Advance File Handling

Shutil, XLSX,CSV,ZIP FILE, PDF, and DOC

The shutil Module

- The shutil (or shell utilities) module has functions to let you copy, move, rename, and delete files in your Python programs. To use the shutil functions, you will first need to use import shutil.

Example

- >>> import shutil,os
- >>> os.chdir('C:\\sp')
- # To copy one file to another folder
- >>> shutil.copy('C:\\sp\\spam.txt', 'C:\\sp1')
- #Copy tree create a new folder
- >>> shutil.copytree('C:\\sp1', 'C:\\bacon_backup')
- 'C:\\bacon_backup'
- #Move to given Folder
- >>> shutil.move('C:\\sp1', 'C:\\bacon_backup')
- 'C:\\bacon_backup\\sp1'

Permanently Deleting Files and Folders

- Calling os.unlink(path) will delete the file at path.
- Calling os.rmdir(path) will delete the folder at path. This folder must be empty of any files or folders.
- Calling shutil.rmtree(path) will remove the folder at path, and all files and folders it contains will also be deleted.

Deleting File

```
>>> import send2trash
baconFile = open('spam1.txt', 'a') # creates the file
>>> baconFile.write('Bacon is not a vegetable.')
25
>>> baconFile.close()
>>> send2trash.send2trash('spam1.txt')
```

Reading ZIP Files

To create a ZipFile object, call the zipfile.ZipFile() function, passing it a string of the .zip file’s filename.

```
>>> import zipfile, os
>>> os.chdir('C:\\')    # move to the folder with example.zip
>>> exampleZip = zipfile.ZipFile('example.zip')
>>> exampleZip.namelist()
  ['spam.txt', 'cats/', 'cats/catnames.txt', 'cats/zophie.jpg']
```

Reading Zip file

```
>>> spamInfo = exampleZip.getinfo('spam.txt')
>>> spamInfo.file_size.
13908
>>> spamInfo.compress_size
3828
```

Excel Sheet

Creating, Reading and Writing XLSX

Excel Sheet

```
>>>import openpyxl
>>> wb = openpyxl.load_workbook('example.xlsx')
>>> wb.get_sheet_names()
>>> sheet = wb.get_sheet_by_name('Sheet3')
>>> sheet
```

Getting cell value

```
>>> sheet.cell(row=1, column=2).value
>>> import openpyxl
    >>> wb = openpyxl.load_workbook('example.xlsx')
>>> sheet = wb.get_sheet_by_name('Sheet1')
>>> sheet.max_row
7
>>> sheet.max_column
3
```

Excel Operation

```
>>> import openpyxl
>>> wb = openpyxl.Workbook()
>>> sheet = wb.active
>>> sheet['A1'] = 200
>>> sheet['A2'] = 200
>>> sheet['A3'] = '=SUM(A1:A2)'
>>> wb.save('writeFormula.xlsx')
>>>
```

PDF and DOC

Reading, Writing, Reading a Password Protected file, and Writing down in Document

PDF Reading

- Framework-- pyPDF2

```
>>> import PyPDF2
>>> pdfFileObj = open('demo.pdf', 'rb')
>>> pdfReader = PyPDF2.PdfFileReader(pdfFileObj)
>>> pdfReader.numPages
24
>>> pageObj = pdfReader.getPage(0)
>>>pageObj.extractText()
```

Read Password Protected File

- >>> import PyPDF2
- >>> pdfReader = PyPDF2.PdfFileReader(open('encrypted.pdf', 'rb'))
- >>> pdfReader.isEncrypted
  - True
- >>> pdfReader.getPage(0) # will get PDF Read Error
- >>> pdfReader.decrypt('Password')
- >>> pageObj = pdfReader.getPage(0)')

Copying a File

- >>> pdfFileObj = open('demo.pdf', 'rb')
- >>> pdf1Reader = PyPDF2.PdfFileReader(pdfFileObj)
- >>> pdfWriter = PyPDF2.PdfFileWriter()
- >>> for pageNum in range(pdf1Reader.numPages):
- 	pageObj = pdf1Reader.getPage(pageNum)
- 	pdfWriter.addPage(pageObj)
- >>> pdfOutputFile = open('combinedminutes.pdf', 'wb')
- >>> pdfWriter.write(pdfOutputFile)
- >>> pdfOutputFile.close()

Reading Word Document

- Framework-- docx

```
>>> doc = docx.Document('sample.docx')
>>> len(doc.paragraphs)
4
>>> doc.paragraphs[0].text
'Sample'
>>> doc.paragraphs[1].text
'Python is a Easy Language'
```

CSV FILE

Reading and writing CSV

CSV File

- >>> import csv
- >>> exampleFile = open('ratings.csv')
- >>> exampleReader = csv.reader(exampleFile)
- >>> exampleData = list(exampleReader)
- >>> exampleData
- [['userId', 'movieId', 'rating', 'timestamp'], ['1', '2', '3.5', '1112486027'], ['1', '29', '3.5', '1112484676'], ['1', '32', '3.5', '1112484819'], ['1', '47', '3.5', '1112484727'], ['1', '50', '3.5', '1112484580'], ['1', '112', '3.5', '1094785740'], ['1', '151', '4', '1094785734'], ['1', '223', '4', '1112485573'], ['1', '253', '4', '1112484940']]

Index and Slice for CSV

- >>> exampleData[0][0]
- 'userId'
- >>> exampleData[0][1]
- 'movieId'
- >>> exampleData[0][1:2]
- ['movieId']
- >>> exampleData[0][1:4]

Write Files

- >>> outputFile = open('output.csv', 'w', newline='')
- >>> outputWriter = csv.writer(outputFile)
- >>> outputWriter.writerow(['spam', 'eggs', 'bacon', 'ham'])
- 21
- >>> outputWriter.writerow(['Hello, world!', 'eggs', 'bacon', 'ham'])
- 32
- >>> outputWriter.writerow([1, 2, 3.141592, 4])
- 16
- >>> outputFile.close()