《Python 网络爬虫程序技术》

测试 3 参考答案

题目 1

```
def spider(url):
    #获取新的地址 newUrl
        if newUrl:
            spider(newUrl)
```

下面说法正确的是：

A.

找不到 newUrl 时会结束递归调用

(正确答案)

B.

不是递归调用

C.

一定会出现死循环

D.

找不到 newUrl 时也不会结束递归调用

题目 2

深度优先爬取说法正确的是

A.

结果与递归调用爬取一样

(正确答案)

B.

结果与递归调用爬取不一样

C.

效率比函数递归调用爬取低

D.

效率比函数递归调用爬取高

题目 3

广度优先爬取数据，说法正确的是：

A.
爬取数据的顺序与深度优先的不同
(正确答案)
B.
爬取数据的顺序与深度优先的相同
C.
爬取数据的顺序与函数递归方法相同
D.
都不对

题目 4

有一个 dowbload(url)函数下载 url 图像:

```
import threading
def download(url):
    pass
```

用多线程调用它，方法是：

A.
T=threading.Thread(target=download,args=[url])
T.start()
(正确答案)
B.
T=threading.Thread(target=download,args=url)
T.start()
C.
T=threading.Thread(target=download,args=(url))
T.start()
D.
都不对

题目 5

爬取网站的很多图片时，说法正确是：

A.
使用多线程效率高，程序复杂
(正确答案)
B.
使用单线程效率高，程序简单
C.
使用单线程效率高，程序复杂
D.

使用多线程效率高，程序简单

题目 6

url="http://www.weather.com.cn/weather/101280601.shtml"
headers={"User-Agent":"Mozilla/5.0 (Windows; U; Windows NT 6.0 x64; en-US; rv:1.9pre) Gecko/2008072421 Minefield/3.0.2pre"}
req=urllib.request.Request(url,headers=headers)
data=urllib.request.urlopen(req)
data=data.read()
其中 headers 的作用是为了模拟浏览器

A. 正确(正确答案)
B. 错误

题目 7

soup.select("body [class] a") 查找<body>下面所有具有 class 属性的节点下面的<a>节点；

A. 正确(正确答案)
B. 错误

题目 8

soup.select("body [class] ") 查找<body>下面所有具有 class 属性的节点；

A. 正确(正确答案)
B. 错误

题目 9

soup.select("body head title") 查找<body>下面<head>下面的<title>节点；

A. 正确(正确答案)
B. 错误

题目 10

soup.select("a[id='link1']") 查找属性 id="link1"的<a>节点；

A. 正确(正确答案)
B. 错误