**CH-12 WEB SCRAPING**

PRACTICE QUESTIONS

1. Briefly describe the differences between the webbrowser, requests, bs4, and selenium modules?

Ans:

Webbrowser- It opens a webpage in our browser.

Requests- Downloads Files or Pages from webs.

Bs4- Parses HTML[Structure of web page]

Selenium- Simulates actions like clicking or typing on a webpage.

1. What type of object is returned by requests.get()? How can you access the downloaded content as a string value?

Ans:

By calling type() on requests.get() , it returns a Response object which contains the response that the web server gave to our request. We can access the downloaded content as a string value using the .text() function.

Eg:- res=requests.get(‘<https://examples.com>’)

As\_string=res.text()

print(As\_string)

1. What requests method checks that the download worked?

Ans:

The status\_code is the request method that checks if the download worked or not.

Eg:- res=requests.get(‘<https://examples.com>’)

If (res.status\_code=200):

print(“Download success”)

else:

print(“Download failed”)

1. How can you get the HTTP status code of a requests response?

Ans:

We can get the HTTP status code by using the raise\_for\_status()

Eg:-res=requests.get(‘https://examples.com’)

res.raise\_for\_status()

1. How do you save a requests response to a file?

Ans:

To save a new file use write mode(wb) using open(). Then use a loop with res.iter\_content() to read file in chunk.

Eg:-

with open(‘file.txt’,’wb’) as file:

for chunk in res.iter\_content(1000000):

file.write(chunk)

1. What is the keyboard shortcut for opening a browser’s developer tools?

Ans:

Chrome: Press F12 or CTRL+SHIFT+C.

Firefox: Press CTRL+SHIFT+C.

Safari: Enable "Develop" in Preferences, then press OPTION+I.

1. How can you view (in the developer tools) the HTML of a specific element on a web page?

Ans:

We can right click on a webpage and select ‘View source’ or ‘View Page Source’ to shows the structure of webpage.

For viewing a specific element on webpage we can use inspect on right click and also use the developer tools if needed.

1. What is the CSS selector string that would find the element with an id attribute of main?

Ans:-

We can use soup.select(‘#attribute\_name’) for this.

Eg:- soup.select(‘#main’)

1. What is the CSS selector string that would find the elements with a CSS class of highlight?

Ans:-

We can use soup.select(‘.attribute\_name’) for this.

Eg:- soup.select(‘.highlight’)

1. What is the CSS selector string that would find all the <div> elements inside another <div> element?

Ans:-

We can use soup.select() for this.

Eg:-soup.select(‘div div’)

1. What is the CSS selector string that would find the <button> element with a value attribute set to favorite?

Ans:-

Eg:-soup.select(‘button[value=”favourite”]’)

1. Say you have a Beautiful Soup Tag object stored in the variable spam for the element <div>Hello, world!</div>. How could you get a string 'Hello, world!' from the Tag object?

Ans:-

spam=soup.div

text=spam.string

print(text)

1. How would you store all the attributes of a Beautiful Soup Tag object in a variable named linkElem?

Ans:-

linkElem=soup.a

attributes=linkElem.attrs

print(attributes)

1. Running import selenium doesn’t work. How do you properly import the selenium module?

Ans:-

from selenium import webdriver

driver=webdriver.Chrome()

driver.get(‘https://www.example.com’)

print(driver.title)

driver.quit()

1. What’s the difference between the find\_element\_\* and find\_elements\_\* methods?

Ans:-

find\_element\_\* returns a single element whereas find\_elements\_\* returns a list of elements. If no elements match fin\_element\_\* will raise NoSuchElementException whereas find\_elements\_\* returns an empty list.

1. What methods do Selenium’s WebElement objects have for simulating mouse clicks and keyboard keys?

Ans:-

click(): This method is used to simulate a mouse click on the element.

send\_keys(\*value): This method is used to simulate typing text or pressing keys inside a text input or text area element.

1. You could call send\_keys(Keys.ENTER) on the Submit button’s WebElement object, but what is an easier way to submit a form with selenium?

Ans:-

form = driver.find\_element\_by\_id('my\_form')

form.submit()

1. How can you simulate clicking a browser’s Forward, Back, and Refresh buttons with selenium?

Ans:-

browser.forward()- Clicks the forward button

browser.back()- Clicks the back button

browser.refresh()- Clicks the refresh/ reload button

**CH-13 WORKING WITH EXCEL SHEETS**

1. What does the openpyxl.load\_workbook() function return?

Ans:-

It returns a workbook object.

