

HiTortoise

我只是个毕业后沦为搬砖的屌丝,谈不了内涵的言论,玩不了文艺的调调,也搞不了高可用的架构,只是在茶前饭后动动手指,敲几个字来消遣罢了。

- 首页
- 联系我

自动加载下一页数据采集chrome浏览器插件制作

Author: [IceCry](#) | July 7, 2017 | Cate: [笔记](#)

最近在数据采集过程中遇到aspx __doPostBack数据加载方式，利用火车头(免费版)或php采集程序无有效方法采集，于是手动编写chrome浏览器插件自动点击“下一页”翻页抓取网址数据，继而用火车头采集数据。


扩展程序

☒ 开发者模式

加载已解压的扩展程序...

打包扩展程序...

立即更新扩展程序



数据采集插件 1.0.1

<http://blog.hitortoise.com>


[详细信息](#) [重新加载 \(Ctrl+R\)](#)

ID : hhfokkbpcmjakgcpemnhcnjpdifcaagm

加载来源 : ~\Desktop\crx

☐ 在隐身模式下启用

☒ 已启用



manifest.json

```
{
  "name": "数据采集",
  "description": "自动点击下一页采集数据",
  "version": "2.0",
  "permissions": [
    "activeTab"
  ],
  "background": {
    "scripts": ["bg.js"],
    "default_popup": "background.html",
    "persistent": true
  },
  "browser_action": {
    "default_title": "SHUJUCAIJI"
  },
  "content_scripts": [
    {
      "matches": ["http://*/*"],
      "js": ["jq.js", "popup.js"],
      "run_at": "document_idle"
    }
  ],
}
```

```
popup.js
//吉林律师
$("#DataGrid1 tr:last td span").next().html("<li>下一页</li>");
//吉林律所

// 开始获取数据

var html = $("#DataGrid1 tr td a:contains('详细信息')");
var arr = [];

for (var i = html.length - 1; i >= 0; i--) {
    arr[i] = html[i].href;
    //获取吉林律师id
    // arr[i] = arr[i].replace("http://www.lawyer-home.com/ls/zcxx/lsjtxx.aspx?id=", '');
    arr[i] = arr[i].replace("http://www.lawyer-home.com/ls/zcxx/lssjtxx.aspx?id=", '');
    //获取吉林律所id
};

$.ajax({
    url: 'http://localhost/gather/public/Index/Jilin/get_num/',
    type: 'post',
    dataType: 'json',
    data: { 'arr': arr },
    success: function(res){
        if(res.status == 200){
            //加载下一页
            //吉林律师
            $("#DataGrid1 tr:last td li").trigger('click');
            //吉林律所
        }
    }
});
```

php端接收数据写入txt

```
public function get_num($arr){
    $str = implode(',', $arr);
    // echo $str;die;

    file_put_contents('jilin.office.txt', $str.PHP_EOL, FILE_APPEND);
    // file_put_contents('jilin.lawyer.txt', $str.PHP_EOL, FILE_APPEND);
    // var_dump($arr);
    return json(['status'=>200, 'message'=>'success']);
}
```

可以完善查询“下一页”文本适应更多页面。

基础版采集下载：[数据采集插件基础演示.rar](#)

Tags: [插件](#)

仅有一条评论

IceCry

July 21st, 2017 at 02:07 pm

可使用contains('下一页')提高通用性

回复

添加新评论

称呼 *

内容 *

提交评论

上一篇: [二维数组中取某一相同字段的值进行,拼接字符串](#)

下一篇: [dedecms织梦手机网站添加上一页/下一页的翻页功能](#)