



南昌大学实验报告

学生姓名：__丁俊__ 学 号：__8003119100__ 专业班级：__信息安全 193 班__

实验类型：☐ 验证 ☒ 综合 ☐ 设计 ☐ 创新 实验日期：__2021.11.18__ 实验成绩：__

一、实验项目名称

数据处理

二、实验目的

1. 掌握 numpy、pandas 等库的使用
2. 掌握数据筛选、数据转换以及时间格式数据的处理
3. 掌握缺失值的检测和处理

三、实验任务

参照房屋资料处理和新闻资料处理案例完成以下任务：

资料筛选、资料转换
检测遗失值和补齐遗失值
处理时间格式数据
重塑资料

四、主要仪器设备及耗材

软件：Anaconda 或者 pycharm 等

五、实验步骤

（请根据实际的操作来填写）

1、访问新浪新闻网页

```
In [158]: import requests
import re
from bs4 import BeautifulSoup
res = requests.get('https://news.sina.com.cn/')
```



2、获取新闻信息详情函数

获取标题 title、内容 content、日期 date、来源 source、关键词 keywords，并返回一个字典。这里已经对一些数据进行了预处理，比如 content，去除一些换行符、回车符等。

```
In [104]: def getAct(url):
            res = requests.get(url)
            res.encoding = 'utf-8'
            bs = BeautifulSoup(res.text, 'lxml')
            actdict = {}
            actdict['title'] = bs.select('.main-title')[0].text
            actdict['content'] = ' '.join(bs.select('.article')[0].text.split())
            actdict['date'] = bs.select('span.date')[0].text
            actdict['source'] = bs.select('.source')[0].text
            if len(bs.select('.keywords')) > 0:
                actdict['keywords'] = bs.select('.keywords')[0].text
            return actdict
```

```
<div class="second-title">北京市发改委主任谈绪祥已任市政府党组成员</div>
<div class="date-source" data-sudaclick="share_1-0">
  <span class="date">2021年11月25日 12:29</span> == $0
  <a href="https://mp.weixin.qq.com/s/Nzff2h8MSeokIYn5HctIw" target="_blank"
    class="source" data-sudaclick="content_media_p" rel="nofollow">新京报</a>
</div>
<div class="page-tools">...</div>
::after
</div>
<!-- 面包屑 search end -->
<h1 class="main-title">北京市发改委主任谈绪祥已任市政府党组成员</h1> == $0
<!-- page-tools start -->
<style type="text/css">...</style>
```

```

<!-- 父早大罐子 begin -->
▼<div class="article-bottom clearfix" id="article-bottom">
  ▼<div class="keywords" id="keywords" data-wbkey="新冠肺炎,河北,新闻" data-
sudaclick="content_keywords_p">
    <label 关键字 : </label> == $0
    <a href="http://tags.news.sina.com.cn/新冠肺炎" target="_blank">新冠肺炎</a>
    <a href="http://tags.news.sina.com.cn/河北" target="_blank">河北</a>
  </div>
  <div class="btns">...</div>
  ::after
</div>
<!-- 文章关键字 end -->
<div id="wxFollow"></div>
<!-- 引文 start -->
<!-- 正文 start -->
▼<div class="article" id="article"> == $0
  <p> 原标题：河北辛集：全面恢复市域内人员正常出行</p>

```

3、将新闻的字典信息存入列表中

```

In [159]: res.encoding = 'utf-8'
bs = BeautifulSoup(res.text, 'lxml')
newsary = []
for ul_link in bs.select('#blk_ndxw_01 ul.list_14'):
    for li_link in ul_link.select('li'):
        if len(li_link.select('a')) > 0:
            # print(li_link.find('a').get('href'))
            newsary.append(getAct(li_link.find('a').get('href')))

```

```

In [161]: for i in newsary:
print(i)

```

至11月24日24时，安徽省尚在医学观察的无症状感染者1例（境外输入）。点击进入专题：新型冠状病毒肺炎疫情 责任编辑：祝加贝，'date': '2021年11月25日 08:08'，'source': '界面新闻'，'keywords': '\n关键字：\n新冠肺炎安徽'}

