一.单选题

1.下面这段代码在Python2下输出结果将是什么？

def div1(x,y):

print(“s%/s%=s%”%(x,y,x/y)

def div2(x,y):

print(“s%//s%=s%”%(x,y,x//y)

div1(5,2)

div1(5.,2)

div2(5,2)

div2(5.,2.)

A.

5/2=2.5

5.0/2=2.5

5//2=2

5.0//2.0=2.0

B.

5/2=2.5

5.0/2=2.5

5//2=2.5

5.0//2.0=2.5

C.

5/2=2

5.0/2=2.5

5//2=2.5

5.0//2.0=2.0

D.

5/2=2

5.0/2=2.5

5//2=2

5.0//2.0=2.5

#A

2.表达式[1]\*2的值为

A.[1] B.[1] C.[1,1] D.[1,2]

#C

3.已知列表x=[1,2,3],那么执行语句x.insert(1,4),x的值为

A.[1,2,3,4] B.[1,4,2,3] C.[4,3,2,1] D.[4,2,3]

#

B

4.已知字符串x=”hello world”,那么执行语句x.replace(‘hello’,’hi’)之后，x的值为

A.”hello world” B.”hillo world” C.”hi world” D.”world”

#

C

5.正则表达式元字符（）用来表示该符号前面的字符或子模块1次或多次出现

A.\* B.None C.% D.defg

#

A

6.语句print(re.match(‘abc’,’defg’))输出结果

A.abc B.None C.abcg D.defg

#

B

二、判断题

1.Python 3.x完全兼容python 2.x #错

2.在Windows平台上编写的Python程序无法在Unix平台运行。 #错

3.不可以在同一台计算机上安装多个Python版本 #错

4.3+4j不是合法的Python表达式 #错

5.Python集合可以包含相同的元素 #错

6.在UTF-8编码中一个汉字需要占用3个字节 #错

7.Python3.x支持使用中文作为变量名 #对

8.函数中必须包含return语句 #错

9.栈和队列都具有先入后出的特点 #错

10.定义类时实现了\_\_eq\_\_方法，该类对象即可支持运算符==. #对

三.XPath解析题

<?xml version=”1.0”encoding=’ISO-8859-1”?>

<bookstore>

<book>

<title lang=”eng”>Harry Potter<title>

<price>29.99</price>

<author>赵六</author>

</book>

<book>

<title lang=”eng”>Learning XML</title>

<price>39.95</price>

<author>张三</author>

</book>

<book>

<title lang=”eng”>ORACLE”</title>

<price>40.32</price>

<author>Lary</author>

</book>

</bookstore>

写出表达式：

1.选取bookstore元素中的book元素的所有title元素，且其中的price元素的值须大于35.0

2.选取属于bookstore元素的后代的所有book元素，而不管它们位于bookstore之下的什么位置

3.选择属于bookstore元素的后代的所有book元素，而不管它们位于bookstore之下的什么位置

4.选取所有带有属性的title元素

四.简答题

1.简述爬虫软件需要具备的功能

**一，获取整个页面数据**

**二，筛选页面中想要的数据**

**三，将页面筛选的数据保存到本地**

2.网页爬取和数据处理的步骤，所碰到的异常情况如何处理

对selenium做异常处理

在使用selenium的过程中我们经常遇到对网页操作的报错，常常程序会停留在这一步，没有办法继续下去，这时候我们就需要对程序做一个异常处理

处理的最终目标是达到，程序在遇到错误的时候，能够继续运行下去，但是同时也会在log日志文件中进行记录。

处理方法如下：

**对于save函数，其内部有使用selenium进行网页内容获取的部分**

# python代码

def save(baseUrl):

driver = webdriver.PhantomJS()

try:

driver.get(baseUrl) # seconds

# 为了进行渲染时间的控制，下面这两种设置都进行才有效

driver.set\_page\_load\_timeout(10)

driver.set\_script\_timeout(10)

# 获取页面数据

data = driver.page\_source

driver.quit() # 对驱动进行关闭

return data # 如果正常，会返回正常的网页内容，作为返回值

except Exception as e:

# 当出现异常的时候，会打开log.txt文件在文件最后添加一行“错误url”

with open('./data/log.txt', 'a') as f\_result:

f\_result.writelines(baseUrl)

return '' # 返回值为 ‘’空字符串，即返回空网页

* 1
* 2
* 3
* 4
* 5
* 6
* 7
* 8
* 9
* 10
* 11
* 12
* 13
* 14
* 15
* 16
* 17
* 18

在以上的代码中实现的就是如何做selenium的简单的异常处理了，只要是在try中某一句发生了错误那么就会跳到except里面执行错误处理的部分。try中错误了的代码的后面是不运行的