



文件读写

北京理工大学计算机学院 高玉金

2019年3月



文件的使用

- Python处理的各种类型的数据，通过内置数据结构如列表、字典等在计算机内存中进行处理加工，最后必须要存入外部存储器中永久保存
- 文件是一个存储在辅助存储器上的数据序列
- 文件是常用的存储形式
- 文件可以包含任何数据内容
- 文件的类型分为文本文件和二进制文件
 - 文本由单一特定编码的字符串构成
 - 二进制文件直接由0和1比特串构成，没有统一编码，需要软件进行编码和解码



文件的打开和关闭

- `file = open(file_name, mode)`, 返回一个stream对象file
- `file.close()`, 关闭文件
- 建议使用with as语句, 可以不用显式关闭文件

打开模式	含义
'r'	只读模式, 如果文件不存在, 返回异常FileNotFoundError, 默认值
'w'	覆盖写模式, 文件不存在则创建, 存在则完全覆盖源文件
'x'	创建写模式, 文件不存在则创建, 存在则返回异常FileExistsError
'a'	追加写模式, 文件不存在则创建, 存在则在原文件最后追加内容
'b'	二进制文件模式
't'	文本文件模式, 默认值
'+'	与r/w/x/a一同使用, 在原功能基础上增加同时读写功能



读取文件

```
With open (fname, "r") as f:
```

```
    lines = f.readlines()
```

```
    for line in lines:
```

```
        print(line)
```

注意每行后面有换行符！

方法	含义
<file>.read(size=-1)	从文件中读入整个文件内容，如果给出参数，读入前size长度的字符串或字节流
<file>.readline(size = -1)	从文件中读入一行内容，如果给出参数，读入该行前size长度的字符串或字节流
<file>.readlines(hint=-1)	从文件中读入所有行，以每行为元素形成一个列表，如果给出参数，读入hint行



文件的编码示例

```
# test.txt 内容为: 北京理工大学  
file = open("test.txt", "rb")  
print(file.readline())
```

```
b' \xb1\xb1\xbe\xa9\xc0\xed  
\xb9\xa4\xb4\xf3\xd1\xa7'
```

```
# word 文件内容为: 北京理工大学 , 红色, 标亮  
file = open("test.docx", "r")  
print(file.readline())
```

UnicodeDecodeError: 'gbk' codec can't
decode byte 0xa2 in position 50: illegal
multibyte sequence



写入文件

```
ls = ["唐诗", "宋词", "元曲"]
```

注意无换行符！

```
with open(fname, "w+") as fo:
```

```
    #fo.writelines(ls)
```

```
    fo.writelines([line+"\n" for line in ls]) #换行
```

```
    fo.Seek(0) # 把指针移到文件头部
```

.....

方法	含义
<file>.write(s)	向文件写入一个字符串或字节流
<file>.writelines(lines)	将一个元素为字符串的列表写入文件
<file>.seek(offset)	改变当前文件操作指针的位置，offset的值： 0: 文件开头； 1: 当前位置； 2: 文件结尾



使用faker库产生模拟数据

- <https://github.com/joke2k/faker>
- 提供模拟数据，前后端分离，适合测试和调试
- 使用faker的智能感知功能提供帮助，或查看相关文档：
<https://faker.readthedocs.io/en/latest/index.html>
- 提供扩展：自定义provider

```
>>> fake.ascii_email()
'usmith@hotmail.com'
>>> fake.words()
['officiis', 'dolorum', 'ad']
```

```
>>> from faker import Faker
```

```
>>> fake = Faker("zh_cn")
```

```
>>> fake.name()
```

```
'毛岩'
```

```
>>> fake.address()
```

```
'山西省丹丹市朝阳荆门路q座 693695'
```



模拟数据存入文件示例

```
from faker import Faker
fake = Faker("zh_cn")
with open("test.txt", "w") as fo:
    for i in range(10):
        fo.writelines([fake.name()+"\n", fake.address()+"\n"])
```

```
石健
辽宁省宇市双滦海街n座 801069
崔建国
新疆维吾尔自治区张家港市和平哈尔滨街Y座 615561
孙慧
内蒙古自治区贵阳县房山梧州路Q座 365182
```