

实用python编程 第6讲

# openpyxl读写excel

2017-11-13

# 本节内容

- 创建excel
- 访问excel

# openpyxl

Openpyxl is a Python library for reading and writing Excel 2010 xlsx/xlsm/xltx/xltm files

<http://openpyxl.readthedocs.io>

查看版本

```
import openpyxl  
print openpyxl.__version__
```

# excel sheet

- pandas风格 (pandas style)

	A	B	C	D
1		Age		
2	0	10		
3	1	20		
4	2	30		
5	3	20		
6	4	15		
7	5	30		
8	6	45		
9				

# 创建excel文件

从pandas DataFrame中导入数据

- `openpyxl.utils.dataframe.dataframe_to_rows()`

```
import pandas as pd
from openpyxl import Workbook
from openpyxl.utils.dataframe import dataframe_to_rows

# Create a Pandas dataframe from the data.
df = pd.DataFrame({'Age': [10, 20, 30, 20, 15, 30, 45]})
wb = Workbook()
ws = wb.active

for r in dataframe_to_rows(df):
    ws.append(r)

wb.save('test.xlsx')
```



	A	B	C
1		Age	
2	0	10	
3	1	20	
4	2	30	
5	3	20	
6	4	15	
7	5	30	
8	6	45	
9			

# openpyxl的数据结构

## 一个workbook

- 包含至少一个worksheet
- 每个worksheet包含多个cell

```
# 新建一个workbook
from openpyxl import Workbook
wb = Workbook()
```

```
# 每个新建的workbook都会自带一个（空的）worksheet
```

```
ws1 = wb.active
print ws1.title
```

```
# 新建一个worksheet
ws2 = wb.create_sheet("Mysheet")
print ws2.title
```

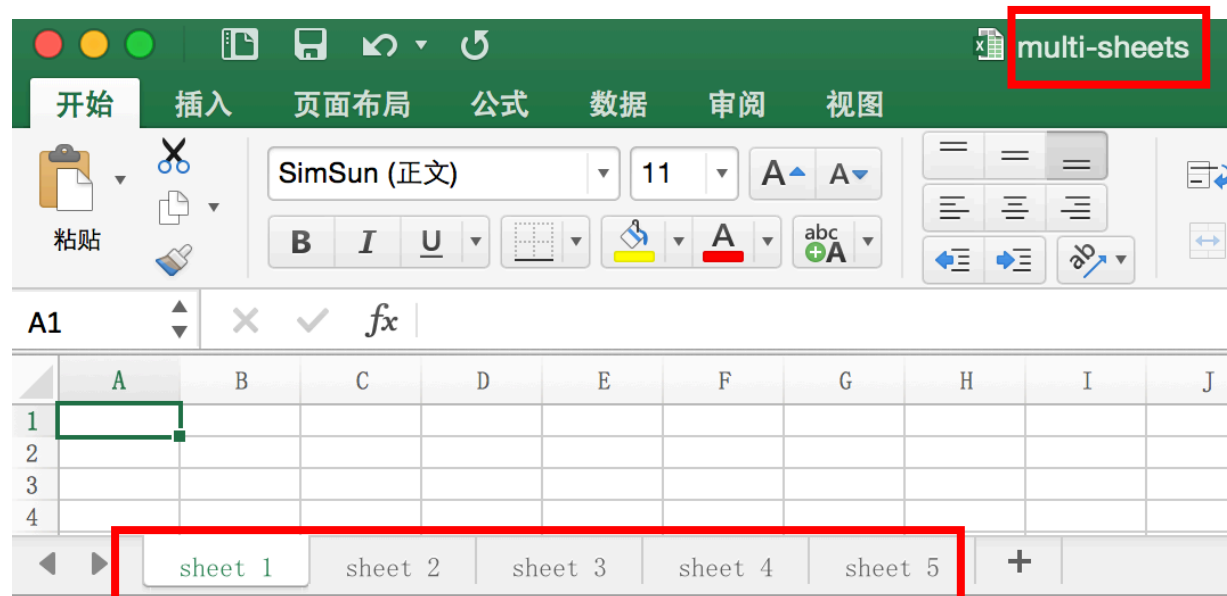


title值，应不同于已有  
worksheet的title

# 课堂练习-1

用openpyxl新建一个excel文件

- 包含5个sheet，sheet名字依次为sheet 1, sheet 2, ..., sheet 5
- excel文件名为 multi-sheets.xlsx