('title': '辽宁本土确诊0新增 本轮疫情首次无新增确诊', 'content': '原标题：辽宁本土确诊0新增 辽宁本轮疫情首次无新增确诊 转发好消息！加油！ 11月24日0时至24时，辽宁省无新增新冠肺炎确诊病例，无新增无症状感染者。当日治愈出院本土新冠肺炎确诊病例20例、本土无症状感染者1例；治愈出院境外输入确诊病例1例、境外输入无症状感染者1例。（总台记者李承泽） 点击进入专题：新型冠状病毒肺炎疫情 辽宁再现本土疫情 责任编辑：张迪'，'date': '2021年11月25日 08:05'，'source': '央视网'，'keywords': '\n关键字：\n新冠肺炎辽宁'}

('title': '北京昨日新增2例境外输入无症状感染者'，'content': '原标题：疫情发布 | 北京11月24日无新增新冠肺炎确诊病例 新增2例境外输入无症状感染者 来源：健康北京 11月24日0时至24时，无新增本土确诊病例、疑似病例和无症状感染者；无新增境外输入确诊病例、疑似病例，新增2例境外输入无症状感染者。全市16区无报告病例天数情况 地区 无报告病例天数 平谷区 无报告病例 延庆区 671 密云区 652 石景山区 528 门头沟区 527 东城区 526 通州区 522 西城区 520 怀柔区 328 顺义区 309 大兴区 299 房山区 106 丰台区 29 昌平区 17 海淀区 12 朝阳区 9 2020年2月29日发生境外输入疫情以来 境外输入确诊病例来源国（地区）情况 来源国（地区） 确诊病例数 英国 72 西班牙 47 意大利 22 美国 20 塞尔维亚 18 巴基斯坦 14 瑞典 14 希腊 11 阿联酋 10 法国 7 加拿大 7 土库曼斯坦 7 菲律宾 5 奥地利 5 匈牙利 4 伊朗 4 俄罗斯 4 泰国 3 瑞士 3 津巴布韦 3 韩国 3 印度尼西亚 2 巴布亚新几内亚 2 摩尔多瓦 2 丹麦 2 波兰 2 卢森堡 1 挪威 1 比利时 1 巴西 1 荷兰 1 墨西哥 1 约旦 1 赤道几内亚 1 日本 1 土耳其 1 中国香港 3 中国台湾 1 合计 307 点击进入专题：新型冠状病毒肺炎疫情 北京再现本土疫情 责任编辑：张建利'，'date': '2021年11月25日 07:51'，'source': '新浪新闻综合'，'keywords': '\n关键字：\n新冠肺炎北京'}

('title': '河北辛集：全面恢复市域内人员正常出行'，'content': '原标题：河北辛集：全面恢复市域内人员正常出行 据“辛集发布”消息，辛集市疫情防控工作指挥部发布关于进一步恢复市民正常出行秩序的公告。根据当前辛集市疫情防控形势，为统筹推进疫情防控和经济社会发展，按照《新

遍历新闻列表中的每一个 ul，并遍历其中的 li a 链接，并访问该链接的详情信息。

4、处理日期

```
In [162]: import pandas as pd
import re
df = pd.DataFrame(newsary)
# df.head(10)
df['date'].str.extract('(\d+\d+\d+ \d+:\d+)')
```

```
Out[162]:
0
0 2021年11月25日 16:35
1 2021年11月25日 16:32
2 2021年11月25日 16:16
3 2021年11月25日 16:03
4 2021年11月25日 15:49
5 2021年11月25日 15:36
6 2021年11月25日 15:36
7 2021年11月25日 15:01
8 2021年11月25日 14:41
9 2021年11月25日 12:29
10 2021年11月25日 12:01
11 2021年11月25日 11:36
12 2021年11月25日 08:47
13 2021年11月25日 08:34
```

```
In [163]: # 转换时间格式
df['date'] = pd.to_datetime(df['date'], format = '%Y年%m月%d日 %H:%M')
# df['date'].head()
df
```

```
Out[163]:
```

