

wdj_crawler_Encapsulation使用说明

背景

爬取[豌豆荚网站](#) [软件分类](#)页面 或者 [游戏分类](#)页面 每一个单独分类的软件图标，爬取软件数量可以自定。

demo

```
if __name__ == '__main__':
    get_app_1 = GetAppid(all_sort_params = {'系统工具' : [5018,0]}, ctoken =
    'QYOVWfSCAluGwfUufI22IC7-', need_app_num = 20, save_path = r'D:\vs code\vs code
    file\test')
    app_ids = get_app_1.get_appid()

    get_page_1 = GetPage(real_data_appids = app_ids , doc_name='系统工具',
    save_path = r'D:\vs code\vs code file\test' )
    page = get_page_1.get_app_page()
    ## 本程序最终输出即为需要爬取的图片文件
```

使用说明

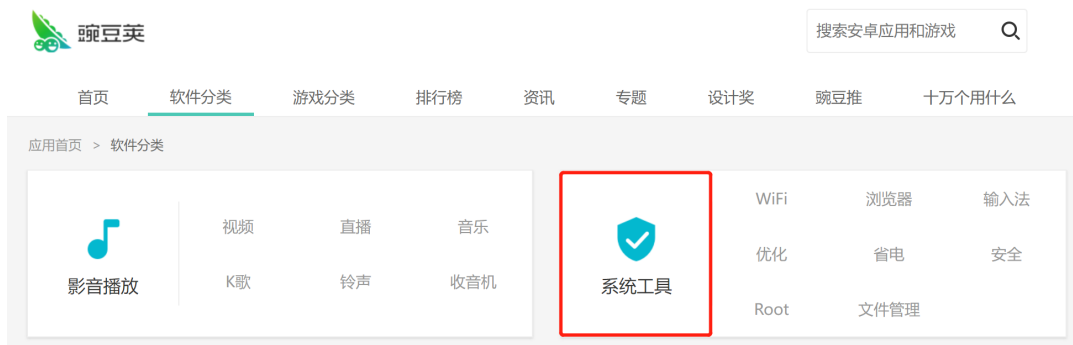
该文件内有三个类，使用者只需要调用 **GetAppid** 和 **Getpage** 这两个类即可获取一个分类的图标。

GetAppid使用

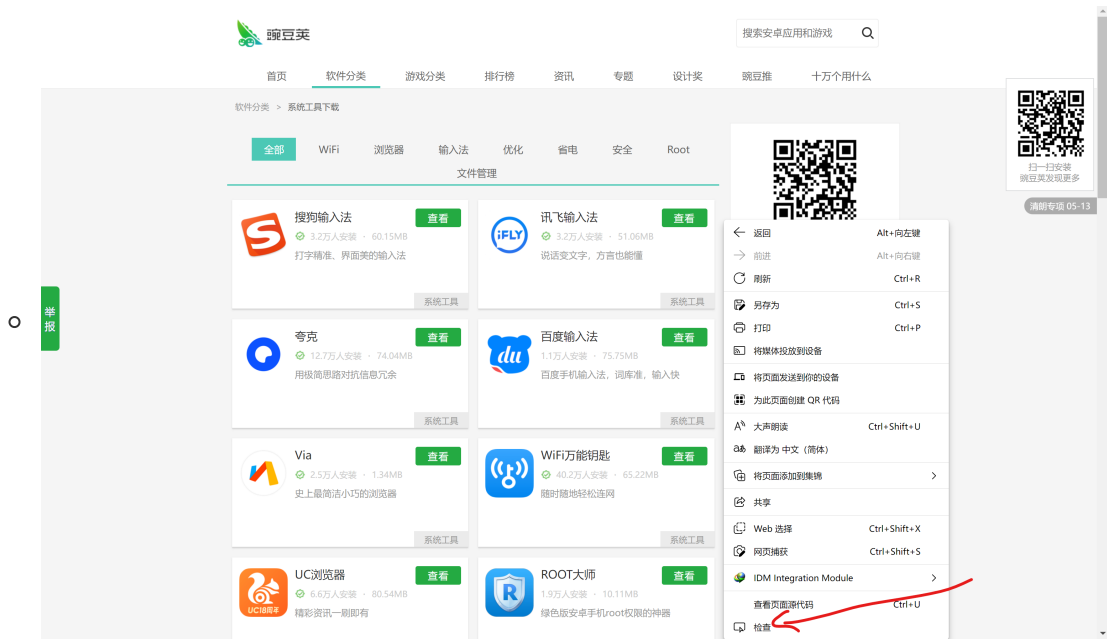
1.GetAppid参数

`all_sort_params` :

