











PRODUCE BY DEFEASE











網易 NETEASE



















TO- WE BY DELFIN









































by dage by delevse









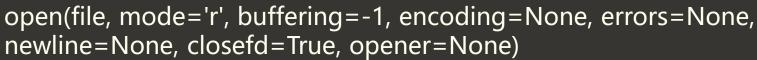


BEGEORGE BA DELEUSE

100

文件读取

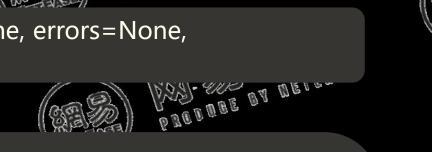






onn 🗐

- •file: 必需,文件路径(相对或者绝对路径)。
- ·mode: 可选,文件打开模式
- •buffering: 设置缓冲
- •encoding: 一般使用utf8
- •errors: 报错级别
- •newline: 区分换行符
- •closefd: 传入的file参数类型
- •opener: 设置自定义开启器,开启器的返回值必须是一个打开的文件描述符。



















文件模式













-					
模式	描述				
r	以只读方式打开文件。文件的指针将会放在文件的开头,这是默认模式				
rb	以二进制格式打开一个文件用于只读。文件指针将会放在文件的开头,这是默认模式				
r+	打开一个文件用于读写。文件指针将会放在文件的开头				
rb+	以二进制格式打开一个文件用于读写。文件指针将会放在文件的开头				
w	打开一个文件只用于写入。如果该文件已存在,就将其覆盖;如果该文件不存在,就创建新文件				
wb	以二进制格式打开一个文件只用于写入。如果该文件已存在,就将其覆盖;如果该文件不存在,就创建新文件				
w+	打开一个文件用于读写。如果该文件已存在,就将其覆盖;如果该文件不存在,就创建新文件				
wb+	以二进制格式打开一个文件用于读写。如果该文件已存在,就将其覆盖;如果该文件不存在,就创建新文件				
a	打开一个文件用于追加。如果该文件已存在,文件指针就会放在文件的结尾。也就是说,新内容将 会被写入已有内容之后。如果该文件不存在,就创建新文件进行写入				
ab	以二进制格式打开一个文件用于追加。如果该文件已存在,文件指针就会放在文件结尾。也就是说,新内容将会被写入已有内容之后。如果该文件不存在,就创建新文件进行写入				
a+	打开一个文件用于读写。如果该文件已存在,文件指针就会放在文件的结尾。文件打开时是追加模式。如果该文件不存在,就创建新文件用于读写				
ab+	以二进制格式打开一个文件用于追加。如果该文件已存在,文件指针将会放在文件结尾;如果该文件不存在,就创建新文件用于读写和追加				



1863



EASE

















四晶



F. S. S.





文件类型	文件说明	读取函数	写入函数	
CSV	该类型文件以纯文本形式存储通常以逗	read_csv	to_csv	88
	号分隔的表格数据(数字和文本)			
HDF	美国国家高级计算应用中心研制的一种	read_hdf	to_hdf	7
	能高效存储和分发科学数据的层级数据			T
	格式			
SQL	一种用结构化查询语言编写的数据库查	read_sql	to_sql	
	询脚本文件			
JSON	一种轻量级的文本数据交换格式文件	read_json	to_json	Q.
HTML	一种由超文本标记语言编写的网页文件	read_html	to_html	4
PICKLE	Python 内部支持的一种序列化文件	read_pickle	to_pickle	















SOUNEE BY DEVEN















PRODUCE BY DETERS

PRODUCE OF DELEV





















PAUL



HH









100

300





具体要求: 爬取豆瓣图书, 并将数据保存到本地txt文件

```
def spyderInfo(data):
    s=etree.HTML(data)
    file=s.xpath('/html/body/div[3]/div[1]/div/div[1]/div')

for div in file:
        title=div.xpath('.//tr/td[2]/div[1]/a/@title')
        info=div.xpath('.//tr/td[2]/p[1]/text()')
        rating=div.xpath('.//tr/td[2]/div[2]/span[1]/@class')
        score=div.xpath('.//tr/td[2]/div[2]/span[2]/text()')
        rating_num=div.xpath('.//tr/td[2]/div[2]/span[3]/text()')
        href=div.xpath('.//tr/td[2]/div[1]/a/@href')
        #print("{}.{};{};{};{}\n".format(title,info,rating,score,rating_num))

with open('./豆瓣图书.txt','a',encoding='utf-8') as f:
    f.write(str(title))
```



PRODUCE.