	title	content	date	source	keywords
0	澳大利亚考虑外交抵制北京冬奥会，外交部：不会得逞	原标题：澳大利亚考虑外交抵制北京冬奥会，外交部：不会得逞 11月25日，外交部发言人赵立坚...	2021-11-25 16:35:00	澎湃新闻	\n关键字：\n澳大利亚北京冬奥会奥林匹克冬奥会
1	国防部：美方所作所为是国际人权史上的耻辱	原标题：国防部：美方所作所为是国际人权史上的耻辱 11月25日下午，国防部举行例行记者会，国...	2021-11-25 16:32:00	环球网	\n关键字：\n国防部叙利亚
2	商务部：外资准入负面清单修订工作正抓紧推进	原标题：商务部：外资准入负面清单修订工作正抓紧推进 在今天（11月25日）商务部举行的例行新...	2021-11-25 16:16:00	央视网	\n关键字：\n商务部党中央国家发展改革委国务院
3	国防部批驳美方军事活动：不要撞了南墙也不回头	原标题：国防部批驳美方军事活动：不要撞了南墙也不回头 “国防部发布”微信公众号消息，11月2...	2021-11-25 16:03:00	界面新闻	NaN
4	廖建宇任四川省监察委员会副主任、代主任	原标题：廖建宇任四川省监察委员会副主任、代主任 11月25日，四川省第十三届人大常委会第三十...	2021-11-25 15:49:00	澎湃新闻	NaN
5	国防部：中俄两军关系没有最好，只有更好	原标题：国防部：中俄两军关系没有最好，只有更好 “国防部发布”微信公众号消息，11月25日下...	2021-11-25 15:36:00	界面新闻	NaN
6	国防部：美方在指责中国之前 应先认真反省一下自己	原标题：国防部：美方在指责中国之前 应先认真反省一下自己 11月25日，国防部例行记者会...	2021-11-25 15:00:00	环球网	NaN

把日期变为 datetime 格式。

5、处理空值

判断有空值的列，可以看到 keywords 列是有空值的。

```
In [127]: df.isnull().any()

Out[127]: title      False
content    False
date       False
source     False
keywords   True
dtype: bool
```

这里我把空值和非空值的行分为两组，进行分别处理，再把两组数据拼接起来，

带空值的行筛选

In [164]:

```
import numpy as np
# df['keywords'].fillna('暂无')
df1 = df[df['keywords'].isnull()] # 先把空值的行记录下来
df1
```

Out[164]:

	title	content	date	source	keywords
3	国防部批驳美方军事活动：不要撞了南墙也不回头	原标题：国防部批驳美方军事活动：不要撞了南墙也不回头 “国防部发布”微信公众号消息，11月2...	2021-11-25 16:03:00	界面新闻	NaN
4	廖建宇任四川省监察委员会副主任、代主任	原标题：廖建宇任四川省监察委员会副主任、代主任 11月25日，四川省第十三届人大常委会第三十...	2021-11-25 15:49:00	澎湃新闻	NaN
5	国防部：中俄两军关系没有最好，只有更好	原标题：国防部：中俄两军关系没有最好，只有更好 “国防部发布”微信公众号消息，11月25日下...	2021-11-25 15:36:00	界面新闻	NaN
6	国防部：美方在指责中国之前 应先认真反省一下自己	原标题：国防部：美方在指责中国之前 应先认真反省一下自己 11月25日下午，国防部举行例行记...	2021-11-25 15:36:00	环球网	NaN
8	处置实验室漏水立功！5名大学生被中科大奖励12万元	原标题：处置实验室漏水立功！5名大学生被中科大奖励12万元 记者从中国科学技术大学获悉，11...	2021-11-25 14:41:00	央视网	NaN
9	北京市发改委主任谈绪祥已任市政府党组成员	原标题：北京市发改委主任谈绪祥已任市政府党组成员 据微信公号“北京京剧院”11月24日消息：...	2021-11-25 12:29:00	新京报	NaN
10	河北衡水居民拨打12345热线反被训斥 涉事人员被停职	原标题：衡水高新区关于保利拉菲公馆小区某业主反映物业费问题的情况通报 来源：衡水高新区 对“...	2021-11-25 12:01:00	政府网站	NaN
14	我国成功发射试验十一号卫星	原标题：我国成功发射试验十一号卫星 北京时间2021年11月25日7时41分，我国在酒泉卫星...	2021-11-25 08:09:00	界面新闻	NaN
23	北京冬残奥会火炬接力将于明年3月2日—4日举行	原标题：北京冬残奥会火炬接力将于明年3月2日—4日举行 今天是北京冬残奥会开幕倒计时100天...	2021-11-24 20:19:00	央视网	NaN
26	印度人也生得少了！人口或于2060年后下降	原标题：印度人也生得少了！人口或于2060年后下降 据外媒24日报道，印度全国家庭健康调查最...	2021-11-25 16:24:00	红星新闻	NaN
29	俄克麦罗沃州一矿井发生冒烟事故 已致1人死	原标题：俄克麦罗沃州一矿井发生冒烟事故 已致1人死数十人伤 中新网11月25日电	2021-11-25	中国新闻	NaN

