# 技术分享

### 电子递交项目

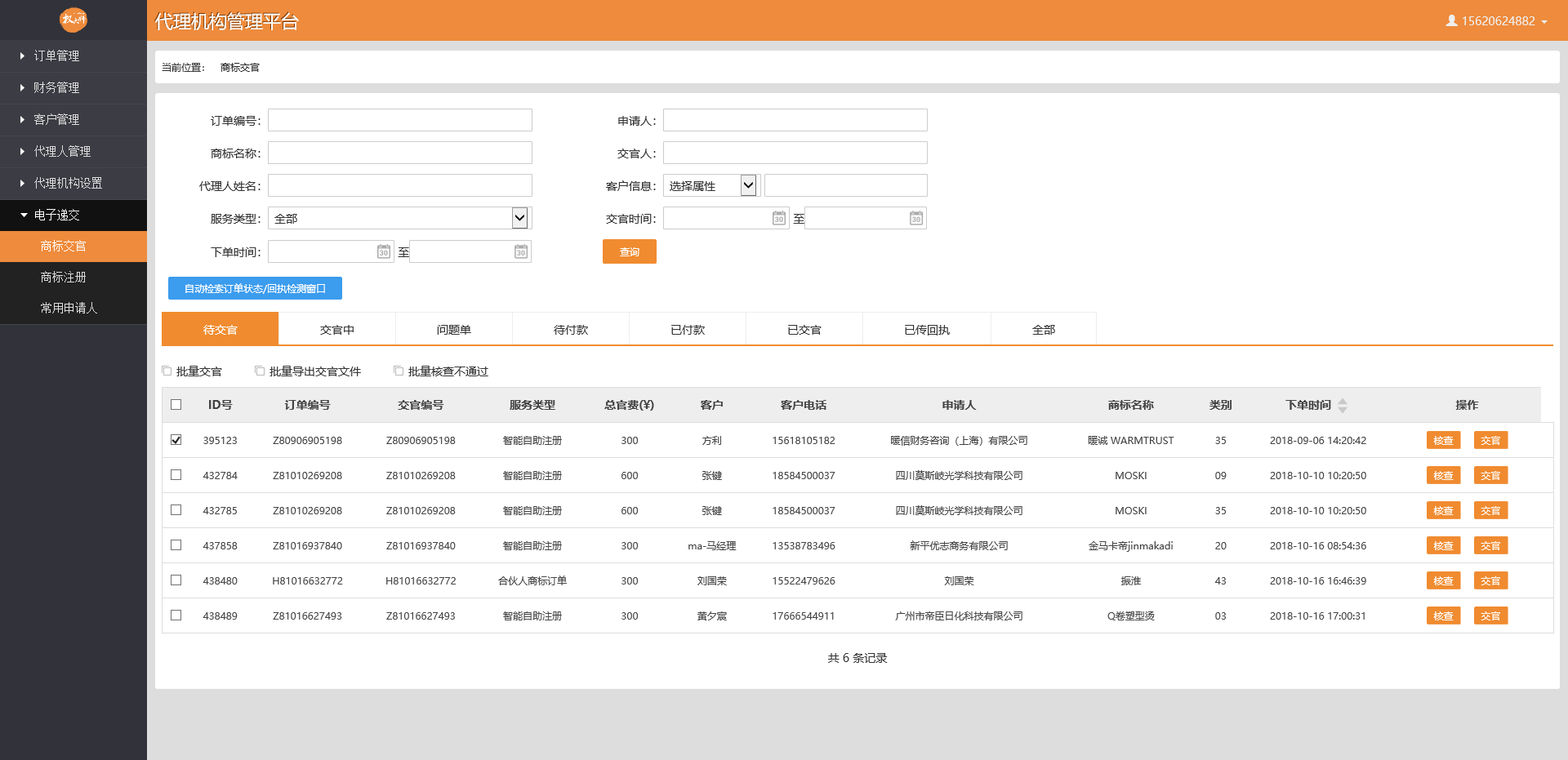


图1：代理机构管理平台

**功能：**在权大师代理机构管理平台做商标交官时，提供自动递交、批量递交功能，省去大量的人工填写时间。

**程序结构：**主要由主程序，守护进程和浏览器插件三部分组成。

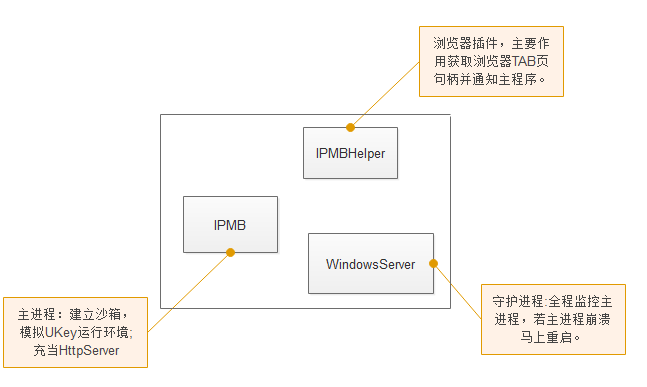


图2：程序组成

IPMBHelper:浏览器插件：用到BHO程序，即IE浏览器拓展程序，用到了COM技术。在这个项目里主要作用是获取管理平台所在TAB页的句柄，并通知到主程序，以便主程序界面的调整。

主程序相当于搭建了一个IE内核浏览器并导航到商标局官网交官页面。因为UKey只能在IE环境运行，所以只能使用IE的接口，用到的技术是WebBrowerControl。同时，为了让管理机构后台和商标局官网页面通信，主程序的这层壳还需要充当HttpServer的角色。

