基于Uiautomation的百度指数获取

**百度指数：**搜索指数是以网民在百度的搜索量为数据基础,以关键词为统计对象,科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和。其在研究关键词搜索趋势、洞察网民需求变化、监测媒体舆情趋势、定位数字消费者特征方面有重要的作用。

一些教程在获取百度搜索指数的时候，使用的是爬虫技术，通过分析请求参数来获取指数信息。一些教程提出了使用截图进行，然后进行图像识别，那么对于不懂这些技术的小伙伴该怎么进行指数的获取呢？

本文将通过uiautomation来进行获取，学好了这个技术你还可以对桌面系统进行自动化。

1. **什么是uiautomation**

Microsoft UI 自动化是一个辅助功能框架，它使 Windows 应用程序能够提供和使用有关用户界面 (UI) 的编程信息。它提供对桌面上大多数 UI 元素的编程访问。它使辅助技术产品（例如屏幕阅读器）能够向最终用户提供有关 UI 的信息并通过标准输入以外的方式操作 UI。UI 自动化还允许自动化测试脚本与 UI 交互。

具体可以从以下链接中学习：

<https://docs.microsoft.com/en-us/dotnet/framework/ui-automation/ui-automation-overview?redirectedfrom=MSDN>

当然了先安装：pip install uiautomation

这个uiautomation库并不是官方的，而是一个作者自己封装的。感谢这位作者，大家可以从这个链接学习：

https://github.com/yinkaisheng/Python-UIAutomation-for-Windows/

1. **浏览器的启动**

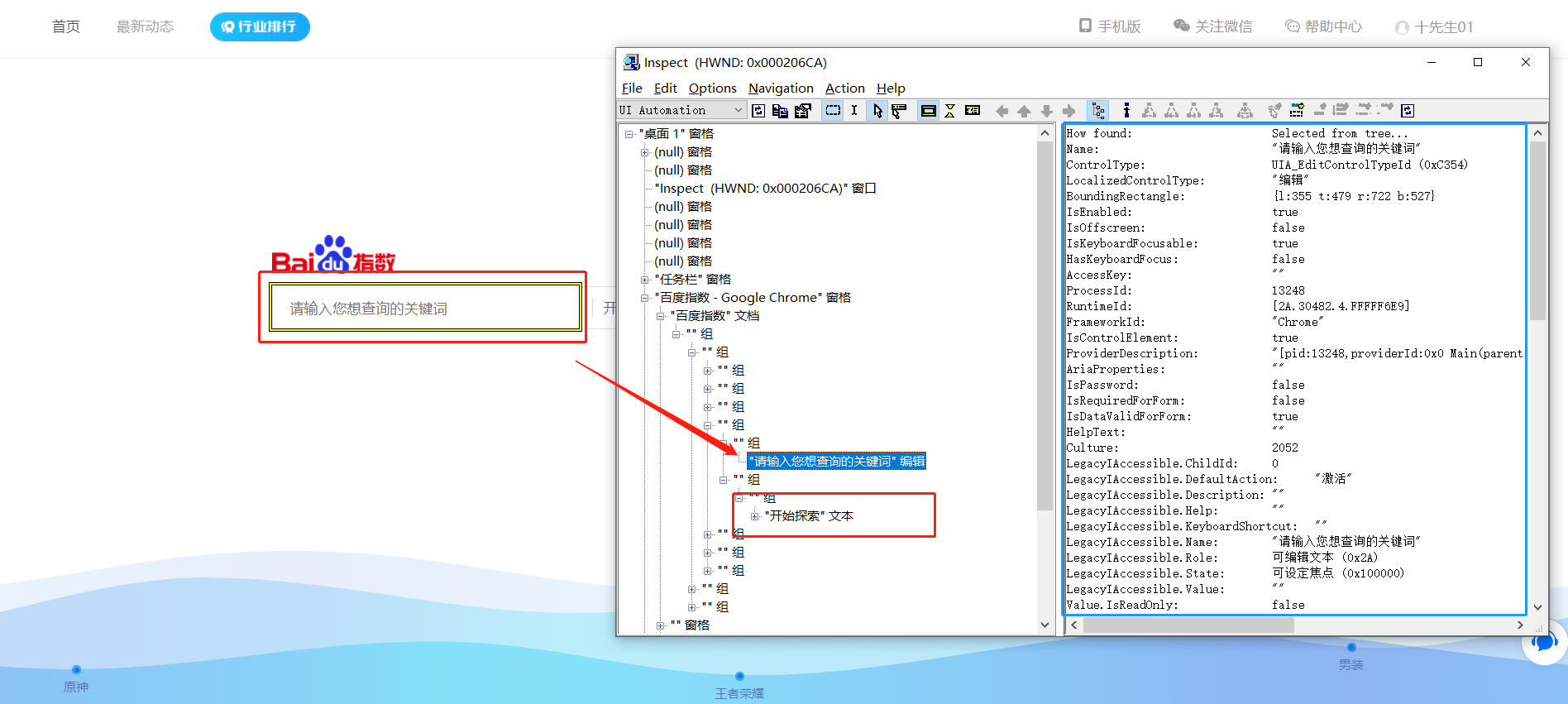
由于我们是要基于UI来进行数据的获取的，如果这个应用程序不是使用微软提供的标准控件来实现的就不能使用uiautomation,也就是说不支持UI框架应用程序是不能使用Uiautomation了，比如 Chrome和基于Electron开发的应用程序。但是也有解决办法：启动时添加参数--force-renderer-accessibility才能支持UIAutomation。