非空值行筛选

]:

```
df = df.dropna(axis = 0) # 筛选非空的行
df
```

29

	title	content	date	source	keywords
0	澳大利亚考虑外交抵制北京冬奥会，外交部：不会得逞	原标题：澳大利亚考虑外交抵制北京冬奥会，外交部：不会得逞 11月25日，外交部发言人赵立坚主...	2021-11-25 16:35:00	澎湃新闻	\n关键字：\n澳大利亚北京冬奥会奥林匹克冬奥会
1	国防部：美方所作所为是国际人权史上的耻辱	原标题：国防部：美方所作所为是国际人权史上的耻辱 11月25日下午，国防部举行例行记者会，国...	2021-11-25 16:32:00	环球网	\n关键字：\n国防部叙利亚
2	商务部：外资准入负面清单修订工作正抓紧推进	原标题：商务部：外资准入负面清单修订工作正抓紧推进 在今天（11月25日）商务部举行的例行新...	2021-11-25 16:16:00	央视网	\n关键字：\n商务部党中央国家发展改革委国务院
7	国防部：推动中美两军关系健康稳定地向前发展	原标题：国防部：推动中美两军关系健康稳定地向前发展 11月25日下午，国防部举行例行记者会，...	2021-11-25 15:01:00	国防部网站	\n关键字：\n国防部中美关系
11	大连发布返校复学预告	原标题：大连发布返校复学预告 来源：半岛晨报 根据全市疫情防控总体形势，如果社会面无新增新冠...	2021-11-25 11:36:00	新浪新闻综合	\n关键字：\n新冠肺炎
12	湖南11月24日新增境外输入无症状感染者1例	原标题：湖南11月24日新增境外输入无症状感染者1例 据湖南卫健委，2021年11月24日0...	2021-11-25 08:47:00	界面新闻	\n关键字：\n新冠肺炎湖南
13	张文宏团队发表最新研究 系中国研究人员首次	原标题：张文宏团队发表最新研究，系中国研究人员首次 据第一财经报道，11月23日，复旦大学附...	2021-11-25 08:34:00	每日经济新闻	\n关键字：\n新冠肺炎
15	安徽11月24日新增境外输入无症状感染者1例	原标题：安徽11月24日新增境外输入无症状感染者1例 据安徽卫健委通报，2021年11月24...	2021-11-25 08:08:00	界面新闻	\n关键字：\n新冠肺炎安徽
16	辽宁本土确诊0新增 本轮疫情首次无新增确诊	原标题：辽宁本土确诊0新增 辽宁本轮疫情首次无新增确诊 转发好消息！加油！ 11月24日0时...	2021-11-25 08:05:00	央视网	\n关键字：\n新冠肺炎辽宁
17	北京昨日新增2例境外输入无症状感染者	原标题：疫情发布 北京11月24日无新增新冠肺炎确诊病例 新增2例境外输入无症状感染者 ...	2021-11-25 07:51:00	新浪新闻综合	\n关键字：\n新冠肺炎北京
18	河北辛集：全面恢复市域内人员正常出行	原标题：河北辛集：全面恢复市域内人员正常出行 据“辛集发布”消息，辛集市疫情防控工作总指挥部...	2021-11-25 07:19:00	界面新闻	\n关键字：\n新冠肺炎河北
19	天宫TV上新！直击“80后”男航天员的在轨日常	原标题：天宫TV上新！ 直击“80后”男航天员的在轨日常 “我是03，我下次出舱，感觉良好！”...	2021-11-24 22:29:00	央视网	\n关键字：\n空间站
20	这次大规模农业种质资源普查 我们收获了什么	原标题：民生直通车 这次大规模农业种质资源普查，我们收获了什么 种业资源是国家的战略资源，...	2021-11-24 22:18:00	新华网	\n关键字：\n农业农村部

