



中國人民大學  
RENMIN UNIVERSITY OF CHINA

## 第2編 Python数据分析

# 第3讲 文件

余力

buaayuli@ruc.edu.cn



中國人民大學  
RENMIN UNIVERSITY OF CHINA



---

# 01. 文本文件

---

# 文本文件处理步骤

---

1. 创建并打开文件，使用`open()`函数返回1个file对象。
2. 调用file对象的`read()`、`write()`等方法处理文件。
3. 调用`close()`关闭文件，释放file对象占用的资源。

# 1. 文件的创建与打开

---

- 文件的**打开或创建**可以使用函数open
- **open(file, [mode[, buffering]])**—>file object
- 0表示无缓冲；1表示行缓冲；
- 大于1则表示缓冲区的大小，
- -1（或者任何负数）代表使用默认的缓冲区大小。

# 文件的打开模式

参数	描述
r	以 <b>只读</b> 的方式打开文件
r+	以 <b>读写</b> 的方式打开文件 <b>不会创建不存在</b> 的文件
w	以 <b>写入</b> 的方式打开文件。先 <b>删除文件原有</b> 的内容，再重新写入新的内容。如果文件不存在，则创建一个新的文件。
w+	以 <b>读写</b> 的方式打开文件。先 <b>删除文件原有</b> 的内容，再重新写入新的内容。如果文件不存在，则创建一个新的文件。
a	以 <b>写入</b> 的方式打开文件，在 <b>文件末尾追加</b> 新的内容。如果文件不存在，则创建一个新文件
a+	以 <b>读写</b> 的方式打开文件，在 <b>文件末尾追加</b> 新的内容。如果文件不存在，则创建一个新文件
b	以 <b>二进制</b> 模式打开文件。可与r、w、a、+结合使用
U	支持所有的换行符号。如：'\r'、'\n'、'\r\n'

# 文件的创建、写入和关闭

---

```
def make_story():  
    f=open('Chp2_story.txt','w')  
    f.write('Marry had a little lamb,\n')  
    f.write('and then she had some more.\n')  
    f.close()  
  
make_story()
```

# 在文件尾添加内容

---

```
def add_to_story(line, fname='Chp2_story.txt'):
    f=open(fname,'a')
    f.write(line)
    f.close()

add_to_story('haha!\n')
```

## 2、文件的读取

---

- `readline()`、`readlines()`、`read()`

### (1) 按行读取方式`readline()`

```
def readfile(file):  
    f=open(file)  
    while True:  
        line=f.readline()  
        if line:  
            print(line)  
        else:  
            break  
    f.close()  
readfile('story.txt')
```



## 2、文件的读取

---

### (2) 多行读取方式readlines()

```
def readfile(file):  
    f=open(file)  
    line=f.readlines()  
    if line:  
        print(line)  
    else:  
        print "No file"  
    f.close()  
readfile('story.txt')
```

## 2、文件的读取

### (3) 一次性读取方式

- 读取文件最简单的方法是使用read(), read()将从文件中一次性读出所有的内容, 并赋值给1个字符串变量。

```
def readall(file):  
    f=open(file)  
    content=f.read()  
    if content:  
        print(content)  
    else:  
        print "No file"  
    f.close()  
readfile('story.txt')
```

### 3、文件的写入

---

- write()、writelines()

```
def writetofile(file):  
    f=open('hello.txt','w+')  
    context=['hello world!\n', 'hello China!\n']  
    f.writelines(context)  
    f.close()  
writetofile('story.txt')
```

# 文件内容的搜索

- 文件内容的搜索和替换可以使用字符串查找和替换来实现。
- 例：从hello.txt文件中统计字符串'hi'出现的次数。

```
def findinfile(filename,content):  
    import re  
    f1=open(filename,'r')  
    count=0  
    for s in f1.readlines():  
        fs=re.findall(content,s)  
        if len(fs)>0:  
            count+=fs.count(content)  
    print ('find '+str(count)+' '+  
content)  
    f1.close()  
findinfile('hello2.txt','hi')
```

# 文件内容的替换

将hello.txt中的字符串'hello'全部替换为'hi'，并将结果存入hello2.txt中

```
import re
f1=open('hello.txt','r')
f2=open('hello2.txt','w')
for s in f1.readlines():
    f2.write(s.replace('hello', 'hi'))
f1.close()
f2.close()
```



中國人民大學  
RENMIN UNIVERSITY OF CHINA



---

## 02. 数据文件

---

# pandas读取数据文件

---

## ■ pandas 模块读取 txt 文件

```
import pandas as pd
```

```
mydata_txt = pd.read_csv('Chp2_txtdata.txt',  
                        ,sep='\t',names=['A','B'])
```

```
print(mydata_txt)
```

写入

```
mydata_txt.to_csv('yahoo_data.txt', index=False, sep=' ')
```

# 从CSV文件导入数据

---

```
import csv  
filename='jiaotan.csv'  
f=open(filename)  
read=csv.reader(f)  
for i in read:  
    print(i)
```



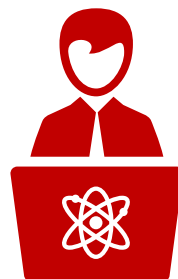
# 读取txt文件

---

```
import numpy as np
filename = 'Chp2_txtdata.txt.txt'
pos = [] ; Efield = []
with open(filename, 'r') as file_to_read:
    while True:
        lines = file_to_read.readline() # 整行读取数据
        if not lines:
            break
        p_tmp, E_tmp = [float(i) for i in lines.split()]
        pos.append(p_tmp) # 添加新读取的数据
        Efield.append(E_tmp)
    pos = np.array(pos) # 将数据从list类型转换为array类型
    Efield = np.array(Efield)
print pos
print Efield
```



中國人民大學  
RENMIN UNIVERSITY OF CHINA



---

# 谢谢大家!

---