1. What does the wb.sheetnames workbook attribute contain?

Ans:-

It contains the list of all sheet names.

1. How would you retrieve the Worksheet object for a sheet named 'Sheet1'?

Ans:-

wb=openpyxl.load\_workbook(‘example.xlsx’)

sheet=wb[‘Sheet1’]

1. How would you retrieve the Worksheet object for the workbook’s active sheet?

Ans:-

active\_sheet=wb.active

1. How would you retrieve the value in the cell C5?

Ans:-

cell=sheet[‘C5’]

print(cell.value)

1. How would you set the value in the cell C5 to "Hello"?

Ans:-

sheet[“C5”]=”Hello”

1. How would you retrieve the cell’s row and column as integers?

Ans:-

print(cell.row)

x=cell.column

print(column\_index\_from\_string(x))

1. What do the sheet.max\_column and sheet.max\_row sheet attributes hold, and what is the data type of these attributes?

Ans:-

sheet.max\_column hold the maximum number of columns in the spreadsheet and sheet.max\_row hold the maximum number of rows in spreadsheet.

The type of these are int.

1. If you needed to get the integer index for column 'M', what function would you need to call?

Ans:-

We need to call the function get\_index\_from\_string(‘M’) to get the integer index for ‘M’.

1. If you needed to get the string name for column 14, what function would you need to call?

Ans:-

We need to call the function get\_column\_letter(14) to get the string name for column 14.

1. How can you retrieve a tuple of all the Cell objects from A1 to F1?

Ans:-

wb=openpyxl.load\_workbook(‘example.xlsx’)

sheet=wb.active

cells=tuple(sheet.iter\_rows(min\_row=1,max\_row=1,min\_col=1,max\_col=6)

for cell in cells:

print(cell)

1. How would you save the workbook to the filename *example.xlsx*?

Ans:-

wb.save(‘example.xlsx’)

1. How do you set a formula in a cell?

Ans:-

sheet[‘A1’]= ’=SUM(A2:A3)’

1. If you want to retrieve the result of a cell’s formula instead of the cell’s formula itself, what must you do first?

Ans:-

sheet[‘A1’]= ’=SUM(A2:A3)’

print(A1.value)

1. How would you set the height of row 5 to 100?

Ans:-

sheet.row\_dimensions[5].height=100

1. How would you hide column C?

Ans:-

sheet.column\_dimensions[‘C’].hidden=True

1. What is a freeze pane?

Ans:-

Freeze pane allows us to see the cell visible even when we scroll the excel sheet.

1. What five functions and methods do you have to call to create a bar chart?

Ans:-

a)openpyxl.Workbook()

b)openpyxl.chart.Reference()

c)openpyxl.chart.BarChart()

d)chart.add\_data()

e)sheet.add\_chart()

**CH 14- WORKING WITH GOOGLE SPREADSHEETS**

1. What three files do you need for EZSheets to access Google Sheets?

Ans:-

The three files which are required for EZSheets to access Google Sheets are

a)credentials.json

b)token-sheets.pickle

c)token-drive.pickle

2. What two types of objects does EZSheets have?

Ans:-

Spreadsheet Objects and Sheet Objects are the two types of objects that EZSheets have.

1. How can you create an Excel file from a Google Sheet spreadsheet?

Ans:-

import ezsheets

ss=ezsheets.Spreadsheet(‘Google Sheet ID or URL’)

ss.downloadAsExcel(‘Myspreasheet.xlsx’)

1. How can you create a Google Sheet spreadsheet from an Excel file?

Ans:-

import ezsheets

ss=ezsheets.upload(‘MyExcelFile.xlsx’)

print(ss.title)

1. The ss variable contains a Spreadsheet object. What code will read data from the cell B2 in a sheet titled “Students”?

Ans:-

sheet=ss[‘Students’]

print(sheet[‘B2’])

1. How can you find the column letters for column 999?

Ans:-

import ezsheets.

print(ezhsheets.getColumnLetterOf(999))

1. How can you find out how many rows and columns a sheet has?

Ans:-

columnCount() and rowCount() will give the number of rows and columns in a sheet.

Eg:-

import ezsheets

sheet=ss[‘Sheet1’]

print(sheet.rowCount)

print(sheet.columnCount)

1. How do you delete a spreadsheet? Is this deletion permanent?

Ans:-

ss.delete() will delete a spreadsheet temporarily by moving it to google drive trash.

ss.delete(permanent=True) permanently deletes a spreadsheet.

