

实验名称：网络爬虫

姓名：廖俊轩

2020 年 11 月 22 日

## 1. 平台环境

Ide: pycharm

编程语言: python

使用库函数: requests、requests\_html、re

Python 环境: python3

操作系统: 一切可以安装 python3 的系统

## 2. 架构设计

程序流程图:

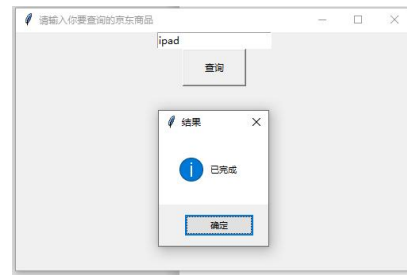


请求网页使用 requests

网页解析则采用 requests-html 库

Gui 则使用 tkinter 库

### 3. 实现界面或测试结果



```
¥ 2499.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 5999.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 5999.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 4799.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 7029.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 2521.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 2099.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 4799.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 4799.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营  
¥ 6229.00  
APPLE 苹果 iPad Pro 12.9英寸平板电脑 2020年新款 256G WLAN版/A14芯片/摄像头/全面屏/5G网络 深空灰色 【官方标配】  
京东 自营
```

数据会存在与查询关键词同名文件夹下

### 4. 总结与展望

本次实验的难点主要是分析目标网页的 html 文件，将所选取对的目标信息解析出来。同时为了让文本更具有可读性。本程序使用了 re 库，对 token 文本进行了正则表达式的匹配。同时为了便于用户使用做了个 gui，让用户可自行输入关键词。

实验名称：聊天室

姓名：廖俊轩

2020 年 11 月 22 日

### 5. 平台环境

Ide: idea ij

注 ideaij 请打开 parallel

编程语言: java

编译环境： java8

架构： sprintboot

辅助软件： RabbitMQ 及其支撑软件

## 6. 架构设计

可能要启动 rabbit mq 的 stomp 插件，启动命令：

```
rabbitmq-plugins enable rabbitmq_stomp
```

```
rabbitmq-plugins enable rabbitmq_web_stomp_examples
```

```
rabbitmq-plugins enable rabbitmq_web_stomp
```

Java 类解释：

WebSocketConfig:

@EnableWebSocketMessageBroker 用于启用我们的 WebSocket 服务器。我们实现了 WebSocketMessageBrokerConfigurer 接口，并实现了其中的方法。在第一种方法中，我们注册一个 websocket 端点，客户端将使用它连接到我们的 websocket 服务器。withSockJS() 是用来为不支持 websocket 的浏览器启用后备选项，使用了 SockJS。方法名中的 STOMP 是来自 Spring 框架 STOMP 实现。STOMP 代表简单文本导向的消息传递协议。它是一种消息传递协议，用于定义数据交换的格式和规则。为啥我们需要这个东西？因为 WebSocket 只是一种通信协议。它没有定义诸如以下内容：如何仅向订阅特定主题的用户发送消息，或者如何向特定用户发送消息。我们需要 STOMP 来实现这些功能。在 configureMessageBroker 方法中，我们配置一个消息代理，用于将消息从一个客户端路由到另一个客户端。第一行定义了以 “/app” 开头的消息应该路由到消息处理方法（之后会定义这个方法）。第二行定义了以 “/topic” 开头的消息应该路由到消息代理。消息代理向订阅特定主题的所有连接客户端广播消息。使用 RabbitMQ 代替内存作为消息代理。

ChatMessage :

ChatMessage 用来在客户端和服务端中交互。

实体中，有三个字段：

type: 消息类型

content: 消息内容

sender: 发送者

类型有三种:

CHAT: 消息

JOIN: 加入

LEAVE: 离开

ChatController:

我们在 websocket 配置中, 从目的地以 /app 开头的客户端发送的所有消息都将路由到这些使用 @MessageMapping 注释的消息处理方法。

WebSocketEventListener:

对 socket 的连接和断连事件进行监听, 这样我们才能广播用户进来和出去等操作. 我们已经在 ChatController 中定义的 addUser () 方法中广播了用户加入事件。因此, 我们不需要在 SessionConnected 事件中执行任何操作。在 SessionDisconnect 事件中, 编写代码用来从 websocket 会话中提取用户名, 并向所有连接的客户端广播用户离开事件。

