

Python 程序设计实验报告

PB20111686 黄瑞轩

Level 1

Level 1.7【统计】时序统计

COVID-19 是由 SARS-CoV-2 导致的传染性疾病，常见症状包括发烧，咳嗽和呼吸急促。现有 [全球疫情统计结果](#)，以数据 `04-01-2022.csv` 为例，编写一个程序，输出确诊人数（Confirmed）最多的 10 个国家详细数据（国家 Country/Region，最新更新时间 Last Update，确认病例 Confirmed，死亡病例 Deaths，恢复病例 Recovered）

思路：阅读 csv 每一行，提取需要输出的数据栏，列表排序，写入 csv 文件。

注意：

- 阅读时同名国家需要合并数据；
- csv写入时，有一个默认的newline参数，如果需要像普通的表格那样，需要设置 `newline=""`。

初始文件：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	FIPS	Admin2	Province_S	Country_R	Last_Update	Lat	Long	Confirmed	Deaths	Recovered	Active	Combined Incident_R	Case_Fatality_Ratio		
2				Afghanistan	2022/4/2 4:20	33.93911	67.70995	177782	7670			Afghanistan	456.6904	4.314273	
3				Albania	2022/4/2 4:20	41.1533	20.1683	273759	3492			Albania	9512.788	1.275575	
4				Algeria	2022/4/2 4:20	28.0339	1.6596	265679	6874			Algeria	605.867	2.587333	
5				Andorra	2022/4/2 4:20	42.5063	1.5218	40024	153			Andorra	51800.94	0.382271	
6				Angola	2022/4/2 4:20	-11.2027	17.8739	99194	1900			Angola	301.811	1.915438	
7				Antarctica	2022/4/2 4:20	-71.9499	23.347	11	0			Antarctica		0	
8				Antigua ar	2022/4/2 4:20	17.0608	-61.7964	7493	135			Antigua ar	7651.54	1.801682	
9				Argentina	2022/4/2 4:20	-38.4161	-63.6167	9039838	128052			Argentina	20001.51	1.41653	
10				Armenia	2022/4/2 4:20	40.0691	45.0382	422540	8616			Armenia	14259.42	2.039097	
11			Australian	Australia	2022/4/2 4:20	-35.4735	149.0124	79515	42			Australian	18573.93	0.05282	
12			New South	Australia	2022/4/2 4:20	-33.8688	151.2093	1863375	2444			New South	22953.62	0.13116	
13			Northern T	Australia	2022/4/2 4:20	-12.4634	130.8456	50713	36			Northern T	20648.62	0.070988	
14			Queenslan	Australia	2022/4/2 4:20	-27.4698	153.0251	785417	746			Queenslan	15353.67	0.094981	
15			South Aust	Australia	2022/4/2 4:20	-34.9285	138.6007	263402	261			South Aust	14995.84	0.090988	
16			Tasmania	Australia	2022/4/2 4:20	-42.8821	147.3272	93192	31			Tasmania,	17402.8	0.033265	
17			Victoria	Australia	2022/4/2 4:20	-37.8136	144.9631	1302047	2752			Victoria, Ai	19635.51	0.211359	
18			Western Ai	Australia	2022/4/2 4:20	-31.9505	115.8605	190502	53			Western Ai	7241.77	0.027821	
19			Austria		2022/4/2 4:20	47.5162	14.5501	3854405	15905			Austria	42796.29	0.412645	
20			Azerbaijan		2022/4/2 4:20	40.1431	47.5769	791979	9694			Azerbaijan	7811.079	1.224022	
21			Bahamas		2022/4/2 4:20	25.02589	-78.0359	33291	788			Bahamas	8465.65	2.367006	
22			Bahrain		2022/4/2 4:20	26.0275	50.55	555003	1471			Bahrain	32616.86	0.265044	
23			Bangladesl		2022/4/2 4:20	23.685	90.3563	1951658	29122			Bangladesl	1185.054	1.492167	
24			Barbados		2022/4/2 4:20	13.1939	-59.5432	59495	375			Barbados	20703.2	0.630305	
25			Belarus		2022/4/2 4:20	53.7098	27.9534	963756	6829			Belarus	10199.21	0.708582	
26			Antwerp	Belgium	2022/4/2 4:20	51.2195	4.4024	608444	0			Antwerp, B	32747.5	0	
27			Brussels	Belgium	2022/4/2 4:20	50.8503	4.3517	428749	0			Brussels, Br	35476.55	0	
28			East Flandr	Belgium	2022/4/2 4:20	51.0362	3.7373	505555	0			East Flandr	33368.56	0	
29			Flemish Br	Belgium	2022/4/2 4:20	50.9167	4.5833	364656	0			Flemish Br	31815.04	0	
30			Hainaut	Belgium	2022/4/2 4:20	50.5257	4.0621	431259	0			Hainaut, Br	32081.97	0	