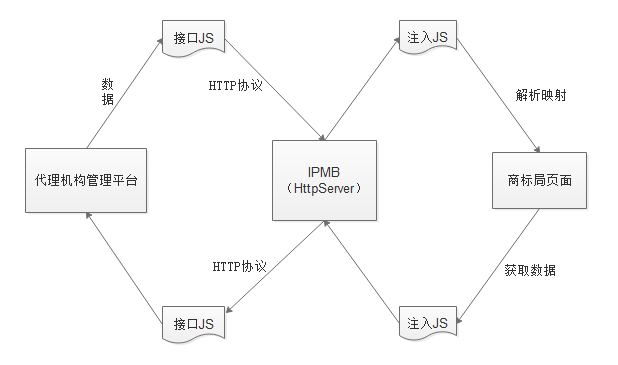


图3：管理平台和商标局通信示意图

管理平台和IPMB之间的通信：使用的是HTTP 1.1的协议，socket套接字。平台使用ajax向HttpServer发送数据。

IPMB和商标局页面之间的通信：WebBrowerControl提供了接口。主要是注入实现解析数据和自动填写逻辑的JS代码。

图4：注入的JS文件

如图所示，大部分的自动递交逻辑都在这些JS文件中，主程序获取JS内容并注入到商标局提交页。



图5：注入示意图

当HttpServer接收到管理平台发来的数据，注入JS到商标局提交页，转发数据给提交页面，就实现了自动将数据填入相应的HTML元素内。

### 爬虫项目

**功能：**从商标局的商标综合查询页爬取商标信息。



图5：爬虫界面

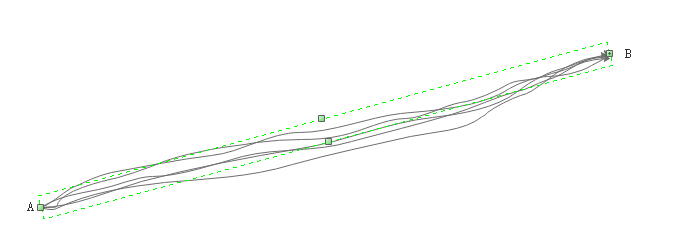
**攻防机制：**

1. User-Agent 用户代理，表明访问源身份的一种方式。整理一些常见的浏览器代理头，每次访问随机使用其中一个就好了。
2. 针对IP的，可以使用代理IP

3、防盗链：按正常人访问网页的顺序，一步步跳转到指定页。

4、控制访问次数。

5、行为分析：计算机程序对鼠标的移动是直线移动，爬虫则是采集人类移动鼠标的轨迹数据并模拟。