网页:

客户端

ChatService:

服务器把消息从 Redis 中拿出来, 并且推送到自己管的用户那边, 我们在 Service 层实现消息的推送。

RedisListenerHandle:

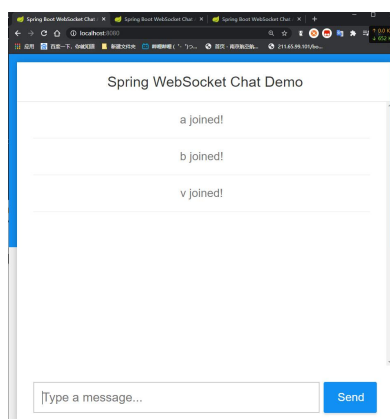
Redis 监听消息处理专用类

RedisListenerBean:

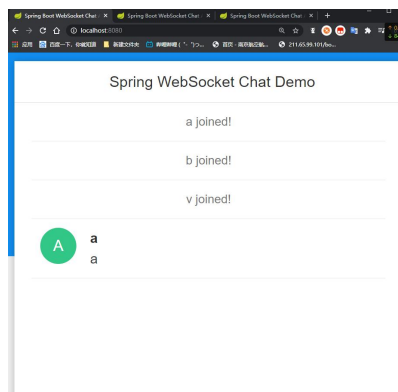
redis 消息监听器容器

前端地址 : localhost:8080

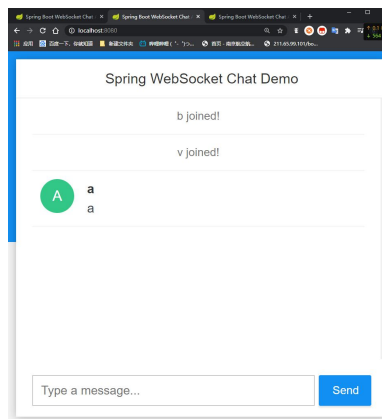
## 7. 实现界面或测试结果



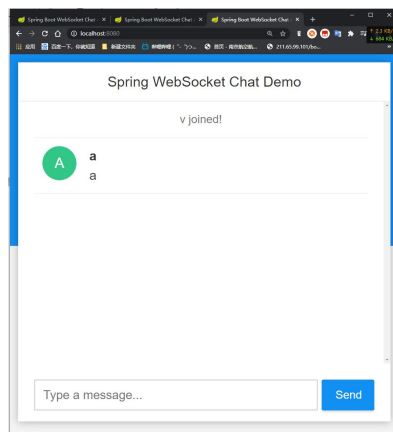
A, b, v 加入聊天室



A 发送消息（a 视角）



A 发送消息（b 视角）



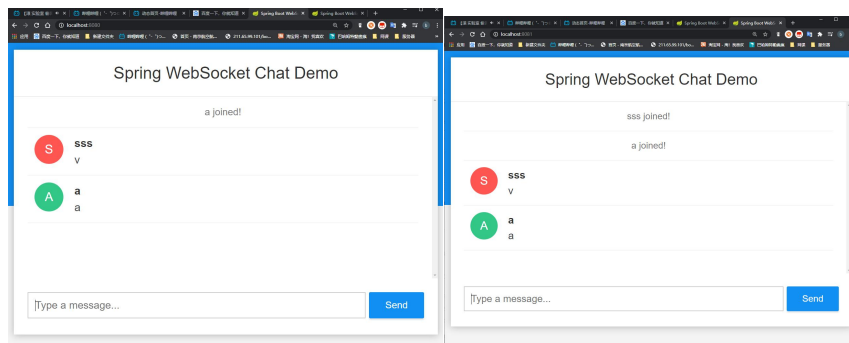
A 发送消息（v 视角）

以上为同端口情况下

不同端口如下

我们将服务器运行在 8080 上，然后打开 localhost:8080，起名 Alice 进入聊天室。随后，我们在 application.properties 中将端口 server.port=8081

再次运行程序。注意将 ideaij 的 parallel 开启



## 8. 总结与展望

本项目采用 spring 架构，有 rabbitmq 管理内存以提高并发量。通过本次实验，我了解到了网络编程的知识。并且通过 java 语言编写出了网络聊天室。通过动手的方式加深了对应用层的理解。帮助我为后续的学习打下基础。