1. What functions will create a new Spreadsheet object and a new Sheet object, respectively?

Ans:-

import ezsheets

ss=ezsheets.createSpreadsheet(‘My New Spreadsheet’)

ss.createSheet(‘My New Sheet’)

1. What will happen if, by making frequent read and write requests with EZSheets, you exceed your Google account’s quota?

Ans:-

It will return the error googleapiclient.errors.HttpError: Quota exceeded for quota group. This error occurs when we exceed 100 requests per 100 seconds. EZSheets will automatically catch this error and retry to make the request. If the request continues to fail the exception may be re-raised.

**CH-15 PDF s**

1. A string value of the PDF filename is not passed to the PyPDF2.PdfFileReader() function. What do you pass to the function instead?

Ans:

A file object opened in binary format is passed to PdfFileReader() instead of the filename String

Eg:-

pdfFileObj=open(‘filename.pdf’ , ’rb’)

pdfReader=PyPDF2.PdfFileReader(pdfFileObj)

1. What modes do the File objects for PdfFileReader() and PdfFileWriter() need to be opened in?

Ans:

PdfFileReader() should be opened in read binary(rb) format whereas PdfFileWriter() should be opened in write binary(wb) format.

3. How do you acquire a Page object for page 5 from a PdfFileReader object?

Ans:

Page=pdfReader.getPage(4)

4. What PdfFileReader variable stores the number of pages in the PDF document?

Ans:

Total=pdfReader.numPages

5. If a PdfFileReader object’s PDF is encrypted with the password swordfish, what must you do before you can obtain Page objects from it?

Ans:

pdfReader.decrypt(“swordfish”)

1. What methods do you use to rotate a page?

Ans:

We use the method roatateClockwise(degree) to rotate the page clockwise and the method rotateCounterClockwise(degree) to rotate the page counter clockwise.

7. What method returns a Document object for a file named demo.docx?

Ans:

We use the docx.Document() method to rteurn a Document object.

Eg:-

import docx

doc=docx.Document(‘demo.docx’)

8. What is the difference between a Paragraph object and a Run object?

Ans:

A paragraph object represents a block of text within a document, ending with a paragraph text.

A Run object is contiguos segment of text within a pargraph that shares the same formatting.

1. How do you obtain a list of Paragraph objects for a Document object that’s stored in a variable named doc?

Ans:

Accesses the paragraphs attribute.

Eg:-

paragraphs=doc.paragraphs

10. What type of object has bold, underline, italic, strike, and outline variables?

Ans:

A Run object has bold, underline, italic ,strike and outline variables.

11. What is the difference between setting the bold variable to True, False, or None?

Ans:

Setting bold variable to:-

True: Applies bold formatting

False: Removes bold formatting

None: No change to the existing bold formatting

12. How do you create a Document object for a new Word document?

Ans:

We use the docx.Document() method without arguments.

Eg:-

doc=docx.Document()

13. How do you add a paragraph with the text 'Hello, there!' to a Document object stored in a variable named doc?

Ans:

We use the add\_paragraph() method

Eg:-

doc.add\_paragraph(“Hello, there!”)

1. What integers represent the levels of headings available in Word documents?

Ans:

The integers 0 to 9 represent the heading levels. Here 0 is the title style and 1 to 9 represents Heading 1 through Heading 9.

**CH-16 CSV AND JSON**

1. What are some features Excel spreadsheets have that CSV spread-sheets don’t?

Ans:

Formatting: Excel supports cell formatting (e.g., font, color, borders, etc.) while CSV does not.

Formulas and Functions: Excel can store and evaluate formulas, which are not supported in CSV.

Multiple Sheets: Excel allows multiple worksheets within a single file, whereas CSV is limited to a single table of data.

Data Validation: Excel supports data validation, dropdowns, and custom rules.

Charts and Graphs: Excel can include charts and graphs, which CSV cannot.

Macros: Excel supports automation through VBA macros; CSV does not.

Data Types: Excel can handle different data types (dates, currency, etc.), while CSV treats all data as plain text.

1. What do you pass to csv.reader() and csv.writer() to create reader and writer objects?

Ans:

· For csv.reader(): Pass a file object opened in text mode (e.g., open('filename.csv', 'r')).

· For csv.writer(): Pass a file object opened in write mode (e.g., open('filename.csv', 'w', newline='')).

