**1. In what modes should the PdfFileReader() and PdfFileWriter() File objects will be opened?**

In [3]:

**Ans: !**pip install PyPDF2

Collecting PyPDF2

Downloading PyPDF2-1.26.0.tar.gz (77 kB)

Building wheels for collected packages: PyPDF2

Building wheel for PyPDF2 (setup.py): started

Building wheel for PyPDF2 (setup.py): finished with status 'done'

Created wheel for PyPDF2: filename=PyPDF2-1.26.0-py3-none-any.whl size=61101 sha256=d4b9460c8c1185e893ff47c4ad1943daedb523050374ed0263b9db65dad70bb9

Stored in directory: c:\users\admin\appdata\local\pip\cache\wheels\b1\1a\8f\a4c34be976825a2f7948d0fa40907598d69834f8ab5889de11

Successfully built PyPDF2

Installing collected packages: PyPDF2

Successfully installed PyPDF2-1.26.0

In [4]:

**import** PyPDF2

The File object passed to PyPDF2.PdfFileReader() needs to be opened in read-binary mode by passing 'rb' as the second argument to open(). Likewise, the File object passed to PyPDF2.PdfFileWriter() needs to be opened in write-binary mode with 'wb'.

**2. From a PdfFileReader object, how do you get a Page object for page 5?**

In [5]:

**Ans:** mypdf **=** open("SLS\_Design\_study\_living\_hinge\_02.pdf","rb")

In [6]:

pdfReader **=** PyPDF2**.**PdfFileReader(mypdf)

PdfReadWarning: Xref table not zero-indexed. ID numbers for objects will be corrected. [pdf.py:1736]

In [7]:

page5 **=** pdfReader**.**getPage(4)

In [8]:

print(page5)