这里我们启动Chrome并在百度搜索指数网址：

那么如何找到这个搜索框的位置和“开始搜索”的按钮呢？



我们可以借助一个工具inspect.exe去寻找，双击这个程序之后，我们找到这个输入框：



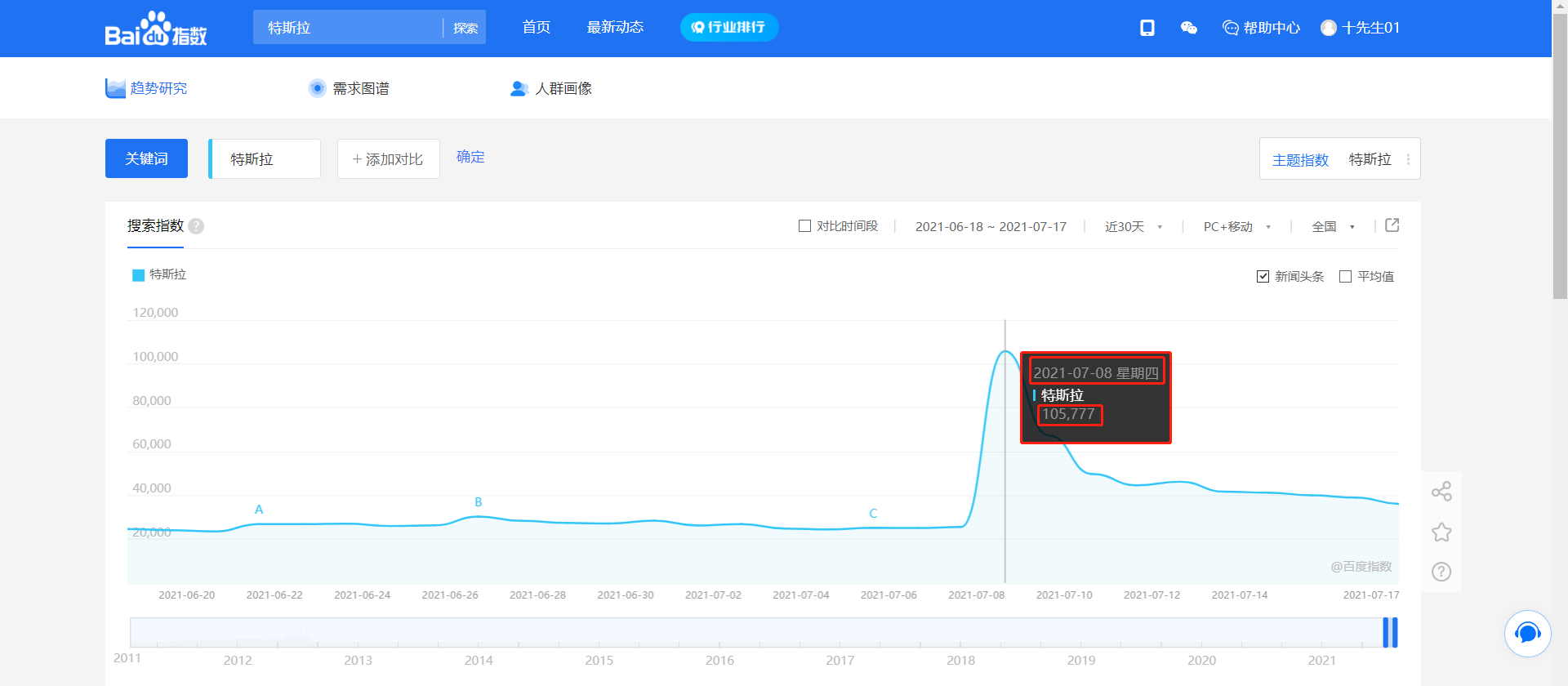
借助inspect可以看出当前页面的程序控件信息，如上图可以看出，其Name为"请输入您想查询的关键词"，ControlType为UIA\_EditControlTypeId，所以可以使用以下代码实现我们的功能了，代码中没有进行控件查找超时的异常处理，大家可以自己添加。

**import** subprocess  
**import** uiautomation **as** auto  
  
**def** show\_index\_window():  
 print(**'root Control:'**, auto.GetRootControl())  
 chromePath = **"C:\\Program Files (x86)\\Google\\Chrome\\Application\\chrome.exe"** url = **r'https://index.baidu.com/v2/index.html#/'** parameter = **'--force-renderer-accessibility'** startmax = **'-start-maximized'** run\_cmd = chromePath + **' '** + url + **' '** + parameter + **' '** + startmax  
 subprocess.Popen(run\_cmd)  
 mainWindow = auto.DocumentControl(ClassName=**'Chrome\_RenderWidgetHostHWND'**)  
 **if** mainWindow.Name == **'百度指数'**:  
 print(**'open success'**)  
  