处理非空值行，把一些符号去掉，保留有用信息和格式

```
: # 处理非空行
df['keywords'] = df['keywords'].map(lambda x:x.split(':')[1].split())
df
```

	title	content	date	source	keywords
5	大连发布返校复学预告	原标题：大连发布返校复学预告 来源：半岛晨报 根据全市疫情防控总体形势，如果社会面无新增新冠...	2021-11-25 11:36:00	新浪新闻综合	新冠肺炎
6	湖南11月24日新增境外输入无症状感染者1例	原标题：湖南11月24日新增境外输入无症状感染者1例 据湖南卫健委，2021年11月24日0...	2021-11-25 08:47:00	界面新闻	新冠肺炎湖南
8	安徽11月24日新增境外输入无症状感染者1例	原标题：安徽11月24日新增境外输入无症状感染者1例 据安徽卫健委通报，2021年11月24...	2021-11-25 08:08:00	界面新闻	新冠肺炎安徽
9	辽宁本土确诊0新增 本轮疫情首次无新增确诊	原标题：辽宁本土确诊0新增 辽宁本轮疫情首次无新增确诊 转发好消息！加油！ 11月24日0时...	2021-11-25 08:05:00	央视网	新冠肺炎辽宁
10	北京昨日新增2例境外输入无症状感染者	原标题：疫情发布 北京11月24日无新增新冠肺炎确诊病例 新增2例境外输入无症状感染者 ...	2021-11-25 07:51:00	新浪新闻综合	新冠肺炎北京
11	河北辛集：全面恢复市域内人员正常出行	原标题：河北辛集：全面恢复市域内人员正常出行 据“辛集发布”消息，辛集市疫情防控工作总指挥部...	2021-11-25 07:19:00	界面新闻	新冠肺炎河北
12	天宫TV上新！直击“80后”男航天员的在轨日常	原标题：天宫TV上新！直击“80后”男航天员的在轨日常 “我是03，我下次出舱，感觉良好！”...	2021-11-24 22:29:00	央视网	空间站

填充空值

```
: df1 = df1.fillna('暂无')
df1
```

合并数据

```
: df = df.append(df1, ignore_index = True)
df
```

保存数据到 csv 文件

```
34]: df.to_csv('t.csv', encoding='utf-8-sig')
```


六、实验数据及处理结果

(请根据实际的数据来填写)

