**《Python程序设计》**

**小作业1：链家**



**学院：计算机学院（国家示范性软件学院）**

**班级：2021211313**

**姓名：吕子健**

**学号：2023523012**

1. **实验目的**

爬取北京链家官网二手房数据 https://bj.lianjia.com/ershoufang/

具体爬取信息及存储要求：

• 要求爬取东城、西城、海淀、朝阳四个城区的二手房数据（每个城 区爬取前5页），提取楼盘名称、平米数、总价、单价 。

• 将信息保存在csv文件中（可以是一个，也可以是每城区一个）

1. **实验内容（代码）**

import requests

from bs4 import BeautifulSoup

import csv

import os

os.chdir(os.path.dirname(\_\_file\_\_))

districts = ["dongcheng", "xicheng", "haidian", "chaoyang"]

pages = 5

with open("链家.csv", "w", encoding="gbk") as f:

    writer = csv.writer(f)

    writer.writerow(["楼盘名称", "平米数", "总价", "单价"])

    for district in districts:

        if district == "dongcheng":

            area\_name = "东城"

        elif district == "xicheng":

            area\_name = "西城"

        elif district == "haidian":

            area\_name = "海淀"

        elif district == "chaoyang":

            area\_name = "朝阳"

        writer.writerow([area\_name])

        for page in range(1, pages + 1):

            url = f"https://bj.lianjia.com/ershoufang/{district}/pg{page}/"

            response = requests.get(url)

            if response.status\_code == 200:

                # 解析响应内容，创建BeautifulSoup对象

                soup = BeautifulSoup(response.text, "html.parser")

                # 找到所有的房源信息

                houses = soup.find\_all("div", class\_="info clear")

                for house in houses:

                    # 提取楼盘名称

                    name = house.find("div", class\_="title").a.text

                    # 提取平米数

                    area = house.find("div", class\_="address").find("div", class\_="houseInfo").text.split("|")[1].strip()

                    # 提取总价

                    total\_price = house.find("div", class\_="priceInfo").find("div", class\_="totalPrice").span.text

                    # 提取单价

                    unit\_price = house.find("div", class\_="priceInfo").find("div", class\_="unitPrice").span.text

                    writer.writerow([name, area, total\_price, unit\_price])

            else:

                print(f"请求失败，状态码为{response.status\_code}")

1. **实验结果**



