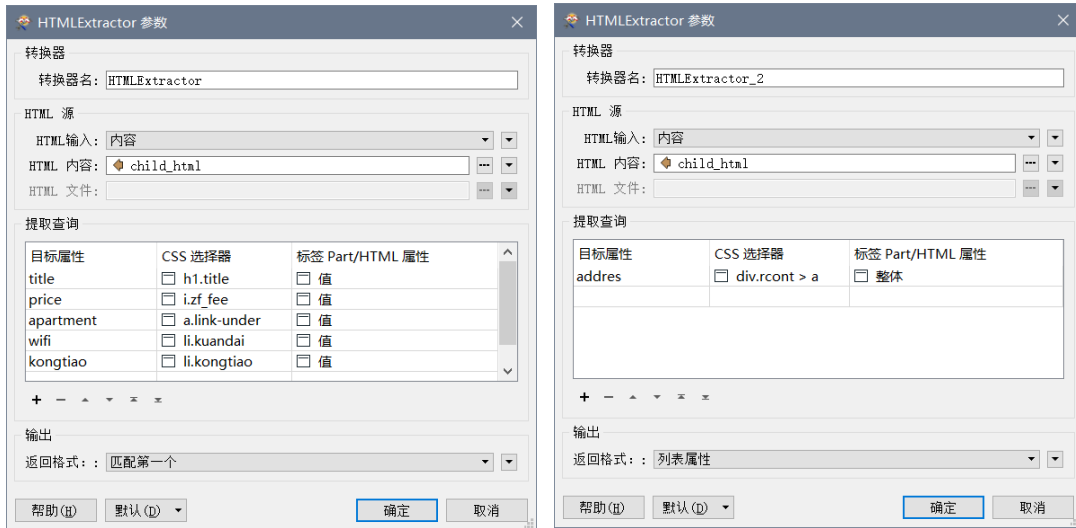
## 1 抓取租房信息

目标是抓取房天下租房前三页的租房信息。

1) 首先利用 Creator 创建 3 个要素，对要素进行计数，计数结果记录在 `page_num` 属性中，通过 HTTPCaller 访问房天下成都租房 500—1000 元的个人合租房源页面，利用正则表达式得到各房源的子网页的网址。

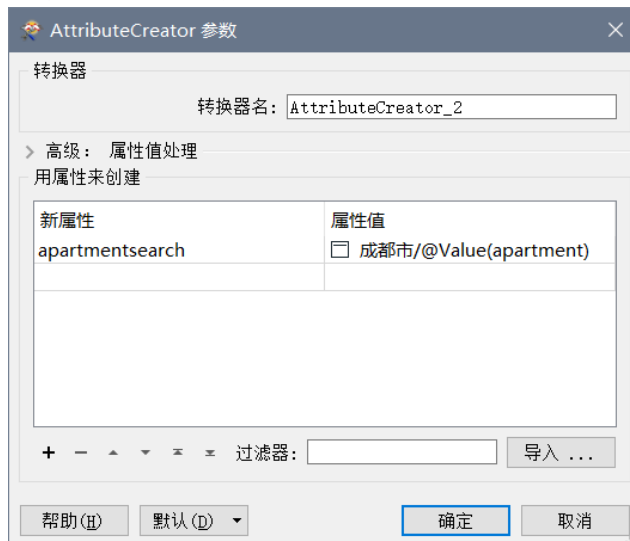


2) HTTPCaller 访问各个房源，并使用 HTMLExtractor，通过 CSS 选择器获取标题、价格、公寓名称、地址、配套设施（主要关心宽带和空调）；使用 StringSearcher，通过正则表达式获取地铁、联系方式、合租户数。



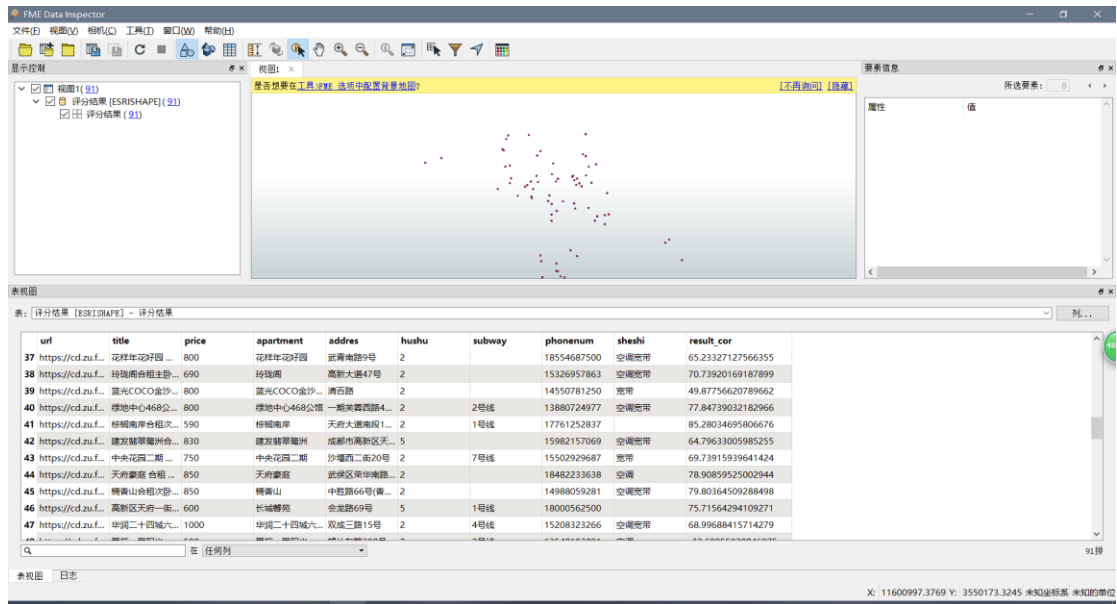
## 2 获取百度地图坐标

1) 通过观察，发现某些地址没有详细到门牌号，所以首选的搜索方式是利用公寓名称搜索，为了防止其他城市有同名的公寓，在公寓名称前加上（成都市）。





## 4 成果展示



## 5 心得体会

大数据是时下的热点词。本文使用 FME 平台实现了互联网数据的抓取、数据矢量化和数据分析。发现 FME 平台在互联网数据方向具有很好的可调整性、很强的拓展性，将数据以要素方式存储到内存中使得数据的获取和使用十分有条理。由于本文抓取的特定网页的信息，如果需要抓取其他网页的信息，则需要修改某些转换器的参数与设置，但相比其他方式，FME 平台的可迁移能力还是十分突出。FME 平台具有使用简便、功能强大、在不同环境中适应能力强等特点。