《Python 网络爬虫程序技术》

期末考试 参考答案

题目 1

1
import urllib.request
resp=urllib.request.urlopen("http://127.0.0.1:5000")
_____
print(data)
获取网站的二进制数据，缺少的语句是：

A.
data=resp.read()
(正确答案)
B.
data=resp.get()
C.
data=resp.getBinary()
D.
data=resp.readBinary()

题目 2

import urllib.request
resp=urllib.request.urlopen("http://127.0.0.1:5000")
_____
print(html)
获取网站的 HTML 文本数据，缺少的语句是：

A.
html=resp.read().decode()
(正确答案)
B.
html=resp.read().encode()
C.
html=resp.read()
D.
html=resp.get()

题目 3

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 4

广度优先爬取数据，说法正确的是：

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

题目 5

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 6

selector.xpath("//bookstore/book") 搜索<bookstore>下一级的<book>元素，找到 2 个；

A. 正确(正确答案)
B. 错误

题目 7

selector.xpath("//body/book") 搜索<body>下一级的<book>元素，结果为空；

A. 正确(正确答案)
B. 错误

题目 8

selector.xpath("//body//book") 搜索<body>下<book>元素，找到 2 个；

A. 正确(正确答案)
B. 错误

题目 9

selector.xpath("/body//book") 搜索文档下一级的<body>下的<book>元素，找结果为空，因为文档的下一级是<html>元素，不是<body>元素；

A. 正确(正确答案)
B. 错误

题目 10

selector.xpath("//book//price")与 selector.xpath("//price")结果一样，都是找到 2 个<price>元素；

A. 正确(正确答案)
B. 错误