|  |
| --- |
| **爬虫16期结业考试** |

考生：

提交考试时间：**2024-11-18 20:03:39**

一 成绩

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 试题总数：25题 | 试题满分：100分 | 通过分数：80分 | 得分： | 考试结果：已通过 |

二 答卷

1(单选题)常用于对称加密的算法是？

A、 RSA

B、 AES

C、 MD5

D、 SHA-256

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

2(单选题)

如何通过Elements面板修改网页中的文本内容？

A、 在Elements面板中直接双击文本进行编辑

B、 在Console中使用document.editText(element, newText)命令

C、 在Sources面板中找到文本所在的文件进行修改

D、 在Network面板中找到文本所在的请求进行修改

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： A |
|
| 正确答案为： A |
| 解析：略 |

3(单选题)验证码的作用是？

A、 加速网站加载速度

B、 防止机器人自动提交表单

C、 提高搜索引擎爬虫的速度

D、 防止用户登录

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

4(单选题) AST遍历是指什么？

A、 在树上寻找叶子节点

B、 逐层查找根节点

C、 遍历整个树结构的每个节点

D、 仅查找深层节点

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： C |
|
| 正确答案为： C |
| 解析：略 |

5(单选题)

如何通过JavaScript读取和修改Cookie？

A、 使用document.cookie属性

B、 使用localStorage对象

C、 使用cookie()函数

D、 使用readCookie()和writeCookie()函数

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： A |
|
| 正确答案为： A |
| 解析：略 |

6(单选题)在Node.js环境中，全局对象是什么？

A、 document

B、 window

C、 global

D、 root

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： C |
|
| 正确答案为： C |
| 解析：略 |

7(单选题)

非对称加密算法相对于对称加密算法的主要优势是什么？

A、 加解密速度快

B、 密钥管理更简单

C、 适用于大量数据加密

D、 安全性更高

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： D |
|
| 正确答案为： D |
| 解析：略 |

8(单选题)User-Agent检测：什么是User-Agent检测？

A、 检测用户的身份编号

B、 检测用户浏览器的标识

C、 检测用户的物理位置

D、 检测用户的登录状态

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

9(单选题)以下哪个方法可以通过Selenium执行JavaScript代码？

A、 execute\_script()

B、 send\_keys()

C、 find\_element\_by\_xpath()

D、 get\_attribute()

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： A |
|
| 正确答案为： A |
| 解析：略 |

10(单选题)

Ajax的主要特点是什么？

A、 同步请求和响应

B、 异步请求和响应

C、 只支持XML格式的数据传输

D、 只能在浏览器端使用

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

11(单选题)什么是IP封锁？

A、 禁止用户使用特定的IP地址

B、 加密用户IP地址

C、 隐藏用户的真实IP地址

D、 提高网络速度

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： A |
|
| 正确答案为： A |
| 解析：略 |

12(单选题)

在JavaScript逆向中，使用哪些工具或技术可以帮助定位加密数据的位置？

A、 使用关键字进行定位

B、 使用hook脚步进行定位

C、 使用代理工具拦截网络请求

D、 所有上述

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： D |
|
| 正确答案为： D |
| 解析：略 |

13(单选题)为什么有些网站会使用反爬虫机制,其中错误的是？

A、 保护数据安全

B、 提高搜索引擎排名

C、 防止非法数据采集

D、 减少服务器负载

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

14(单选题)

在Web浏览器中，访问一个使用TLS保护的网站时，URL的协议前缀是什么？

A、 HTTP

B、 HTTPS

C、 TCP

D、 SSL

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

15(单选题) 在RPC中，通信的两端通常分别被称为什么？

A、 客户端和服务器端

B、 发送者和接收者

C、 主机和从机

D、 控制器和执行器

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： A |
|
| 正确答案为： A |
| 解析：略 |

16(单选题) 在JavaScript中，常见的加密算法包括哪些？

A、 MD5

B、 SHA-256

C、 AES

D、 所有上述

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： D |
|
| 正确答案为： D |
| 解析：略 |

17(单选题) 以下哪种JavaScript混淆技术用于将可读的代码变成难以理解的形式？

A、 Minification

B、 Obfuscation

C、 Compression

D、 Encryption

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

18(单选题) 以下不属于HTTP请求方法的是

A、GET

B、DELETE

C、PAT

D、POST

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： C |
|
| 正确答案为： C |
| 解析：略 |