Lcsv - Excel										
文件 开始 插入 页面布局 公式 数据 审阅 视图 帮助 PDF工具集 百度网盘 操作说明搜索										
D13 2021/11/24 22:29:00										
	A	B	C	D	E	F	G	H	I	J
1		title	content	date	source	keywords				
2	0	澳大利亚考虑外交抵制北京冬奥	澳大利亚考虑外交抵制北京冬奥会，外交部：不会得逞 11月25日，外交部发	2021/11/25 16:35	澎湃新闻	[澳大利亚北京冬奥会奥林匹克冬奥会]				
3	1	国防部：美方所作所为是国际人权史上的耻辱	国防部：美方所作所为是国际人权史上的耻辱 11月25日下午，国防部举行例	2021/11/25 16:32	环球网	[国防部叙利亚]				
4	2	商务部：外资准入负面清单修订工作正抓紧推进	商务部：外资准入负面清单修订工作正抓紧推进 在今天（11月25日）商务	2021/11/25 16:16	央视网	[商务部党中央国家发展改革委员会]				
5	3	国防部：推动中美两军关系健康稳定向前发展	国防部：推动中美两军关系健康稳定向前发展 11月25日下午，国防部举行	2021/11/25 15:01	国防部网站	[国防部中美关系]				
6	4	大连发布返校复学预告	大连发布返校复学预告 来源：半岛晨报 根据全市疫情防控总体形势，如果社	2021/11/25 11:36	新浪新闻综合	[新冠肺炎]				
7	5	湖南11月24日新增境外输入无症状感染者1例	湖南11月24日新增境外输入无症状感染者1例 据湖南卫健委，2021年11月24	2021/11/25 8:47	界面新闻	[新冠肺炎湖南]				
8	6	张文宏团队发表最新研究	张文宏团队发表最新研究 系中国研究人员首次 据第一财经报道，11月23日	2021/11/25 8:34	每日经济新闻	[新冠肺炎]				
9	7	安徽11月24日新增境外输入无症状感染者1例	安徽11月24日新增境外输入无症状感染者1例 据安徽卫健委通报，2021年11	2021/11/25 8:08	界面新闻	[新冠肺炎安徽]				
10	8	辽宁本土确诊0新增 本轮疫情首次无新增确诊	辽宁本土确诊0新增 辽宁本轮疫情首次无新增确诊 转发好消息！加油！ 11月	2021/11/25 8:05	央视网	[新冠肺炎辽宁]				
1	9	北京昨日新增2例境外输入无症病例	疫情发布 北京11月24日无新增新冠肺炎确诊病例 新增2例境外输入无症病	2021/11/25 7:51	新浪新闻综合	[新冠肺炎北京]				
2	10	河北辛集：全面恢复市内人员正常出行	河北辛集：全面恢复市内人员正常出行 据辛集发布“消息，辛集市疫情防	2021/11/25 7:19	界面新闻	[新冠肺炎河北]				
3	11	天宫TV上新！直击“80后”男航天员	天宫TV上新！直击“80后”男航天员的在轨日常“我是03，我下出舱，感觉良	2021/11/24 22:29	央视网	[空间站]				
4	12	这次大规模农业种质资源普查	民生直通车 这次大规模农业种质资源普查，我们收获了什么 农业种质资源	2021/11/24 22:18	新华网	[农业农村部]				
5	13	辽宁大连市将4个中风险地区调整为低风险地区	辽宁大连市将4个中风险地区调整为低风险地区 11月24日大连市新冠肺炎疫	2021/11/24 22:10	央视网	[辽宁省大连市风险地区低风险新冠肺炎]				
6	14	根治欠薪工作仍存问题	根治欠薪工作仍存问题 国务院约谈有关地方负责人 中新网11月14日电 据人	2021/11/24 22:09	中国新闻网	[国务院党中央农民工资]				
7	15	11月24日外交部例行记者会	2021年11月24日外交部发言人赵立坚主持例行记者会一、应柬埔寨首相洪森	2021/11/24 20:17	外交部网站	[赵立坚外交部发言人上合组织]				
8	16	索马里首都发生自杀式汽车炸弹袭击	索马里首都发生自杀式汽车炸弹袭击致9死17伤 新华社内罗毕11月25日电 (2021/11/25 17:02	新华网	[索马里汽车炸弹首都内罗毕摩加迪沙]				
9	17	美媒：德国成最新一个“新冠”原	美媒：德国成最新一个“新冠”死亡人数超10万人“国家（环球网报道）实	2021/11/25 15:42	环球网	[德国美联社新冠肺炎]				
10	18	美媒民调：57%美国人认为政府应	美媒民调：57%美国人认为政府应为通胀问题负责 资料图 海外网11月25日电	2021/11/25 15:31	海外网	[美国人]				
