

Python 实验报告

实验八：爬虫与发布版本

姓名：余泓锸 学号：2016326603046

实验目的

了解编写爬虫的相关python库及其注意事项
了解各个网站对爬虫的态度和要求
掌握基本爬虫编写方法及实际可行
了解并测试项目发布Debug和Release版本的区别

实验内容

实验题1：

1. 请你写一个爬虫，能够爬取百度图片，搜索python后显示的图片的网址
解答：

In [8]:

```
import os
a = os.system('pip install --upgrade pip')
b = os.system('pip install requests')
c = os.system('pip install faker')
if not a and not b and not c:
    print('库安装成功')
```

库安装成功

In [2]:

```
import requests
import os
from faker import Faker

def imgUrls(keyWord, userAgent, pn=0, rn=30):
    url = 'https://image.baidu.com/search/acjson'
    params = {
        'tn': 'resultjson_com',
        'ipn': 'rj',
        'ct': '201326592',
        'is': '',
        'fp': 'result',
        'queryWord': keyWord,
        'cl': 2,
        'lm': -1,
        'ie': 'utf-8',
        'oe': 'utf-8',
        'adpicid': '',
        'st': -1,
        'z': 0,
        'ic': 0,
        'hd': 0,
        'latest': 0,
        'copyright': 0,
        'word': keyWord,
        's': '',
        'se': '',
        'tab': '',
        'width': '',
        'height': '',
        'face': 0,
        'istype': 2,
        'qc': '',
        'nc': 1,
        'fr': '',
        'expermode': '',
        'selected_tags': '',
        'cg': 'girl',
        'pn': pn,
        'rn': rn,
        'gsm': 'le',
        '1543417707210': ''
    }
    rep = requests.get(url, headers={'User-Agent': userAgent}, params=params)
    if(int(rep.status_code) == 200):
        try:
            imgs = rep.json()
            """解析返回的url地址"""
            def decodeUrl(imgUrl):
                longDic={'_z2C$q': ":", '_z&e3B': ".", 'AzdH3F': "/" }
                mapDic={'w': "a", 'k': "b", 'v': "c", 'l': "d", 'j': "e", 'u': "f", '2': "g", 'i': "h", 't': "i"}
                for k in longDic:
                    imgUrl=imgUrl.replace(k, longDic[k])
                imgUrl=list(imgUrl)
                tmp=[]
                for i in imgUrl:
                    if i in mapDic:
                        tmp.append(mapDic[i])
                    else:
                        tmp.append(i)
                return ''.join(tmp)
            for img in imgs['data']:
                img['url'] = decodeUrl(img['url'])
                print(img['url'])
```

```

        tmp.append(i)
    return ''.join(tmp)
imgUrls = [decodeUrl(imgs['data'][sec]['objURL'])
            for sec in range(len(imgs['data']) - 1)]
result = imgUrls
status = True
except Exception as err:
    result = str(err)
    status = False
finally:
    return {'result': result, 'status': status}

if __name__ == '__main__':
    fake = Faker(locale='zh-CN')
    keyword = input('input a key:')
    dirpath = os.getcwd() + '\\img'
    if not os.path.isdir(dirpath):
        os.mkdir(dirpath)
    keypath = dirpath + '\\ ' + keyword
    if not os.path.isdir(keypath):
        os.mkdir(keypath)
    index = 0
    pn = 0
    rn = 30
    print(dirpath)
    while True:
        print(pn, rn)
        urls = imgUrls(keyWord=keyword, userAgent=fake.user_agent(), pn=pn, rn=rn)
        if urls['status']:
            for url in urls['result']:
                print("Downloading:", url)
                try:
                    res = requests.get(url)
                    if str(res.status_code)[0] == "4":
                        print("未下载成功: ", url)
                        continue
                except Exception as e:
                    print("未下载成功: ", url)
                    continue
                filename = os.path.join(dirpath, keyword + str(index) + ".jpg")
                with open(filename, 'wb') as f:
                    f.write(res.content)
                index += 1
    print('当前已下载{0}图片{1}张'.format(keyword, index))
    ch = input('是否继续下载 (y/n) ? ')
    if ch[0] == 'y':
        pn += 30
    else:
        break

```

input a key:python

C:\Users\herry\python_kj\10. Spider\img

0 30

Downloading: <http://img.cwq.com/201611/581c95c35ca62.png> (<http://img.cwq.com/201611/581c95c35ca62.png>)

Downloading: http://pic.51yuansu.com/pic3/cover/01/77/36/59663d7b05b59_610.jpg (http://pic.51yuansu.com/pic3/cover/01/77/36/59663d7b05b59_610.jpg)

