

专栏：002：python 文件操作

E - 爬虫技术 文件操作

文件访问模式

序号	文件访问模式	解释
01	r	读
02	w	写
03	a	追加
04	r+	读写模式
05	w+	读写模式
06	a+	读写模式
07	b	二进制文件操作：图像,视频之类的

代码示例

```
"读文件内容"

filename = "test.txt"
f = opne(filename,"r")
print(f.read())
f.close()

"写文件内容"
filename_2 = "write.txt"
f = opne(filename_2,"w")
f.write("I love python")
f.close()
```

文件对象方法

序号	对象方法	操作及解释
01	file.close()	关闭文件
02	file.read()	读取文件内容，一次加载所有内容至内存中
03	file.readline(size= num)	读取从起始位置到num位置的所有内容
04	file.readlines()	按行读取内容
05	file.write()	向文件中写入内容
06	file.tell()	返回当前在文件中的位置
07	file.seek()	在文件中移动文件指针

注意事项

- open

单纯使用open函数，及时将文件关闭close.

解释：文件使用完毕后必须关闭，因为文件对象会占用操作系统的资源，并且操作系统同一时间能打开的文件数量也是有限的

- with + codecs模块

上下文解释器

```
with codecs.open(filename,"r",encoding="utf8") as f:
```

不需要使用close关闭文件，而可以安全的使用。

- EOF: end of file

操作演示

```
# test.txt

---
i write this file just for fun .
how old are you .
and?
can i help you .
---

# 读文件操作

import codecs
filename = "test.txt"
with codecs.open(filename,"r",encoding="utf8") as f:
    all_content = f.readline(20)
    print(all_content)

---
i write this file ju      # 数下刚好20
---

# 二进制文件写入
# 将文件夹下首页的24张图片下载至本地
import requests
from lxml import etree
import codecs
url = "https://alpha.wallhaven.cc/random"
con = requests.get(url)
html = con.content

Html = etree.HTML(html)
pattern_figure = Html.xpath('//figure//@data-src')
i= 0
print(pattern_figure)
for one in pattern_figure:
    print(one)
    with codecs.open("%s.jpg" %i , "wb") as f:
        one_pic = requests.get(one)
        f.write(one_pic.content)
        i+=1
```

如图示图片文件：



参考资料

教程1

xpath语法

关于本人，国内小硕，自学编程
半路出家的IT学习者，所学编程语言：python
感兴趣领域：爬虫及数据科学
Github：[wuxiaoshen](#)
weibo: [乌小小申](#)