

Reading Web Data From Python

TOTAL POINTS 13

1. Which of the following Python data structures is most similar to the value returned in this line of Python:

1 point

```
1 x = urllib.request.urlopen('http://data.pr4e.org/romeo.txt')
```

- ☐ socket
- ☒ file handle
- ☐ list
- ☐ regular expression
- ☐ dictionary

2. In this Python code, which line actually reads the data?

1 point

```
1 import socket
2
3 mysock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
4 mysock.connect(('data.pr4e.org', 80))
5 cmd = 'GET http://data.pr4e.org/romeo.txt HTTP/1.0\n\n'.encode()
6 mysock.send(cmd)
7
8 while True:
9     data = mysock.recv(512)
10    if (len(data) < 1):
11        break
12    print(data.decode())
13 mysock.close()
```

- ☒ mysock.recv()
- ☐ socket.socket()
- ☐ mysock.close()
- ☐ mysock.connect()
- ☐ mysock.send()

3. Which of the following regular expressions would extract the URL from this line of

1 point

HTML:

```
1 <p>Please click <a href="http://www.dr-chuck.com">here</a></p>
```

- ☒ href="(.+)"
- ☐ href=".+"
- ☐ http://.*
- ☐ <.*>

4. In this Python code, which line is most like the open() call to read a file:

1 point

```
1 import socket
2
3 mysock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
4 mysock.connect(('data.pr4e.org', 80))
5 cmd = 'GET http://data.pr4e.org/romeo.txt HTTP/1.0\n\n'.encode()
6 mysock.send(cmd)
7
8 while True:
9     data = mysock.recv(512)
10    if (len(data) < 1):
11        break
12    print(data.decode())
13 mysock.close()
```

- ☒ mysock.connect()
- ☐ import socket
- ☐ mysock.recv()
- ☐ mysock.send()
- ☐ socket.socket()

5. Which HTTP header tells the browser the kind of document that is being returned?

1 point

- ☐ ETag:
- ☒ Content-Type:
- ☐ Metadata:
- ☐ HTML-Document:

☐ Document-Type:

6. What should you check before scraping a web site?

1 point

- ☐ That the web site returns HTML for all pages
- ☐ That the web site only has links within the same site
- ☐ That the web site supports the HTTP GET command
- ☒ That the web site allows scraping

7. What is the purpose of the BeautifulSoup Python library?

1 point

- ☐ It allows a web site to choose an attractive skin
- ☐ It optimizes files that are retrieved many times
- ☒ It repairs and parses HTML to make it easier for a program to understand
- ☐ It animates web operations to make them more attractive
- ☐ It builds word clouds from web pages

8. What ends up in the "x" variable in the following code:

1 point

```
1 html = urllib.request.urlopen(url).read()
2 soup = BeautifulSoup(html, 'html.parser')
3 x = soup('a')
```

- ☒ A list of all the anchor tags (<a..) in the HTML from the URL
- ☐ True if there were any anchor tags in the HTML from the URL
- ☐ All of the externally linked CSS files in the HTML from the URL
- ☐ All of the paragraphs of the HTML from the URL

9. What is the most common Unicode encoding when moving data between systems?

1 point

- ☐ UTF-128

- ☐ UTF-64
- ☐ UTF-16
- ☐ UTF-32
- ☒ UTF-8

10. What is the ASCII character that is associated with the decimal value 42?

1 point

- ☐ !
- ☐ +
- ☐ /
- ☒ *
- ☐ ^

11. What word does the following sequence of numbers represent in ASCII:

1 point

108, 105, 110, 101

- ☐ func
- ☐ lost
- ☒ line
- ☐ ping
- ☐ tree

12. How are strings stored internally in Python 3?

1 point

- ☒ Unicode
- ☐ UTF-8
- ☐ ASCII

- ☐ EBCDIC
- ☐ Byte Code

13. When reading data across the network (i.e. from a URL) in Python 3, what method must be used to convert it to the internal format used by strings?

1 point

- ☐ find()
- ☐ encode()
- ☐ trim()
- ☐ upper()
- ☒ decode()



I, **JAY PATEL**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

[Learn more about Coursera's Honor Code](#)