Downloading: http://upload-images.jianshu.io/upload_images/13692432-e5659988f3c7d6f

f.jpg (http://upload-images.jianshu.io/upload_images/13692432-e5659988f3c7d6ff.jpg)
Downloading: <http://img.51test.net/uploadfile/images/2018/0411/15234194104601835.jpg>
(<http://img.51test.net/uploadfile/images/2018/0411/15234194104601835.jpg>)
Downloading: <http://editerupload.eepw.com.cn/201411/87051416902987.jpg> (<http://editerupload.eepw.com.cn/201411/87051416902987.jpg>)
Downloading: <http://insights.thoughtworkers.org/wp-content/uploads/2015/07/718-%E4%B D%9F%E8%BE-%E5%85%A8%E6%A0%88%E5%BC%80%E5%8F%91%E8%AF%AD%E8%A8%80Python.png> (<http://insights.thoughtworkers.org/wp-content/uploads/2015/07/718-%E4%BD%9F%E8%BE-%E5%85%A8%E6%A0%88%E5%BC%80%E5%8F%91%E8%AF%AD%E8%A8%80Python.png>)
Downloading: <http://s3.sinaimg.cn/mw690/0070ps9Lzy7iwJXwL8md2&690> (<http://s3.sinaimg.cn/mw690/0070ps9Lzy7iwJXwL8md2&690>)
Downloading: http://s4.51cto.com/oss/201810/28/b05a1f587833f4983f373bec574682b1.jpg-wh_651x-s_3181838570.jpg (http://s4.51cto.com/oss/201810/28/b05a1f587833f4983f373bec574682b1.jpg-wh_651x-s_3181838570.jpg)
Downloading: <http://p99.pstatp.com/large/dfic-imagehandler/f707d7e6-b40a-4c19-bf20-2a0537d85c25> (<http://p99.pstatp.com/large/dfic-imagehandler/f707d7e6-b40a-4c19-bf20-2a0537d85c25>)
Downloading: <http://www.suzxms.com.cn/UploadFiles/2018-05-29/15275775467548780.png>
(<http://www.suzxms.com.cn/UploadFiles/2018-05-29/15275775467548780.png>)
Downloading: <http://il.51cto.com/images/201705/a83234480c656daff21659cc686eec972ae81b.jpg> (<http://il.51cto.com/images/201705/a83234480c656daff21659cc686eec972ae81b.jpg>)
Downloading: http://bb.tedu.cn/upload/20180904/20180904102251_724.png (http://bb.tedu.cn/upload/20180904/20180904102251_724.png)
Downloading: <http://ask.qcloudimg.com/http-save/developer-news/ezygmxcxd.jpeg> (<http://ask.qcloudimg.com/http-save/developer-news/ezygmxcxd.jpeg>)
Downloading: http://picl.zhimg.com/80/v2-2fe99d57ca95e51509a62b822b96b8da_qhd.jpg (http://picl.zhimg.com/80/v2-2fe99d57ca95e51509a62b822b96b8da_qhd.jpg)
Downloading: <http://5b0988e595225.cdn.sohucs.com/images/20180523/d411a933c9984f1a979fdf71ff2d1927.jpeg> (<http://5b0988e595225.cdn.sohucs.com/images/20180523/d411a933c9984f1a979fdf71ff2d1927.jpeg>)
Downloading: <http://5b0988e595225.cdn.sohucs.com/images/20180702/3e86a148fa684b4ca8ba94e0c9271d77.png> (<http://5b0988e595225.cdn.sohucs.com/images/20180702/3e86a148fa684b4ca8ba94e0c9271d77.png>)
Downloading: <http://ask.qcloudimg.com/http-save/developer-news/z64fnzredt.jpeg> (<http://ask.qcloudimg.com/http-save/developer-news/z64fnzredt.jpeg>)
Downloading: <http://www.djyyw.cn/upload/img/9N0IgYAfxd6w3hfPpivZI472LowS6LQE7Ht/Qj4aTk9aRJ66Y2d9ewDLe8-K9CT6K/ePFQ6Vnt9kU1x3zVlqVd125n3A4/s0AclaGg.jpg> (<http://www.djyyw.cn/upload/img/9N0IgYAfxd6w3hfPpivZI472LowS6LQE7Ht/Qj4aTk9aRJ66Y2d9ewDLe8-K9CT6K/ePFQ6Vnt9kU1x3zVlqVd125n3A4/s0AclaGg.jpg>)
Downloading: http://images.cnblogs.com/cnblogs_com/xuanye/WindowsLiveWriter/PythonZZ_F4E4/python_cheatsheet_Cn_r2_c1_2.jpg (http://images.cnblogs.com/cnblogs_com/xuanye/WindowsLiveWriter/PythonZZ_F4E4/python_cheatsheet_Cn_r2_c1_2.jpg)
Downloading: <http://n.sinaimg.cn/sinacn00/39/w546h293/20180821/cedb-hhxaafz5470697.png> (<http://n.sinaimg.cn/sinacn00/39/w546h293/20180821/cedb-hhxaafz5470697.png>)
Downloading: <http://www.ujiuye.com/uploadfile/2018/0712/20180712112500231.jpg> (<http://www.ujiuye.com/uploadfile/2018/0712/20180712112500231.jpg>)
Downloading: <http://dtimage.oss-cn-shanghai.aliyuncs.com/2016/1125/25025605sypp.jpg>
(<http://dtimage.oss-cn-shanghai.aliyuncs.com/2016/1125/25025605sypp.jpg>)
未下载成功: <http://dtimage.oss-cn-shanghai.aliyuncs.com/2016/1125/25025605sypp.jpg>
(<http://dtimage.oss-cn-shanghai.aliyuncs.com/2016/1125/25025605sypp.jpg>)
Downloading: http://pic2.zhimg.com/50/5c14c6b7a_hd.jpg (http://pic2.zhimg.com/50/5c14c6b7a_hd.jpg)
Downloading: http://image.codes51.com/Article/image/20160406/20160406234959_6093.png
(http://image.codes51.com/Article/image/20160406/20160406234959_6093.png)
未下载成功: http://image.codes51.com/Article/image/20160406/20160406234959_6093.png
(http://image.codes51.com/Article/image/20160406/20160406234959_6093.png)
Downloading: <http://aliyunzixunbucket.oss-cn-beijing.aliyuncs.com/png/20180316193055821114.png> (<http://aliyunzixunbucket.oss-cn-beijing.aliyuncs.com/png/20180316193055821114.png>)
Downloading: <http://jiuye-res.jikexueyuan.com/zhiye/showcase/attach-/20160824/25840a8e-72b5-4d38-ba9f-5bc8c5777ec2.jpg> (<http://jiuye-res.jikexueyuan.com/zhiye/showcase/>