{'/CropBox': [27.61, 61, 567.61, 781], '/Parent': IndirectObject(1231, 0), '/StructParents': 4, '/Contents': IndirectObject(26, 0), '/Rotate': 90, '/MediaBox': [0, 0, 595.22, 842], '/Resources': {'/XObject': {'/Im30': IndirectObject(27, 0), '/Im31': IndirectObject(28, 0), '/Im32': IndirectObject(30, 0), '/Im33': IndirectObject(31, 0), '/Im34': IndirectObject(32, 0), '/Im35': IndirectObject(34, 0), '/Im36': IndirectObject(35, 0), '/Im37': IndirectObject(36, 0), '/Im38': IndirectObject(38, 0), '/Im39': IndirectObject(40, 0), '/Im40': IndirectObject(42, 0), '/Im41': IndirectObject(44, 0), '/Im42': IndirectObject(46, 0), '/Im43': IndirectObject(48, 0), '/Im44': IndirectObject(50, 0), '/Im45': IndirectObject(52, 0), '/Im46': IndirectObject(54, 0), '/Im47': IndirectObject(56, 0), '/Im48': IndirectObject(58, 0), '/Im49': IndirectObject(60, 0), '/Im50': IndirectObject(62, 0), '/Im51': IndirectObject(64, 0), '/Im52': IndirectObject(66, 0), '/Im53': IndirectObject(68, 0), '/Im54': IndirectObject(70, 0), '/Im55': IndirectObject(72, 0), '/Im56': IndirectObject(74, 0), '/Im57': IndirectObject(76, 0), '/Im10': IndirectObject(78, 0), '/Im11': IndirectObject(79, 0), '/Im12': IndirectObject(80, 0), '/Im13': IndirectObject(81, 0), '/Im14': IndirectObject(82, 0), '/Im15': IndirectObject(84, 0), '/Im16': IndirectObject(86, 0), '/Im17': IndirectObject(87, 0), '/Im18': IndirectObject(88, 0), '/Im19': IndirectObject(90, 0), '/Im0': IndirectObject(3621, 0), '/Im1': IndirectObject(91, 0), '/Im2': IndirectObject(92, 0), '/Im3': IndirectObject(956, 0), '/Im4': IndirectObject(94, 0), '/Im5': IndirectObject(96, 0), '/Im6': IndirectObject(98, 0), '/Im7': IndirectObject(99, 0), '/Im8': IndirectObject(100, 0), '/Im9': IndirectObject(102, 0), '/Im20': IndirectObject(104, 0), '/Im21': IndirectObject(106, 0), '/Im22': IndirectObject(108, 0), '/Im23': IndirectObject(960, 0), '/Im24': IndirectObject(110, 0), '/Im25': IndirectObject(112, 0), '/Im26': IndirectObject(114, 0), '/Im27': IndirectObject(116, 0), '/Im28': IndirectObject(118, 0), '/Im29': IndirectObject(119, 0)}, '/ColorSpace': {'/CS10': IndirectObject(963, 0), '/CS11': IndirectObject(964, 0), '/CS12': IndirectObject(965, 0), '/CS13': IndirectObject(966, 0), '/CS14': IndirectObject(967, 0), '/CS15': IndirectObject(968, 0), '/CS16': IndirectObject(969, 0), '/CS17': IndirectObject(929, 0), '/CS18': IndirectObject(930, 0), '/CS19': IndirectObject(931, 0), '/CS20': IndirectObject(932, 0), '/CS21': IndirectObject(933, 0), '/CS22': IndirectObject(934, 0), '/CS23': IndirectObject(935, 0), '/CS24': IndirectObject(936, 0), '/CS25': IndirectObject(937, 0), '/CS26': IndirectObject(938, 0), '/CS27': IndirectObject(939, 0), '/CS28': IndirectObject(940, 0), '/CS29': IndirectObject(941, 0), '/CS30': IndirectObject(942, 0), '/CS31': IndirectObject(943, 0), '/CS32': IndirectObject(944, 0), '/CS33': IndirectObject(945, 0), '/CS34': IndirectObject(946, 0), '/CS35': IndirectObject(947, 0), '/CS36': IndirectObject(948, 0), '/CS37': IndirectObject(949, 0), '/CS38': IndirectObject(950, 0), '/CS0': IndirectObject(3620, 0), '/CS1': IndirectObject(957, 0), '/CS2': IndirectObject(958, 0), '/CS3': IndirectObject(959, 0), '/CS4': IndirectObject(961, 0), '/CS5': IndirectObject(951, 0), '/CS6': IndirectObject(952, 0), '/CS7': IndirectObject(953, 0), '/CS8': IndirectObject(955, 0), '/CS9': IndirectObject(962, 0)}, '/Font': {'/TT0': IndirectObject(924, 0), '/TT1': IndirectObject(3612, 0), '/TT2': IndirectObject(921, 0), '/TT3': IndirectObject(3618, 0), '/C2\_0': IndirectObject(954, 0)}, '/ProcSet': ['/PDF', '/Text', '/ImageB', '/ImageC', '/ImageI'], '/Properties': {'/MC0': IndirectObject(3610, 0)}, '/ExtGState': {'/GS0': IndirectObject(3614, 0)}}, '/Type': '/Page'}

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

**Ans:** The total number of pages in the document is stored in the numPages attribute of a PdfFileReader object

In [10]:

pdfReader**.**numPages

Out[10]:

29

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

In [ ]:

**Ans: import** PyPDF2

pdfReader **=** PyPDF2**.**PdfFileReader(open('encrypted.pdf', 'rb'))

pdfReader**.**isEncrypted

True

In [ ]:

pdfReader**.**getPage(0)

In [ ]:

pdfReader **=** PyPDF2**.**PdfFileReader(open('encrypted.pdf', 'rb'))

pdfReader**.**decrypt('swordfish')

pageObj **=** pdfReader**.**getPage(0)

**5. What methods do you use to rotate a page?**

In [ ]:

**Ans:** minutesFile **=** open('meetingminutes.pdf', 'rb')

pdfReader **=** PyPDF2**.**PdfFileReader(minutesFile)

page **=** pdfReader**.**getPage(0)

page**.**rotateClockwise(90)

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

**Ans:** To Paragraph objects, character styles can be applied to Run objects, and linked styles can be applied to both kinds of objects. You can give both Paragraph and Run objects styles by setting their style attribute to a string. This string should be the name of a style. If style is set to None, then there will be no style associated with the Paragraph or Run object.