 *# 输入"特斯拉"并点击"开始探索按钮"* edit = auto.EditControl(mainWindow, Name=**'请输入您想查询的关键词'**)  
 **try**:  
 edit.SendKeys(**'特斯拉'**)  
 *# editConttol.GetPattern(auto.PatternId.ValuePattern).SetValue('特斯拉')* **except** LookupError **as** ex:  
 **return "find control time out "** *# 点击按钮  
 # time.sleep(1)* textControl = auto.TextControl(mainWindow, Name=**'开始探索'**)  
 **try**:  
 print(textControl.Name)  
 **except** LookupError **as** ex:  
 **return** **"find control time out "**

rect = textControl.BoundingRectangle  
 print(rect)  
 left = rect.left  
 top = rect.top  
 right = rect.right  
 bottom = rect.bottom

*# 进行点击*  
 *# 作差取得控件中间位置也可* auto.Click(left + 10, top + 10)

这样就自动进入到“特斯拉”百度指数详情界面。



接下来我们就来提取其中的搜索指数数字信息。

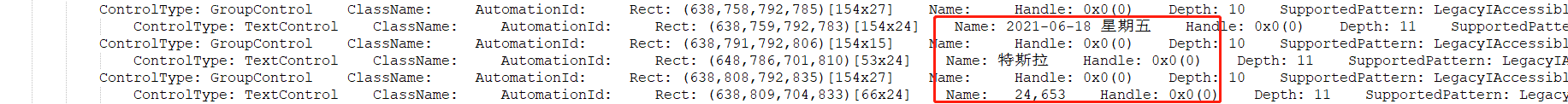
1. **提取指数信息**

通过inspect发现搜索指数信息的控件是图像类型的即，ControlType为ImageControl。因此打开开发者模式是无法提取到的。但是我们又需要里面的信息，我们这样来操作：

因为这些数字是通过悬浮窗口才显示的，我的做法是这样的：先运行程序：

python D:\Python\Python38\Scripts\automation.py -t 8

然后切换到指数界面，将鼠标移动图像曲线开始处的位置上比如2021-06-18那天的指数数据(也不一定是曲线)显示指数信息上图所示。然后可以在automation.py同一目录出现控件信息文件@AutomationLog.txt。在这个控件信息文件中我们找到目标：



可以看出，目标数值24653就显示出来了：

那么怎么将鼠标移动到开始的位置，很简单。先获取控件然后获取属性BoundingRectangle即可，跟上述点击“**开始探索**”按钮一样。这里留给读者自己实现。在不能直接获取到目标控件的时候，可以先获取某一个元素，比如“新闻头条”等控件，然后获取其父元素，在获取子元素。



接下来我们就通过“新闻头条”控件来获取目标信息：

代码如下：

*# -\*- coding: utf-8 -\*-***import** time  
**import** uiautomation **as** auto  
  
**def** get\_index\_baidu():  
 mainWindow = auto.PaneControl(ClassName=**'Chrome\_WidgetWin\_1'**)  
 **if** mainWindow.Exists(3,1):  
 handle = mainWindow.NativeWindowHandle  
 *# auto.SetWindowTopmost(handle, 'True')  
 # auto.SwitchToThisWindow(handle)* auto.ShowWindow(handle, auto.SW.Maximize)  
 news = auto.ListItemControl(mainWindow, Name = **'新闻头条'**)  
 **try**:  
 print(**'news:'**,news)  
 *# editConttol.GetPattern(auto.PatternId.ValuePattern).SetValue('特斯拉')* **except** LookupError **as** ex:  
 **return "find control time out"**  
 *# 获取父元素* f\_new = news.GetParentControl()  
 ff\_new = f\_new.GetParentControl()  
 *# 下一元素* target = ff\_new.GetNextSiblingControl()  
 *# 第一个子元素* target\_z = target.GetChildren()[0]  
 *# 等待元素加载* **"""  
 这里调用鼠标移动到曲线上。  
 如果想要自动化，则需要获取内置控件的信息，然后通过坐标来移动。  
 """** *# 这里等待移动鼠标到曲线上* time.sleep(2)  
 f\_target\_z = target\_z.GetChildren()[1]  
 target\_son = f\_target\_z.GetChildren()  
  
 **for** each **in** target\_son:  
 text\_control = each.GetChildren()[0]  
 print(text\_control.Name)

输出结果如下：

2021-07-08 星期四

特斯拉

105,777

1. **总结**

UIautomation在自动化方面有着很广泛的应用，不仅可以对浏览器自动化而且还可以对一些可视化界面进行自动化。网上的资料也不是很多，大家可以对源码研究来进行更深的学习，当然了还得多实践。