attach-/20160824/25840a8e-72b5-4d38-ba9f-5bc8c5777ec2. jpg)
Downloading: http://ckimg.baidu.com/course/2017-08/09/0c768840b6ce9d82978ded501f3b590a.jpg (http://ckimg.baidu.com/course/2017-08/09/0c768840b6ce9d82978ded501f3b590a.jpg)
Downloading: http://www.tiejiang.org/wp-content/uploads/2014/08/1.jpg?imageView2/1/w/175/h/126/q/100 (http://www.tiejiang.org/wp-content/uploads/2014/08/1.jpg?imageView2/1/w/175/h/126/q/100)
Downloading: http://image.mamicode.com/info/201804/20180420171427621122.png (http://image.mamicode.com/info/201804/20180420171427621122.png)
Downloading: http://image20.it168.com/201801_670x502/3129/b5aa801fb9890b30.jpg (http://image20.it168.com/201801_670x502/3129/b5aa801fb9890b30.jpg)
当前已下载python图片28张
是否继续下载 (y/n) ? n

实验题2：

2. 项目发布Debug和Release

解答：

Debug版本是调试版本，其中会包含大量的调试文件，
而且为防止使源代码和生成的指令间关系更复杂导致调试过程复杂化，
一般不进行任何优化，便于程序员调试。

Release版本是发布版本，不对源代码进行调试，编译时对应用程序的速度进行优化，
从而保证程序在代码大小和运行速度上都是最优的。

值得一提的是，通过百度了解，Debug与Release编译方式的本质区别在于一组编译选项

Debug 版本：

/MDd /MLd 或 /MTd 使用 Debug runtime library(调试版本的运行时刻函数库)
/Od 关闭优化开关
/D "_DEBUG" 相当于 #define _DEBUG,打开编译调试代码开关(主要针对assert函数)
/ZI 创建 Edit and continue(编辑继续)数据库，这样在调试过程中如果修改了源代码不需重新编译
/GZ 可以帮助捕获内存错误
/Gm 打开最小化重链接开关，减少链接时间

Release 版本：

/MD /ML 或 /MT 使用发布版本的运行时刻函数库
/O1 或 /O2 优化开关，使程序最小或最快
/D "NDEBUG" 关闭条件编译调试代码开关(即不编译assert函数)
/GF 合并重复的字符串，并将字符串常量放到只读内存，防止被修改

因此实际上Debug与Release并没有本质的界限，其编译过程只是编译器通过预定的选项行动。

由于debug和release版的dll不能混合使用，因此实际生产环境中，尽量使用Release版本的DLL。

In []: