***Json，lxml模块***

一.JSON模块

Json是一种网络中常用的数据交换类型，一个文件要想在网络进行传输，需要将文件转换为一种便于在网络之间传输的类型，便于人们进行阅读，json就是这样应运而生的。Json中的数据是由键值对构成的，与python中字典不同的是，json将数据转换为一种字符串的形式。

在电脑上如何安装json呢？

打开电脑的cmd，输入pip install json，然后在python命令行中运行 import json,如果没有出现什么错误，说明已经成功安装了。

Json中有许多模块，我目前在爬虫中用到的就两个方法，其他的

方法等碰见了再讲解。

   json.loads() #把json字符串转换为python类型

def loads(s,encoding=None, cls=None, object\_hook=None, parse\_float=None, parse\_int=None,parse\_constant=None, object\_pairs\_hook=None, \*\*kw):

 这是loads的源代码，可以参考一下。

>>json.loads('{"name":"Shara", "school":”MIT”}')

{'name': ‘Shara’, 'school': 'MIT'}

   json.dumps() #把python类型转换为json字符串类型

def dumps(obj, skipkeys=False, ensure\_ascii=True, check\_circular=True,

        allow\_nan=True, cls=None, indent=None, separators=None,

        default=None, sort\_keys=False, \*\*kw):

这是dumps的源代码。

>> dict ={'name': ‘Shara’, 'school': 'MIT'}

>>json.dumps(dict)

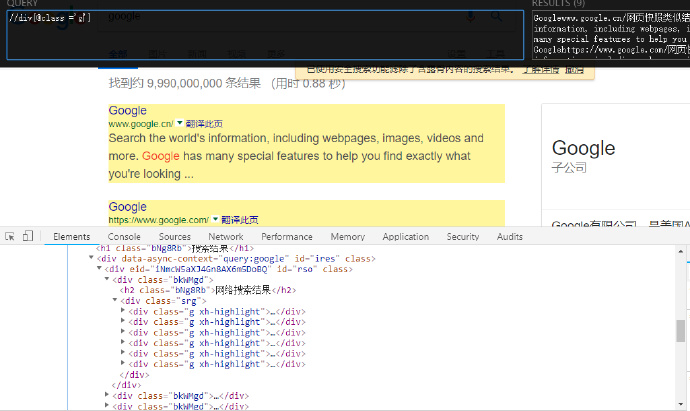
{"name":"Shara", "school":”MIT”}

二.lxml模块

   介绍这个模块之前，首先要讲一点有关xpath的内容。

   xpath，用于在html文件中查找某些内容的一种语法结构。html即之前讲过的服务器返回给浏览器的响应文件，或者说是xml文件，在网页“检查”中的Elements中也找召见。

具体用法：



输入//div[@class = ‘g’]  ,把html中所有的这个元素选中，图中颜色部分是我选中的部分，后面还可加一些提取的符号，比如@href，就是把href中的元素提取出来，/text(),是把指定的目录下的文本提取出来。

https://www.w3.org/TR/xpath20/

 这个网站中包含了所有有关xpath的相关应用,我就不做过多的翻译了，如果有人需要我可以下次找个时间进行个全文翻译之类的。

下面进入lxml环节。

安装，与前面许多包的安装是相同的：

pip install lxml

在python中使用时，通常写作

from lxml import etree

elements = etree.HTML(html字符串) #加入响应即可

elements.xpath() #这是就可以使用xpath模块了

将lxml重新起名为etree，国际通用惯例，记住就行。Lxml模块的作用过程是先获取一个html字符串，然后使用xpath来处理该字符串，使之成为一个elements对象，最后在对其数据进行整理，以便观察的一个流程。