**Lab3：ESI学科数据获取及分析**

姓名：\_\_\_\_\_\_徐卓尔\_\_\_\_ 学号：\_\_\_\_10245102419\_\_\_\_\_\_

日期：\_\_\_\_9.30\_\_\_\_\_\_

# 一、实验目的

1. 熟悉 Selenium 自动化测试工具的使用方法。  
2. 学习如何通过 Python 脚本与浏览器交互，实现数据采集。  
3. 掌握从动态网页中滚动加载表格数据并保存到本地的流程。  
4. 实现对 ESI 数据库中不同学科的全球大学排名数据采集，为后续分析打下基础。

# 二、实验环境

操作系统：Windows 10 / 11  
编程语言：Python 3.12  
第三方库：selenium、pandas、openpyxl  
浏览器：Microsoft Edge   
ESI 数据库访问地址：https://esi.clarivate.com

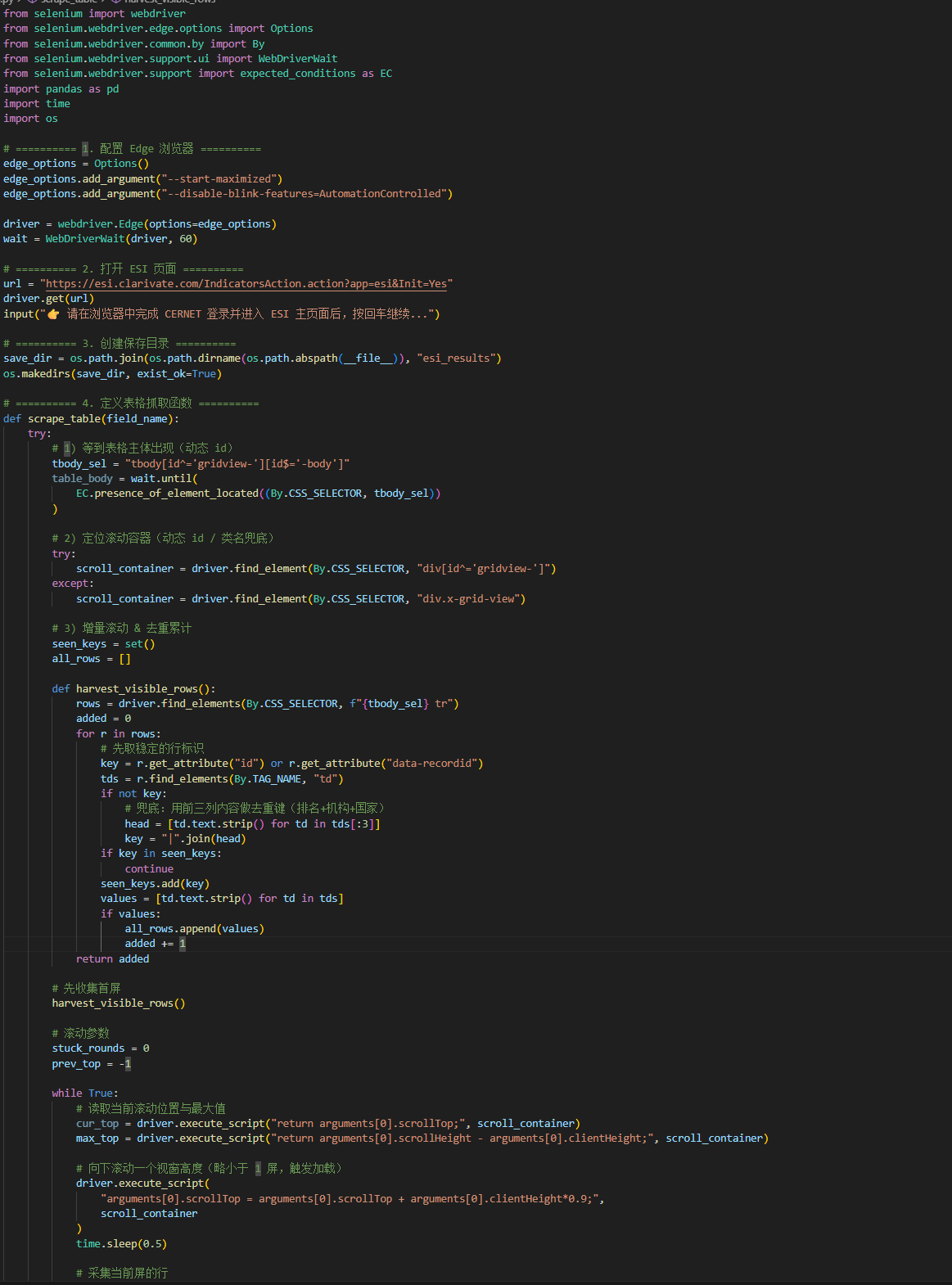
# 三、实验内容

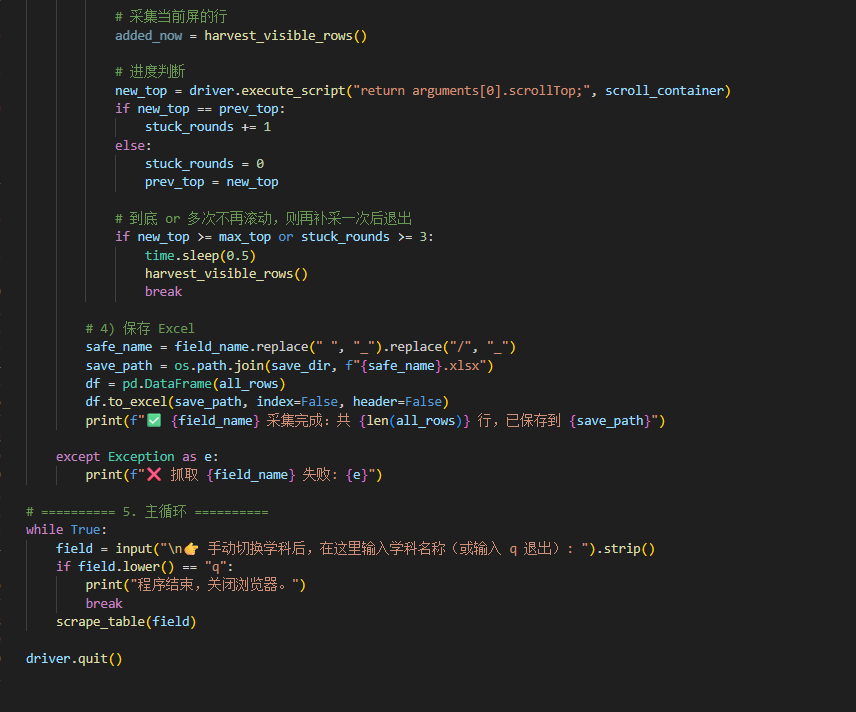
1. 通过 Selenium 启动浏览器并打开 ESI 数据库页面。  
2. 在终端提示下，手动完成 CERNET 登录并进入 ESI 主页面。  
3. 通过程序实现对 ESI 表格的滚动加载和数据抓取。  
4. 采用人工与程序结合的方式：  
 - 人工在浏览器界面手动切换学科；  
 - 在终端输入对应学科名称；  
 - 程序自动抓取该学科下的全球大学排名数据并保存为 Excel 文件。  
5. 所有数据统一保存到 esi\_results 文件夹中，文件名为对应学科名称。