19(单选题)robots.txt 是用于什么目的的文件？

A、 配置网站的主页

B、 限制搜索引擎爬虫访问的页面

C、 存储网站的静态资源

D、 用于网站用户登录验证

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： B |
|
| 正确答案为： B |
| 解析：略 |

20(单选题) 在Python中，用于解析HTML文档的库是什么？

A、 requests

B、 json

C、 beautifulsoup

D、 selenium

|  |
| --- |
| 本题得分: 2分 满分2分 |
|
| 学员回答为： C |
|
| 正确答案为： C |
| 解析：略 |

21(问答题)使用多线程采集本站点数据

地址：https://adworld.xctf.org.cn/contest/list

字段：标题、时间、主办方

交付：提交各功能核心截图

|  |
| --- |
| 本题得分: 12分 满分12分 |
|
| 学员回答为： import requests from concurrent.futures import ThreadPoolExecutor, as\_completed headers = { "Accept": "application/json, text/plain, \*/\*", "Accept-Language": "zh-CN", "Cache-Control": "no-cache", "Connection": "keep-alive", "Pragma": "no-cache", "Sec-Fetch-Dest": "empty", "Sec-Fetch-Mode": "cors", "Sec-Fetch-Site": "same-origin", "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/127.0.0.0 Safari/537.36", "sec-ch-ua": "\"Not)A;Brand\";v=\"99\", \"Google Chrome\";v=\"127\", \"Chromium\";v=\"127\"", "sec-ch-ua-mobile": "?0", "sec-ch-ua-platform": "\"Windows\"" } url = "https://adworld.xctf.org.cn/api/event/release\_event/list/" def get\_info(page): params = { "page": f"{page}", "page\_size": "20", "search": "", "isSearch": "false", "event\_status": "", "event\_type": "" } response = requests.get(url, headers=headers, params=params) data = response.json() rows = data.get('data', {}).get('rows', []) items = {} for row in rows: title = row.get('release\_name') time = row.get('competition\_start\_time') host = row.get('release\_sponsor') if title: items[title] = [time, host] return items # print(get\_info(1)) if \_\_name\_\_ == '\_\_main\_\_': with ThreadPoolExecutor(max\_workers=5) as pool: futures = [pool.submit(get\_info, page) for page in range(5)] # 并发返回: as\_completed会返回先完成的结果 for future in as\_completed(futures): # 生成器 print(future.result()) |
|
| 正确答案为： w |
| 解析：略 |

22(问答题)采集本站首页数据top10

地址：https://www.chinaindex.net/ranklist/5

字段：字段不限、存储使用mongodb

交付：提供数据和代码截图即可说明：禁止使用自动化等技术

|  |
| --- |
| 本题得分: 12分 满分12分 |
|
| 学员回答为： |
|
| 正确答案为： 无 |
| 解析：略 |

23(问答题)公共服务数据采集，打开站点采集数据100条、要求对标题进行hash处理去重

地址:https://www.ynjzjgcx.com/dataPub/enterprise

字段：企业名称  信用代码 企业注册地

交付：提交各功能核心截图

|  |
| --- |
| 本题得分: 12分 满分12分 |
|
| 学员回答为： |
|
| 正确答案为： w |
| 解析：略 |

24(问答题)使用网络爬虫技术采集该站点10000条数据，且需要设计data去重算法，使用多线程采集 ,字段不限，mongo存储 ,

采集地址：https://www.oklink.com/zh-hans/btc/tx-list

交付：提供数据和代码截图即可。说明：禁止使用自动化等技术

|  |
| --- |
| 本题得分: 12分 满分12分 |
|
| 学员回答为： |
|
| 正确答案为： 无 |
| 解析：略 |

25(问答题)描述：使用代码模拟网站登陆、需要解决验证码等验证问题

地址：https://www.fcbox.com/pages/user/login.html

要求：需要协议请求、禁用自动化

|  |
| --- |
| 本题得分: 12分 满分12分 |
|
| 学员回答为： |
|
| 正确答案为： 无 |
| 解析：略 |