图：人类移动鼠标轨迹图

**功能模块：**

主程序：使用MFC框架，嵌入CEF和DuiLib控件，实现一个谷歌内核的浏览器。类似之前的电子递交，自己实现的浏览器方便注入和读取数据。

模拟鼠标轨迹：从已经的采集的人类使用鼠标时的轨迹坐标中，随机选取一部分模拟出一条从A到B的轨迹线。

读取和保存：使用脚本启动程序，指定商标的申请号、超时、Agent、是否显示鼠标轨迹、保存文件名等。

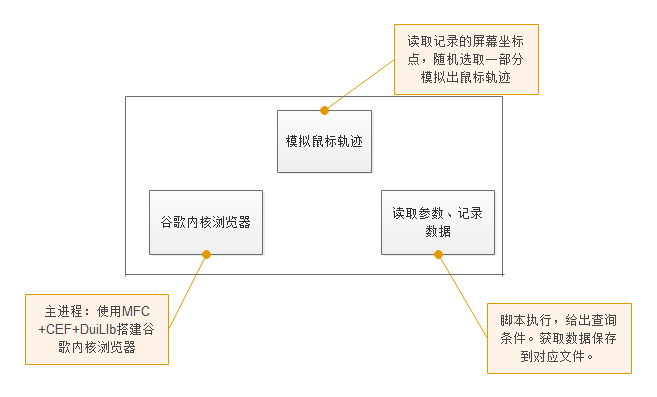
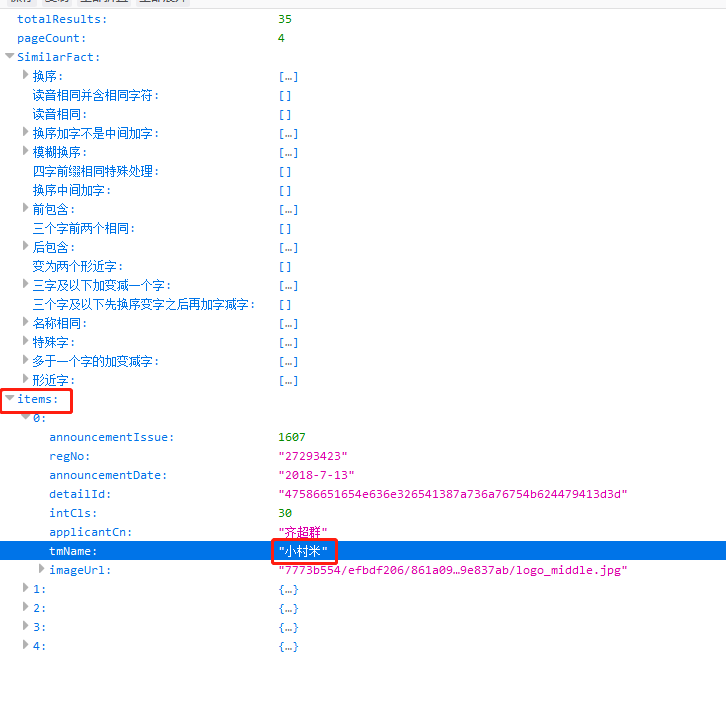
****

图6：爬虫功能模块

### 三、商标近似查询项目

功能：目的为商标代理部提供服务。客户申请的商标需要查询某些商标类别下与其近似的商标。目前查询的方法是根据条件手动拼接URL，然后访问。URL如下格式，需要手动替换tmName（商标名），pageNo（起始查询页），pageSize（总查询页），announcelssue（商标局公告期数）。

http://115.182.99.149:9999/api/searchForAnnounce?tmName=abc&pageNo=3&pageSize=10&announceIssue=1607



