

## Forum:[Hack Forums](#)

Thread: [very strange issue with python lists and dict's i need help with this, explanatio](#)

**degado** - March 4, 2020, 12:59 pm

Hello people, I have python programming problem with lists. I am trying to write a script that extracts the bookmark names. I am using PyPDF3 for this to work.

Help me out if you can

so, in its raw form the PyPDF3 returns me the following structure from the pdf.

```
{'/Title': 'Preface', '/Page': IndirectObject(45, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E2030E48>, '/Top': 576, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E2030F60>}  
[{'/Title': 'Mission', '/Page': IndirectObject(51, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E2030FD0>, '/Top': 349.12354, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D0F0>}, {'/Title': 'Review', '/Page': IndirectObject(55, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E203D160>, '/Top': 367.72357, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D278>}, {'/Title': 'Conventions Used in This Book', '/Page': IndirectObject(55, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E203D2E8>, '/Top': 192.3996, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D400>}, {'/Title': 'Using Code Examples', '/Page': IndirectObject(65, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E203D470>, '/Top': 228.42026, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D588>}, {'/Title': 'Safari® Books Online', '/Page': IndirectObject(78, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E203D5C0>, '/Top': 430.72357, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D710>}, {'/Title': 'How to Contact Us', '/Page': IndirectObject(96, 0), '/Type': '/XYZ', '/Left': <PyPDF3.generic.NullObject object at 0x00000181E203D780>, '/Top': 581.32361, '/Zoom': <PyPDF3.generic.NullObject object at 0x00000181E203D860>}]
```

the output is bigger, but it is sufficient for illustration purposes

so as we can see it returns a dict or a array of dicts depending if the bookmarks are stored in the multi level

```
1  
2  
3  
3.1  
3.2  
etc
```

I want to print them all out in the list so there is my attempt below, but it does not work correctly I suspect it something wrong with the loop, can't get this to work,

```
from PyPDF3 import*
```

```
pdf_input = PdfFileReader(open("test.pdf", "rb"))  
pdf_output = PdfFileWriter()
```

```

pdf_bookmarks = pdf_input.getOutlines(node=None, outlines=None)
bookmark_name = "/Title"
bookmark_list = []

for count1 in range(pdf_bookmarks.__len__()):
    if isinstance(pdf_bookmarks[count1], dict):
        bookmark_sub = False
        bookmark_list.append(pdf_bookmarks[count1][bookmark_name])
        print(pdf_bookmarks[count1][bookmark_name])
        print("l1")

    if isinstance(pdf_bookmarks[count1], list):
        bookmark_sub = True
        for count2 in range(pdf_bookmarks[count1].__len__()):
            bookmark_list.append(pdf_bookmarks[count1][count2][bookmark_name])
            print(pdf_bookmarks[count1][count2][bookmark_name])
            print("l2")

            if isinstance(pdf_bookmarks[count1][count2], list):
                for count3 in range(pdf_bookmarks[count1][count2].__len__()):
                    bookmark_list.append(pdf_bookmarks[count1][count2][count3][bookmark_name])
                    print(pdf_bookmarks[count1][count2][count3][bookmark_name])
                    print("l3")
                    if isinstance(pdf_bookmarks[count1][count2][count3], list):
                        for count4 in range(pdf_bookmarks[count1][count2][count3].__len__()):
                            bookmark_list.append(pdf_bookmarks[count1][count2][count3][count4][bookmark_name])
                            print(pdf_bookmarks[count1][count2][count3][count4][bookmark_name])
                            print("l4")

                            if isinstance(pdf_bookmarks[count1][count2][count3][count4], list):
                                for count5 in range(pdf_bookmarks[count1][count2][count3][count4].__len__()):
                                    bookmark_list.append(pdf_bookmarks[count1][count2][count3][count4][count5][bookmark_name])
                                    print(pdf_bookmarks[count1][count2][count3][count4][count5][bookmark_name])
                                    print("l5")

for bookmark_counter in range(bookmark_list.__len__()):
    print(bookmark_list[bookmark_counter])

```

output

Microsoft Windows [Version 10.0.17763.1039]  
 © 2018 Microsoft Corporation. All rights reserved.

D:\python\projects\pdf-tool>python pdf-helper.py

Copyright

l1

Table of Contents

l1

Foreword

I1

Preface

I1

Mission

I2

Review

I2

Conventions Used in This Book

I2

Using Code Examples

I2

Safari® Books Online

I2

How to Contact Us

I2

Chapter 1. ES? Now & Future

I1

Versioning

I2

Transpiling

I2

Traceback (most recent call last):

File "pdf-helper.py", line 19, in <module>

bookmark\_list.append(pdf\_bookmarks[count1][count2][bookmark\_name])

TypeError: list indices must be integers or slices, not str

You see my pdf that i am testing with has more than 2 bookmark sublevels, and it gets stuck when trying to read the third level

I have a suspicion that I am not accessing lists correctly

Thanks for help

I1, I2 etc text is the indicator on which loop it gets stuck for debug Purposes only.

Thanks for help

**h0lyI337** - March 4, 2020, 2:15 pm

one of those indexes are coming back as str instead of an int. Convert it into an int.

**degado** - March 5, 2020, 9:43 am

(Mar 4, 2020 03:15 PM)h0lyI337 Wrote: [\[CITE\]](#) one of those indexes are coming back as str instead of an int. Convert it into an int. I am using array length to go through the loop and there are no strings involved except the actual bookmark names of course.

it works for 1st and 2nd level bookmarks but does not for 3rd so it is not the str issue.

i checked the index with the following statement

```
if isinstance(pdf_bookmarks[count1][count2], str):
```

```
    print("string")
```

it is all good