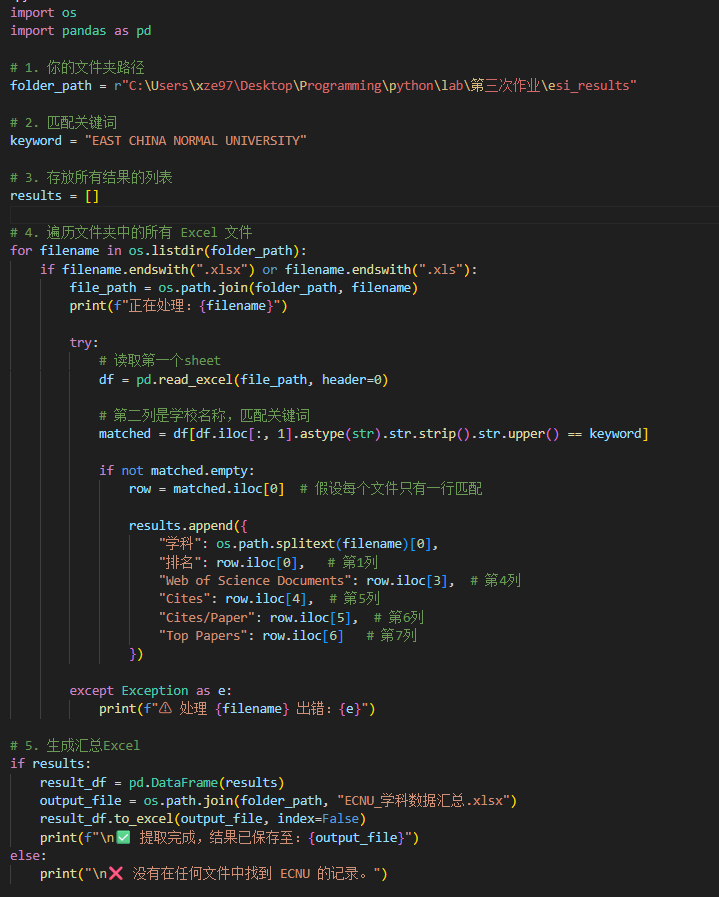
6.提取所有表格中ECNU的数据

**实验代码截图**

**1.爬取网络数据**

****

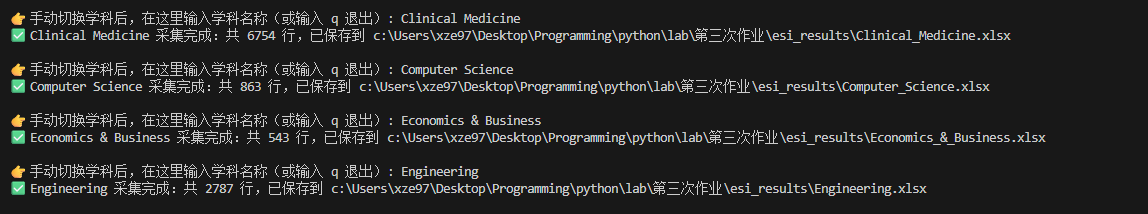
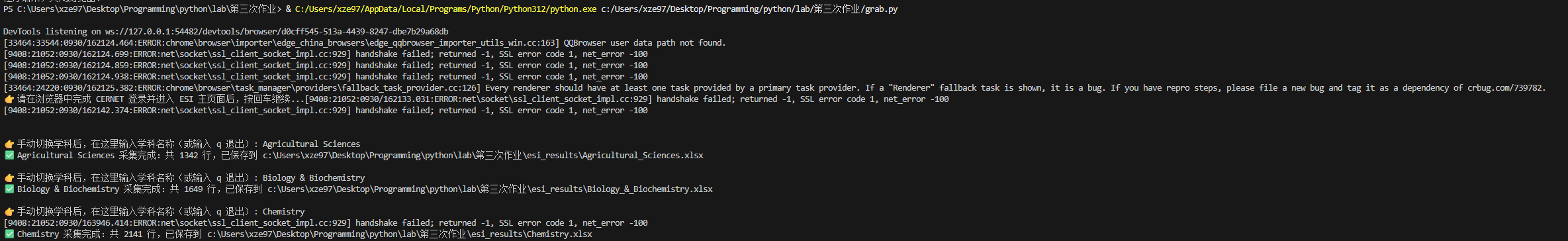


**2.提取ECNU数据**

# **四、实验步骤**

1. 环境准备：安装 Python 和所需第三方库。  
2. 编写爬虫脚本：配置 Selenium、打开网页、滚动加载数据。  
3. 数据采集与保存：提取表格数据，保存为 Excel 文件。  
4. 用户交互逻辑：用户切换学科并输入学科名称，程序保存数据。

# 五、实验结果与分析

1. 实验成功实现了 ESI 数据的分学科采集，并保存为 Excel 文件。  
2. 每个学科数据独立保存，便于单学科分析。  
3. 程序支持多次学科切换采集，健壮性较好。  
4. 若网络延迟，程序会自动等待数据加载完成。



# 

# 六、实验总结

1. 本实验通过 Python + Selenium 成功实现了对 ESI 数据库的半自动化采集。  
2. 解决了动态 ID 变化、滚动加载、登录认证等问题。  
3. 加深了对网页结构解析、自动化操作、数据采集与存储的理解。  
4. 实验结果为后续学科比较和科研实力分析提供了数据支持。