

# 数据需求

## 目的

根据输入项目名称定向抓取唐山住建局中各类预售信息。

## 数据来源

从住建局 <http://zhujianju.tangshan.gov.cn/> 找到“在线服务”进入“预售查询” (图 1)，在第二层页面 (图 2) 根据“项目名称”输入关键字“项目名称”进行查询。



图 1



图 2

在第三层查询结果页面根据预售证号的排列进行逐条数据抓取，首先在第三层查询结果页面（图 3）选择“详情”进入第四层页面（图 4），抓取第四层“详情”页面的“预售证号、房屋坐落位置、项目名称、发证日期、售房单位”；



图 3



图 4



同时存在在第三层查询结果页面（图 3）选择“楼盘信息”进入第四层页面（图 5），根据第四层页面中的“幢号”逐一对“房屋列表”下一级页面（图 6）进行全部数据抓取，包括“幢号（第四层页面）、单元号、房名、建筑面积、套内面积 当前层、状态、是否抵押”。



唐山市住房和城乡建设局  
BUREAU OF HOUSING & URBAN-RURAL DEVELOPMENT.TANGSHAN

首页 | 住建动态 | 政务公开 | 公示公告 | 政民互动 | 办事指南 | 在线服务

请输入关键词 本站搜索

预售证楼栋列表

幢号	预售许可证	总套数	总面积	总层数	状态	楼盘表
0007	(唐)房预售证第1887号	127	9647.94	18	可售	房屋列表
0011	(唐)房预售证第1887号	114	8272.69	20	可售	房屋列表
0010	(唐)房预售证第1887号	116	8251.52	20	可售	房屋列表

图 5



唐山市住房和城乡建设局  
BUREAU OF HOUSING & URBAN-RURAL DEVELOPMENT.TANGSHAN

首页 | 住建动态 | 政务公开 | 公示公告 | 政民互动 | 办事指南 | 在线服务

请输入关键词 本站搜索

单元号	房名	建筑面积	套内面积	当前层	状态	是否抵押
01	1门101号	123.24	97.27	01	未售	
01	1门102号	122.28	96.51	01	未售	
01	1门201号	123.24	97.27	02	未售	
01	1门202号	122.28	96.51	02	未售	
01	1门301号	123.24	97.27	03	未售	
01	1门302号	122.28	96.51	03	未售	
01	1门401号	123.24	97.27	04	未售	

图 6

以此类推，逐条抓取预售证详情数据、预售证楼盘表内各幢号的房屋列表数据，直到将所查询项目数据全部抓取。

报表格式

报表（图 7）格式为参考，对报表只要求数据详实，文件输出格式为 EXCEL 格式。其他无要求

售房单位	项目名称	房屋坐落位置	预售许可证	发证日期	幢号	单元号	房名	建筑面积	套内面积	当前层	状态	是否抵押

图 7