

# 15.常用文件读写

## 一、文件操作函数

---

### 1. 打开和创建文件对象函数open

```
open(file, mode='r', buffering=None, encoding=None, errors=None,
      newline=None, closefd=True)
```

- 参数说明

- file 你要打开的文件名（字符串），它可以是相对路径也可以是绝对路径
- mode 文件的打开方式，默认是以只读的方式打开
- buffering 访问文件的缓冲方式。0表示不缓冲，1表示只缓冲1行数据，任何其它大于1的值代表使用给定值作为缓冲区大小。不提供该参数或者给定负值代表使用系统默认缓冲机制，一般情况下采用系统默认的就好。
- encoding 文件编码，gbk utf-8
- errors 文件出错处理
- newline 文件换行符，默认是跟操作系统的一致。在windows下\r\n mac和Linux\n
- closefd 关闭文件描述符

文件打开方式说明

模式	说明
r	以只读方式打开
rU或U	以读方式打开，同时提供通用换行符支持
w	以写方式打开，如果文件已经存在，会覆盖掉原来的数据
a	以追加模式打开，在文件后面追加数据，不能读
r+	以读写模式打开
w+	以读写模式打开，参见w
a+	以读写追加模式打开，参见a
rb	以二进制读模式打开
wb	以二进制写模式打开
ab	以二进制追加模式打开
rb+	以二进制读写模式打开
wb+	以二进制读写模式打开
ab+	以二进制读写追加模式打开

2. 工厂函数file---python3中已经不用了  
这个实现的功能跟open函数是一样的。

## 二、文件的读写操作

---

1. 读取文件中的数据
  - 使用read方法一次性读取所有的数据
  - 使用readline方法每次只读取一行数据
  - 使用readlines方法将每行读取出来得到一个列表
  - 使用for循环迭代读取
  - 使用列表推导式来将文件的每一行都读取到列表中去
2. 写入数据到文件
  - 使用write方法写入字符串
  - 使用writelines一次性从序列中写入多个字符串到文件
3. 将缓冲区数据写入到文件
  - 使用flush方法可以将缓冲区中的数据写入到文件中
4. 关闭文件

- 文件打开之后一定要记得关闭，否则会造成资源泄漏
- 使用close方法关闭文件

## 三、文件对象的内建方法

方法名	说明
file.close()	关闭文件
file.fileno()	返回文件的描述符
file.flush()	刷新文件的内部缓冲区
file.isatty()	判断file是否是一个类tty设备
file.next()	返回文件的下一行，类似readline，但是没有其它行时会引发异常
file.read(size=-1)	从文件中读取size个字节，为给定或负值的时候表示读取剩余所有字节，然后作为字符串返回
file.truncate(size=file.tell())	截取文件到最大size字节，默认为当前文件位置
file.readline(size=-1)	从文件中读取并返回一行（包括换行符），或返回最大size字节
file.readlines()	读取文件所有行并作为一个列表返回（包含所有换行符）
file.seek(off, whence=0)	在文件中移动文件指针，从whence（0代表文件起始，1代表当前位置，2代表文件末尾）偏移off字节
file.tell()	返回文件指针在当前文件的位置
file.write(str)	向文件写入字符串
file.writelines(seq)	向文件写入字符串序列seq，seq应该是一个返回字符串的可迭代对象

## 四、使用上下文管理器操作文件

python的上下文管理器会自动的清理一些资源，比如打开的文件和关闭文件，当执行完成上下文管理器的代码之后，文件会被自动的关闭。

## 1. with 上下文管理器

### ◦ 语法

```
with open(filename, mode) as variable:  
    # 对文件进行各种操作
```

### ◦ 例子:

```
# 打开一个文件然后读取文件中的内容加工之后写入到列表  
text = []  
with open('1.txt') as f:  
    for line in f:  
        text.append(line.strip())
```

# 作业

---

1. 对gbk编码的文件进行读写
2. 将之前的第二个爬虫程序采集下来的内容写入到txt文件里面，文件名是标题名称，文件内容就是正文的内容

注意事项：windows的文件名不能包含<>/\|:"\*?