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

In [ ]:

**Ans:** paragraphObj**.**style **=** 'doc'

<https://automatetheboringstuff.com/2e/chapter15/>

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

**Ans:** Run type of object has bold, underline, italic, strike, and outline variables.

In [ ]:

**>>>** **import** docx

**>>>** doc **=** docx**.**Document('demo.docx')

**>>>** doc**.**paragraphs[0]**.**text

'Document Title'

**>>>** doc**.**paragraphs[0]**.**style *# The exact id may be different:*

\_ParagraphStyle('Title') id: 3095631007984

**>>>** doc**.**paragraphs[0]**.**style **=** 'Normal'

**>>>** doc**.**paragraphs[1]**.**text

'A plain paragraph with some bold and some italic'

**>>>** (doc**.**paragraphs[1]**.**runs[0]**.**text, doc**.**paragraphs[1]**.**runs[1]**.**text, doc**.**

paragraphs[1]**.**runs[2]**.**text, doc**.**paragraphs[1]**.**runs[3]**.**text)

('A plain paragraph with some ', 'bold', ' and some ', 'italic')

**>>>** doc**.**paragraphs[1]**.**runs[0]**.**style **=** 'QuoteChar'

**>>>** doc**.**paragraphs[1]**.**runs[1]**.**underline **=** **True**

**>>>** doc**.**paragraphs[1]**.**runs[3]**.**underline **=** **True**

**>>>** doc**.**save('restyled.docx')

**9. What is the difference between False, True, and None for the bold variable?**

**Ans:** True (the attribute is always enabled, no matter what other styles are applied to the run), False (the attribute is always disabled), or None (defaults to whatever the run’s style is set to).

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

In [2]:

**Ans: !**pip install docx

Collecting docx

Downloading docx-0.2.4.tar.gz (54 kB)

Requirement already satisfied: lxml in c:\users\admin\anaconda3\lib\site-packages (from docx) (4.6.3)

Requirement already satisfied: Pillow>=2.0 in c:\users\admin\anaconda3\lib\site-packages (from docx) (8.3.1)

Building wheels for collected packages: docx

Building wheel for docx (setup.py): started

Building wheel for docx (setup.py): finished with status 'done'

Created wheel for docx: filename=docx-0.2.4-py3-none-any.whl size=53924 sha256=6457e5faafa95fb5d1ec3b1d24fe0101c2c704c1759aae5c82814e467a5bc630

Stored in directory: c:\users\admin\appdata\local\pip\cache\wheels\78\ae\2e\5ce789557ca59281a463dabe9d6522e429be240322148e5379

Successfully built docx

Installing collected packages: docx

Successfully installed docx-0.2.4

In [ ]:

**import** docx

doc **=** docx**.**Document()

doc**.**add\_paragraph('Hello, world!')

**<**docx**.**text**.**Paragraph object at 0x0000000003B56F60**>**

doc**.**save('helloworld.docx')

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

In [ ]:

**Ans: import** docx

doc **=** docx**.**Document()

doc**.**add\_paragraph('Hello world!')

**<**docx**.**text**.**Paragraph object at 0x000000000366AD30**>**

paraObj1 **=** doc**.**add\_paragraph('This is a second paragraph.')

paraObj2 **=** doc**.**add\_paragraph('This is a yet another paragraph.')

paraObj1**.**add\_run(' This text is being added to the second paragraph.')

**<**docx**.**text**.**Run object at 0x0000000003A2C860**>**

doc**.**save('multipleParagraphs.docx')

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

In [ ]:

**Ans:** doc **=** docx**.**Document()

doc**.**add\_heading('Header 0', 0)

**<**docx**.**text**.**Paragraph object at 0x00000000036CB3C8**>**

doc**.**add\_heading('Header 1', 1)

**<**docx**.**text**.**Paragraph object at 0x00000000036CB630**>**

doc**.**add\_heading('Header 2', 2)

**<**docx**.**text**.**Paragraph object at 0x00000000036CB828**>**

doc**.**add\_heading('Header 3', 3)

**<**docx**.**text**.**Paragraph object at 0x00000000036CB2E8**>**

doc**.**add\_heading('Header 4', 4)

**<**docx**.**text**.**Paragraph object at 0x00000000036CB3C8**>**

**>>>** doc**.**save('headings.docx')