ES-AND BA BELFIN













BUGDUE BA DELEVSE







PRODURE BY DETERSE











PRODURE OF MELETSE









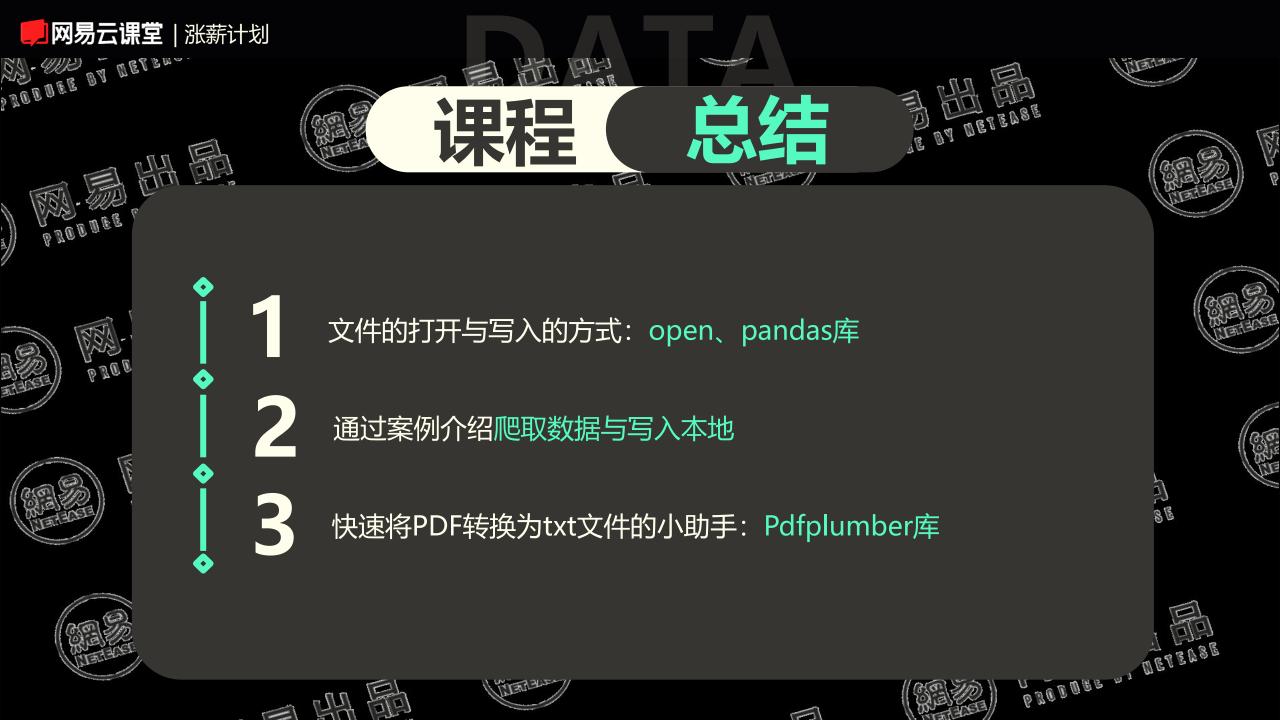


BEGEORGE BA DELEUSE

Out[3]: 342

Pdfplumber洋: 完文女小助手 主要用途: 快速将PDF转化为TXT文件

#PDF格式转化为 txt文件,数据量大的统计非常方面 In [3]: v import pdfplumber from openpyxl import Workbook #打开excel, 统计表格使用 with pdfplumber.open("/Users/samuelzhan/Downloads/中国平安2020财报.PDF") as p: page count = len(p.pages) #统计文档的页数 for i in range(0,page_count): #提取每页的对象并存储 page =p.pages[i] #提取每页的文字信息 textdata=page.extract text() data=open('/Users/samuelzhan/Downloads/平安财报2020.text','a') #将: #文档写入 data.write(str(textdata)) page count



ED-ANG BA WELFA





























ENODER DA DELEVSE





100









