Assignment 12 Solutions

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

ANS: For `PdfFileReader()` file objects WILL be opened in `rb` -> read binary mode, Whereas for `PdfFileWriter()` file objects WILL be opened in `wb` -> write binary mode.

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

ANS: Calling getPage(4) will return a Page object for page 5, since page 0 is the first page.

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

ANS: The numPages variable stores an integer of the number of pages in the PdfFileReader object.

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?

ANS: If pdf is encrypted we will Call the `decrypt('swordfish')` .

In [ ]:

```
# for Example Code:
from PyPDF import PdfFileReader
pdf_reader = PdfFileReader(file_path)
if pdf_reader.isEncrypted: # to check whether the pdf is encrypted or not
    pdf_reader.decrypt("swordfish")
for page in pdf_reader.pages:
    print(page.extractText())
```

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

ANS: PyPDF2 Package provides 2 methods to rotate a page

1. The `rotateClockwise()` and 2. `rotateCounterClockwise()` methods. The degrees to rotate is passed as an integer argument.

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

ANS: The structure of a document is represented by three different data types in `python-Docx` . At the highest level, a `Document` object represents the entire document. The Document object contains a list of `Paragraph` objects for the paragraphs in the document. (A new paragraph begins whenever the user presses `ENTER` or `RETURN` while typing in a Word document.) Each of these Paragraph objects contains a list of one or more `Run` objects.

The text in a Word document is more than just a string. It has font, size, color, and other styling information associated with it. A style in Word is a collection of these attributes. A Run object is a contiguous run of text with the same style. A new Run object is needed whenever the text style changes.

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

ANS: Use `print(doc.paragraphs)` .

In [ ]:

```
1  # Example Program
2  from docx import Document
3  doc = Document("sample_file.docx")
4  print(doc.paragraphs)
5  for paragraph in doc.paragraphs:
6      print(paragraph.text)
```

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

ANS: Run object has bold, underline, italic, strike, and outline variables. The text in a Word document is more than just a string. It has font, size, color, and other styling information associated with it.

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

ANS: True always makes the Run object bolded and False makes it always not bolded, no matter what the style's bold setting is. None will make the Run object just use the style's bold setting.

In [ ]:

```
1  # Example Program
2  bold = True  # Style Set to Bold
3  bold = False # Style Not Set to Bold
4  bold = None # Style is Not Applicable
```

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

ANS: Call the `docx.Document()` function.

In [ ]:

```python
# Example Program
from docx import Document
document = Document()
document.add_paragraph("iNeuron Full Stack DataScience Course")
document.save('mydocument.docx')
```

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

ANS: Use doc.add_paragraph('Hello there!')

In [ ]:

```python
# Example Program
from docx import Document
doc = Document()
doc.add_paragraph('Hello, there!')
doc.save('hello.docx')
```

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

ANS: The levels for a heading in a word document can be specified by using the level attribute inside the add_heading method. There are a total of 5 levels statring for 0,1, 2, 3 and 4.