1	19	美国将出售3200万桶战略储备原油	美国将出售3200万桶战略储备原油 以降低油价 资料图：美国原油价格上涨	2021/11/25 13:39	海外网	[原油美国原油储备]				
2	20	埃塞俄比亚总理自曝要亲赴前线	埃塞俄比亚总理自曝要亲赴前线，美国警告：埃局势“无军事解决方案” [环	2021/11/25 11:04	环球网	[埃塞俄比亚埃塞俄比亚总理美国埃塞国防军]				
3	21	外媒：美国政府欲说服国会取消对	外媒：美国政府欲说服国会取消对“北溪-2”部分制裁“北溪-2”天然气管道项	2021/11/25 10:42	海外网	[美国政府国会俄罗斯制裁]				
4	22	元宇宙尽头在哪里 科技巨头们用	元宇宙尽头在哪里，科技巨头们用脚投票 一千个科技从业者眼中有一千个元	2021/11/25 8:19	新京报	[美元]				
5	23	克宫介绍普京接种鼻喷新冠疫苗情	克宫介绍普京接种鼻喷新冠疫苗情况：他接种后体温正常，无不良反应 俄罗	2021/11/25 8:19	环球网	[新冠疫苗佩斯科夫普京疫苗俄罗斯总统新冠肺炎]				
6	24	伊拉克军方：驻伊外国军队作战人	伊拉克军方：驻伊外国军队作战人员将于15天内撤离 伊拉克军方11月24日表	2021/11/25 8:08	央视网	[伊拉克伊朗美国]				
7	25	普京：我接种疫苗整整6个月体内	普京：我接种疫苗整整6个月体内抗体滴度下降，专业人士建议接种加强针	2021/11/25 7:51	环球网	[普京疫苗新冠疫苗新冠肺炎]				
8	26	元宇宙热潮：一场既有技术突破也	元宇宙热潮：一场既有技术突破也有泡沫的游戏 从某种意义上来说，元宇	2021/11/25 7:42	澎湃新闻	[技术突破]				
9	27	载34人移民船只穿越英吉利海峡时	载34人移民船只穿越英吉利海峡时倾覆，31人死亡，2人获救，1人下落不明	2021/11/25 7:08	环球网	[移民法国欧盟路透社]				
10	28	任内第二次！意大利总理是否原	意大利总理任内第二次否决中企收购案，又拿半导体行业安全当借口 [环球	2021/11/25 6:51	环球网	[意大利意大利总理美国浙江省德拉克]				
1	29	印度女子指控家人与警察等合谋绑	印度女子指控家人与警察等合谋绑架自己孩子送出寄养，当地法院作出判	2021/11/25 6:51	环球网	[印度喀拉拉邦法院]				
2	30	局势不断升级 北约用“蟒蛇”战	北约用“蟒蛇”战术围困白俄罗斯，俄回应：俄白将继续对西方挑衅行为作出	2021/11/25 6:50	环球网	[白俄罗斯北约]				
3	31	国防部批驳美方军事活动：不要撞	国防部批驳美方军事活动：不要撞了南墙不回头“国防部发布“微信公众	2021/11/25 16:03	界面新闻	暂无				
4	32	廖建宇任四川省监察委员会副主	廖建宇任四川省监察委员会副主任、代主任 11月25日，四川省第十三届人	2021/11/25 15:49	澎湃新闻	暂无				
5	33	国防部：中俄两军关系没有最好，	国防部：中俄两军关系没有最好，只有更好“国防部发布“微信公众消息，	2021/11/25 15:36	界面新闻	暂无				
6	34	国防部：美方在指责中国之前应先	国防部：美方在指责中国之前应先认真反省一下自己 11月25日下午，国防	2021/11/25 15:36	环球网	暂无				
7	35	处置实验室漏水立功！5名大	处置实验室漏水立功！5名大学生被中科院大奖励12万元 记者从中国科学	2021/11/25 14:41	央视网	暂无				
8	36	北京市发改委主任谈续任已任原	北京市发改委主任谈续任已任原北京市政府党组成员 据微信公众号“北京	2021/11/25 12:29	新京报	暂无				
9	37	河北衡水居民拨打12345热线	衡水高新区关于保利拉菲公馆小区业主反映物业费问题的情况通报 来源：	2021/11/25 12:01	政府网站	暂无				
10	38	我国成功发射试验十一号卫星	我国成功发射试验十一号卫星 北京时间2021年11月25日7时41分，我国在	2021/11/25 8:09	界面新闻	暂无				

七、思考讨论题或体会或对改进实验的建议

(请根据实际的数据来填写)

如果想做有关于数据处理的实验，需要找到适合的网站，但是现在网站大部分比较完善，很难找到那种可以锻炼的网页，建议实机演示比较好。

八、参考资料

- [1] 朝乐门 著. Python 编程 从数据分析到数据科学. 电子工业出版社. 2019.1
- [2] [美] 阿曼多·凡丹戈 (Armando Fandango) 著, 韩波 译. Python 数据分析(第2版). 人民邮电出版社. 2018.6
- [3] 嵩天, 礼欣, 黄天羽 著. Python 语言程序设计基础 (第2版). 高等教育出版社. 2017.2