- **传参形式：** { '分类名称' : ['catId' , 'subcatId'] }
- **参数说明：**
 - 1.分类名称即为需要爬取的分类名称
 - 2.catId 和 subcatId 是每一个分类自己唯一的参数
- **catId 和 subcatId 获取参数方式如下**
 - 1.此处以爬取系统工具为例，参数中的分类名称可以自定义



2. 右击页面空白处，点击检查

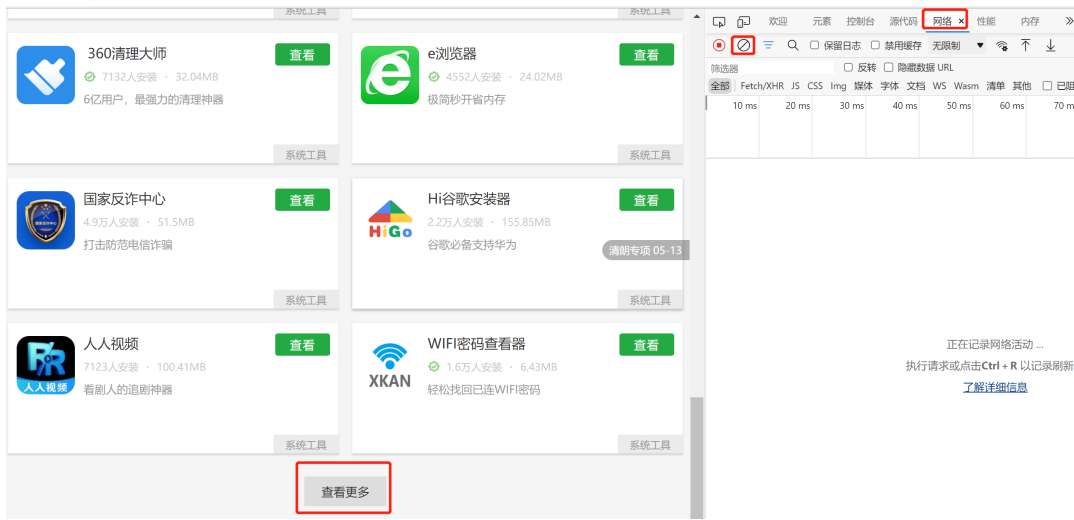


3. 点击检查后

1. 将主页面下拉到查看更多位置
2. 选择网络
3. 并点击禁止标志清空该页面



4. 点击查看更多，右侧会刷新出新的文件包，双击右侧网络中以more开头的文件包



5. 双击该文件包后，我们会看到这个页面，其中就有catId和 subcatId 和 ctoken



ctoken：

- 传参形式：string
- 参数说明：ctoken获取方法和上面一样（注意：豌豆荚的ctoken每天都会换，需要及时更换）

need_app_num：

- 传参形式：int
- 参数说明：需要获取的软件数量

save_path：

- 传参形式：string
- 参数说明：将文件保存在电脑什么位置（路径）

2.GetAppid输出

豌豆荚中每一个软件都有自己唯一的appid，GetAppid返回一个包含所有需要爬取的软件的appid列表。

GetPage使用

1.GetPage参数

`real_data_ppids:`

- **传参形式:** `list`
- **参数说明:** 将GetAppid返回的列表赋给这个参数即可

`doc_name:`

- **传参形式:** `string`
- **参数说明:** `doc_name` 的名字务必和 GetAppid中的all_sort_params中的分类名称**务必一模一样**

`save_path:`

- **传参形式:** `string`
- **参数说明:** 和GetAppid中的save_path**务必一模一样**

2.GetPage输出

GetPage输出即为需要爬取的图片文件