redorange@AcA:~/桌面$ ls

' ' class0430.docx google-chrome.desktop test.py 未命名

apps courseAndStudy python-latest-news.csv 钉钉.desktop 未命名.c

redorange@AcA:~/桌面$ cat test.py

#!/usr/bin/env python3.6

# -\*- coding: utf-8 -\*-

from urllib import request as ur

import re

import csv

def get\_html(url):

# 请求站点获得一个HTTPResponse对象

response = ur.urlopen(url)

# 网页内容

html = response.read().decode('utf-8')

#print(html)

return html

def deal\_with(data):

pattern = re.compile('<li>\s+<time\sdatetime="(\d{4})-(\d{2})-(\d{2}).\*?</time>\s+<a\shref=".\*?>(.\*?)</a></li>',re.S)

items = re.findall(pattern, data)

#print(items[:5])

return items[:5]

def write\_data(items):

headers = ['日期','标题']

# python-latest-news.csv

# 用utf-8 会导致中文乱码

with open('python-latest-news.csv','w+',encoding='utf-8',newline='') as f:

f\_csv = csv.writer(f)

f\_csv.writerow(headers)

rows=[]

for cnt in items:

datetime = cnt[0]+"-"+cnt[1]+"-"+cnt[2]

news = cnt[-1]

#print(news)

row = [datetime,news]

rows.append(row)

#print(rows)

f\_csv.writerows(rows)

f.close()

def main():

url = 'https://www.python.org'

# 获取html文件

data = get\_html(url)

# 处理数据

items = deal\_with(data)

# 写入数据

write\_data(items)

print("写入数据成功~")

return None

if \_\_name\_\_ == '\_\_main\_\_':

main()redorange@AcA:~/桌面$ python3 test.py

写入数据成功~

redorange@AcA:~/桌面$ ./test.py

bash: ./test.py: Permission denied

redorange@AcA:~/桌面$ sudo ./test.py

[sudo] password for redorange:

sudo: ./test.py: command not found

redorange@AcA:~/桌面$ chmod +x test.py

redorange@AcA:~/桌面$ ls

' ' class0430.docx google-chrome.desktop test.py 未命名

apps courseAndStudy python-latest-news.csv 钉钉.desktop 未命名.c

redorange@AcA:~/桌面$ sudo ./test.py

写入数据成功~

redorange@AcA:~/桌面$ cat python-latest-news.csv

日期,标题

2020-05-10,CPython on Mobile platforms - Python Language Summit 2020

2020-05-07,Core Workflow updates - Python Language Summit 2020

2020-05-05,Python Developers Survey 2019 Results

2020-05-04,Python’s migration to GitHub - Request for Project Manager Resumes

2020-05-03,Property-Based Testing for Python builtins and the standard library - Python Language Summit 2020

redorange@AcA:~/桌面$