# 读取excel文件

```
from openpyxl import load_workbook
```

```
wb = load_workbook('test.xlsx')
```



# 访问workbook中的各个worksheet

- 方法1: 通过指定 title

```
ws.title = 'my sheet'  
...  
ws = wb['my sheet']
```

- 方法2: 循环遍历

```
for ws in wb:  
    print ws.title
```

```
# 获得worksheet的有效行数和列数  
ws.max_row  
ws.max_column
```

# 访问worksheet中的cell

- 方法1：通过指定key(类似访问dict)

```
c1 = ws['B2']  
print c1.value
```

- 方法2：cell()函数

```
c2 = ws.cell(row=2,column=2)  
print c2.value
```

# 修改worksheet中的cell

- 方法1：通过指定key (类似修改dict)

```
c1 = ws['B2']  
print c1.value  
ws['B2'] = 33  
print c1.value
```

- 方法2: cell()函数

```
c2 = ws.cell(row=2,column=2)  
print c2.value  
c2.value = 66  
print ws.cell(row=2,column=2).value
```

# 访问worksheet中的多个cell

- 注意excel的行索引从1开始（而不是0）

访问范围	示例代码	解释
特定一行	<code>ws[1]</code>	第1行
特定一列	<code>ws['B']</code>	第B列
特定行列	<code>ws['A2':'B4']</code>	A2与B4构成的矩形区域

```
cell_range = ws['A2':'B4']  
for row in cell_range:  
    for cell in row:  
        print cell, cell.value
```

```
<Cell u'Sheet'.A2> 0  
<Cell u'Sheet'.B2> 66  
<Cell u'Sheet'.A3> 1  
<Cell u'Sheet'.B3> 20  
<Cell u'Sheet'.A4> 2  
<Cell u'Sheet'.B4> 30
```

# 设置pandas style

- 让index与header粗体显示

```
import pandas as pd
from openpyxl.utils.dataframe import dataframe_to_rows

df = pd.DataFrame({'Age': [10, 20, 30, 20, 15, 30, 45]})
wb = Workbook()
ws = wb.active

for r in dataframe_to_rows(df):
    ws.append(r)

for cell in ws['A'] + ws[1]:
    cell.style = 'Pandas'

wb.save('test.xlsx')
```



	A	B	C
1		Age	
2	0	10	
3	1	20	
4	2	30	
5	3	20	
6	4	15	
7	5	30	
8	6	45	
9			

# 将worksheet导入DataFrame

```
from pandas import DataFrame
col_name = ws['B1'].value
vals = [x.value for x in ws['B'][1:]]
df = DataFrame({col_name:vals})
print df.shape
print df.Age
```

# 课堂练习-2

将作业3的数据data20171023.txt导入到excel文件中

	A	B	C
1		x1	x2
2	0	0.811991	-0.46155
3	1	-1.50515	1.00333
4	2	0.197914	0.662164
5	3	1.03889	-0.58688
6	4	1.09047	-0.14467
7	5	-0.183808	0.864504
8	6	-1.5803	-1.05173
9	7	-0.329522	1.2748
10	8	-0.284141	-2.44655
11	9	1.16204	1.16582
12	10	0.442525	-1.7029
13	11	-1.15792	-0.91186
14	12	0.923053	-0.6982
15	13	0.0888079	0.005465
16	14	0.567213	0.280248
17	15	-0.055587	-0.30675
18	16	-0.825939	-1.96446
19	17	0.027196	0.003907
20	18	-0.577381	-0.23149
21	19	-1.7388	-0.02135

# 参考资料

<http://openpyxl.readthedocs.io/en/default/tutorial.html>