Python笔记（一）：安装+爬虫环境配置+打包为EXE文件

1、     安装

https://www.python.org/downloads/windows/  到官网下载安装程序

 Windows x86        32位操作系统

 Windows x86-64       64位操作系统

web-based installer     基于网络的安装程序

executable installer     可执行的安装程序

embeddable zip file     python的压缩包

上面3个任意下载一个就行了，都是一样的东西。我是下载第2个

help file  帮助文档

安装时记得勾上 ADD…to path。

安装成功后在命令提示符中输入 PYTHON，如图显示版本号就是安装成功，不是可执行命令的话，在系统变量path 中加上PYTHON的安装路径

PYTHON比较重要的几个环境变量（不是必须，根据需要配置）：

2、     爬虫环境配置

说明：Pip模块会自动安装，可以在命令提示符窗口输入PIP测试是否安装成功。（直接执行PIP报错的话，可以CD 到PYTHON安装目录的Scripts文件夹下执行，或者直接执行 python -m pip）

 确定pip安装好后，执行以下命令

1、pip install beautifulsoup4

2、到https://pypi.python.org/pypi/lxml/4.1.1    下载lxml库，记得下载适合自己操作系统和PYTHON版本的，比如我就是下载Windows64位+PYTHON3.6的

3、然后执行这个命令    pip install lxml的文件名及路径（例如：d:\xx.whl）

4、安装成功会提示 Successfully installed

3、     将脚本打包为EXE文件

（一）环境配置

1、 执行pip install pypiwin32

2、 执行 pip install pyinstaller

第二步如果安装不成功可以到http://www.pyinstaller.org/downloads.html下载安装包，然后执行

Python 解压后文件路径\setup.py install

（二）将源文件打包为EXE文件

pyinstaller -F -w 文件名及路径，成功后会提示存放位置。

-F：打包为单独的Exe文件 -w:不显示命令窗口

4、简单的爬虫程序和一个简单的GUI界面

import urllibimport refrom urllib import requestimport tkinter as tkdef getHtml(url): page = urllib.request.urlopen(url) html = page.read() return htmldef getImg(html): imgre = re.compile(r'src="(.+?\.jpg)"') html = html.decode('utf-8') # python3 imglist = re.findall(imgre,html) x = 0 for imgurl in imglist: urllib.request.urlretrieve(imgurl,'%s.jpg' % x) x+=1def start(url): html = getHtml(url) print(getImg(html))x = tk.Tk()label1 = tk.Label(x,text = "请输入网址：")label1.grid(row=0,column=0)label2 = tk.Label(x,text = "请输入文件存放路径：")label2.grid(row=1,column=0)var1 = tk.StringVar()entry1 = tk.Entry(x,textvariable=var1)entry1.grid(row=0,column=1)entry2 = tk.Entry(x)entry2.grid(row=1,column = 1)def seturl(): url="https://tieba.baidu.com/p/5475267611" # 本来想动态获取文本框输入的，不知道为什么WINDOWS10上运行返回值是NONE， # 另一台电脑WIN7却又可以，不知道是不是环境配置问题 # print(var1.get()) start(url)cbbtn1 = tk.Checkbutton(x,text = "同意协议")cbbtn1.grid(row = 2, column = 0)btn1 = tk.Button(x,text = "开始抓取",command = seturl)btn1.grid(row = 2,column = 2)btn2 = tk.Button(x,text = "取消")btn2.grid(row = 2,column = 3)#img = tk.PhotoImage(file = "C:\Users\123456\Pictures\lovewallpaper\11.jpg")#imgview = tk.Label(x,image = img)#imgview.grid(row = 0,column = 2)x.mainloop()