1. What modes do File objects for reader and writer objects need to be opened in?

Ans:

Reader(‘r’) and Writer(‘w’)

1. What method takes a list argument and writes it to a CSV file?

Ans:

writerow() method

1. What do the delimiter and lineterminator keyword arguments do?

Ans:

· delimiter**:** Specifies the character used to separate fields in the CSV file (default is a comma ,).

· lineterminator**:** Specifies the character(s) used to terminate lines in the CSV file (default is '\r\n').

1. What function takes a string of JSON data and returns a Python data structure?

Ans:

json.loads()

1. What function takes a Python data structure and returns a string of JSON data?

Ans:

json.dumps()

**Ch-17 Keeping Time,Scheduling Tasks and Launching programs**

1. What is the Unix epoch?

Ans:

The Unix epoch is the starting point for time in Unix systems, defined as 00:00:00 Coordinated Universal Time (UTC) on January 1, 1970.

1. What function returns the number of seconds since the Unix epoch?

Ans:

time.time()

1. How can you pause your program for exactly 5 seconds?

Ans:

time.sleep(5)

1. What does the round() function return?

Ans:

round() returns a floating point number or an integer rounded to a specific number of decimal places.

1. What is the difference between a datetime object and a timedelta object?

Ans:

· A datetime object represents a specific **date and time** (e.g., January 7, 2019, 12:30 PM).

· A timedelta object represents a **duration** or difference between two points in time (e.g., 5 days, 3 hours).

1. Using the datetime module, what day of the week was January 7, 2019?

Ans:

import datetime

date = datetime.date(2019, 1, 7)

day\_of\_week = date.strftime("%A")

print(day\_of\_week)

# Output: "Monday"

1. Say you have a function named spam(). How can you call this function and run the code inside it in a separate thread?

Ans:

Using threading module.

import threading

thread = threading.Thread(target=spam)

thread.start()

1. What should you do to avoid concurrency issues with multiple threads?

Ans:

To avoid concurrency issues:

1. Use thread-safe data structures or libraries.
2. Protect shared resources with synchronization primitives like threading.Lock or threading.RLock.
3. Avoid accessing shared variables without proper synchronization.
4. Consider using the queue module for thread-safe queues.

**CH-18 Sending Email and Text Messages**

1. What is the protocol for sending email? For checking and receiving email?

Ans:

SMTP(Simple Mail Transfer Protocol) for sending emails. For checking and recieveing emails IMAP(Internet Message Access Protocol).

1. What four smtplib functions/methods must you call to log in to an SMTP server?

Ans:

· smtplib.SMTP() or smtplib.SMTP\_SSL() to create an SMTP connection.

· connect() (if not using SMTP\_SSL()) to connect to the SMTP server.

· ehlo() to identify your program to the server.

· login(username, password) to log in with your email credentials.

1. What two imapclient functions/methods must you call to log in to an IMAP server?

Ans:

IMAPClient() to create connection to the server.

Login(username,password) to authenticate with the server.

1. What kind of argument do you pass to imapObj.search()?

Ans:

We pass a String argument to imapObj.search().

1. What do you do if your code gets an error message that says got more than 10000 bytes?

Ans:

We use the fetch() method

imapObj.fetch(message\_ids, ['BODY.PEEK[HEADER]', 'BODY.PEEK[TEXT]'])

1. The imapclient module handles connecting to an IMAP server and finding emails. What is one module that handles reading the emails that imapclient collects?

Ans:

The pyzmail module is commonly used for parsing and reading emails collected by imapclient.

1. When using the Gmail API, what are the credentials.json and token.json files?

Ans:

· credentials.json**:** A file that contains the OAuth 2.0 client credentials from the Google Cloud Console.

· token.json**:** A file generated after the first authorization that stores the user's access and refresh tokens for subsequent API calls.

1. In the Gmail API, what’s the difference between “thread” and “message” objects?

Ans:

* **Thread:** Represents a collection of related messages grouped together (e.g., a conversation).
* **Message:** Represents an individual email within a thread.

1. Using ezgmail.search(), how can you find emails that have file attachments?

Ans:

has.attachment query is used

ezgmail.search('has:attachment')

1. What three pieces of information do you need from Twilio before you can send text messages?

Ans:

1. **Account SID**: Your Twilio account identifier.
2. **Auth Token**: A secret key for authenticating API requests.
3. **Twilio Phone Number**: The phone number provided by Twilio for sending messages.