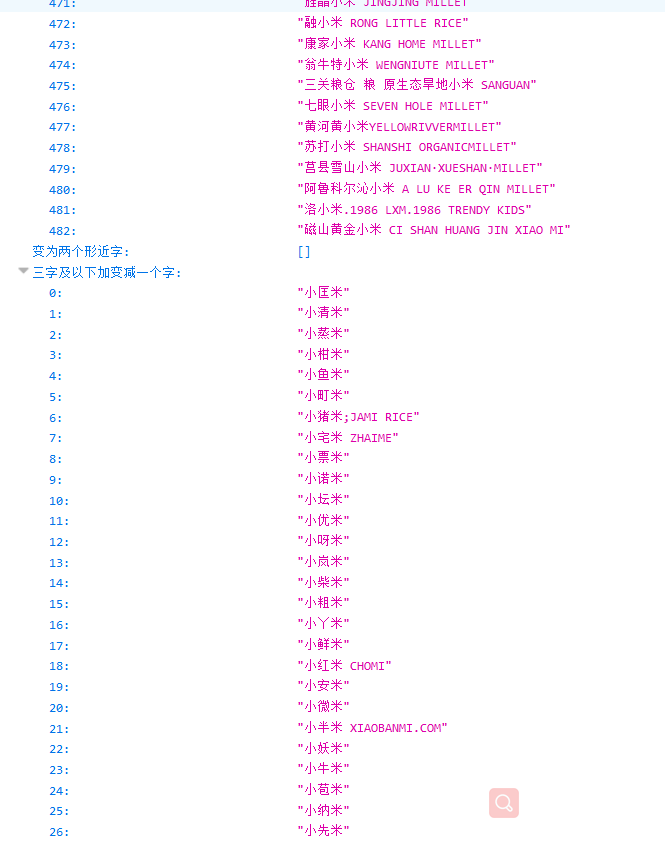


图7：网页查询结果示意图

在items数组里保存着所有的近似商标，需要先从关心的商标类别里筛选一部分，然后用tmName到上面的SimilarFact对象里查找对应的匹配条件。

程序设计的目的是将这些繁琐的操作实现自动化。设计流程图如下：

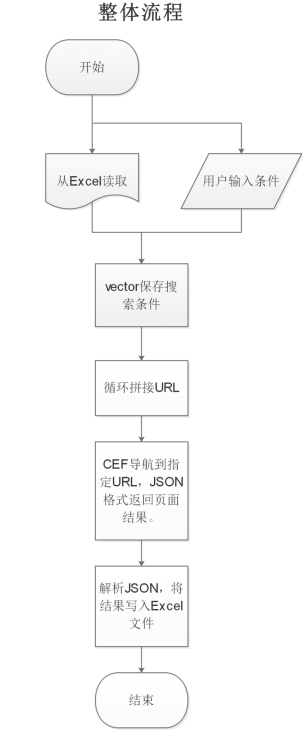


图8：流程图

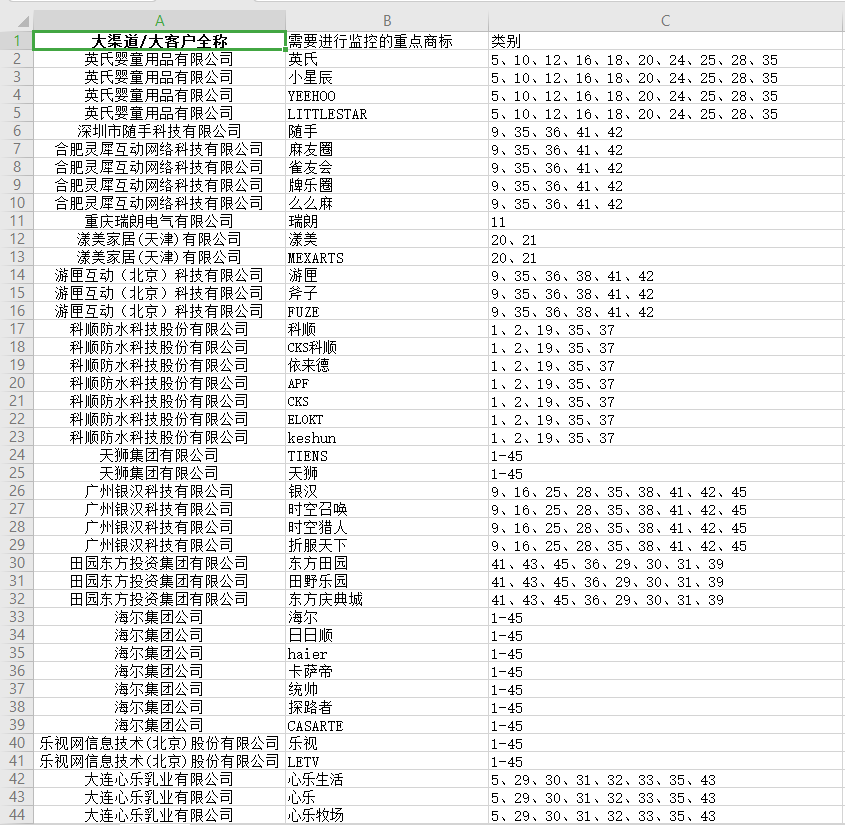
+-

图9：搜索条件



图10：保存的结果

项目仍然采用MFC+CEF+DuiLIb搭建。

查询条件支持用户手动输入，也支持从Excel文件中自动读取。对Excel的操作采用COM技术。

加入多线程，每条线程完成拼接URL、命令CEF新建Frame导航指定URL、返回JSON结果、解析JSON、按格式自动写入Excel。线程同步采用信号量的PV操作以及临界区。