获得文件：

	A	B	C	D	E	F	G	H
1	Rank	Country_R	Last_Update	Confirmed	Deaths	Recovered		
2	1	US	2022/4/2 4:20	80141546	981701	0		
3	2	India	2022/4/2 4:20	43027035	521264	0		
4	3	Brazil	2022/4/2 4:20	29978513	660265	0		
5	4	France	2022/4/2 4:20	25952226	143441	0		
6	5	United Kin	2022/4/2 4:20	21379545	166168	0		
7	6	Germany	2022/3/19 23:20	21357039	129695	0		
8	7	Russia	2022/4/2 4:20	17601907	361679	0		
9	8	Turkey	2022/4/2 4:20	14873927	98078	0		
10	9	Italy	2022/4/2 4:20	14719394	159537	0		
11	10	Korea, Sou	2022/4/2 4:20	13639915	16929	0		
12								
13								
14								
15								
16								
17								
18								

Level 2

Level 2.13 【计算机】

基于 Python，使用浏览器引擎 Selenium 爬取科大首页中的科大要闻第一条新闻，要求爬取结果至少有新闻标题、时间、以及新闻内容。

思路：Selenium 使用 Chrome 来做 (chromedriver)，通过 Selenium 的 `find_element_by_xpath` 方法找到对应的内容。

关键代码：

```
browser = webdriver.Chrome()
news = []

browser.get("http://ustc.edu.cn/")
browser.find_element_by_xpath("/html/body/main/div[1]/section[1]/ul/li[1]/div[2]/h4[1]/a").click()
news.append(browser.find_element_by_xpath("/html/body/main/div/div/div[1]/form/div[1]").text)
news.append(browser.find_element_by_xpath("/html/body/main/div/div/div[1]/form/div[2]").text)
news.append(browser.find_element_by_xpath('//*[@id="vsb_content_2"]/div').text)
```

爬取结果：

```
['2022年05月21日', '中国科大举办2022科技活动周开幕式暨全国科普教育基地揭牌仪式', '5月21日上午，中国科大2022科技活动周开幕式暨全国科普教育基地揭牌仪式在东区水上报告厅举行。党委书记舒歌群、校长包信和、党委常务副书记蒋一出席仪式，党委常委、副校长罗喜胜主持仪式。相关职能部门、院系负责人、部分感兴趣师生参加仪式。\\n罗喜胜介绍了学校2022科技活动周的组织筹备情况。他表示，为加快推进科技规划政策扎实落地，广泛宣传科技创新成果，开展科学普及惠民活动，根据科技部、中宣部、中国科协及中科院总体部署，学校于5月21日举办科技活动周暨第十八届公众科学日活动。本次活动采用线上线下相结合、线上即时互动的方式，让更多观众享受这一科学盛宴。\\n日前中国科协公布《2021-2025年第一批全国科普教育基地名单》，我校国家同步辐射实验室和火灾科学国家重点实验室入选。舒歌群、包信和、国家同步辐射实验室书记李良彬、火灾科学国家重点实验室主任刘乃安共同为全国科普教育基地揭牌。\\n舒歌群在致辞中强调，值此第二个百年奋斗目标新征程扬帆起航之际，要深入学习领会习近平总书记关于科技创新系列重要指示精神，坚持“四个面向”，认真贯彻落实习近平总书记“科技创新与科学普及两翼齐飞”的重要论述精神，深刻把握全民科学素质建设的责任使命，推动在全社会形成讲科学、爱科学、学科学、用科学的良好氛围。他表示，学校将办好此次科技活动周作为深入贯彻《科技进步法》《科普法》、坚决落实党中央关于科技自立自强的战略部署、推进科技强国建设的一项重要任务抓好抓实，精心组织筹办27个科普点共105个方案和九场大型科普报告，科普点数、活动方案数均达到历史新高。舒歌群指出，要以举办此次科技活动周为契机，进一步充分发挥我校在教学、科研、人才、基地方面的综合优势，大力培育青少年尊崇科学的人生价值观，激发青少年热衷科学探索的兴趣爱好，引导青少年树立实现高水平科技自立自强的理想志向，在青春的赛道上奋力奔跑，跑出当代青年的最好成绩。\\n\\n(党委宣传部)']
```

存在 `news` 列表中，`news[0]` 为时间，`news[1]` 为标题，`news[2]` 为内容。

Level 3

Level 3.6 【物理】

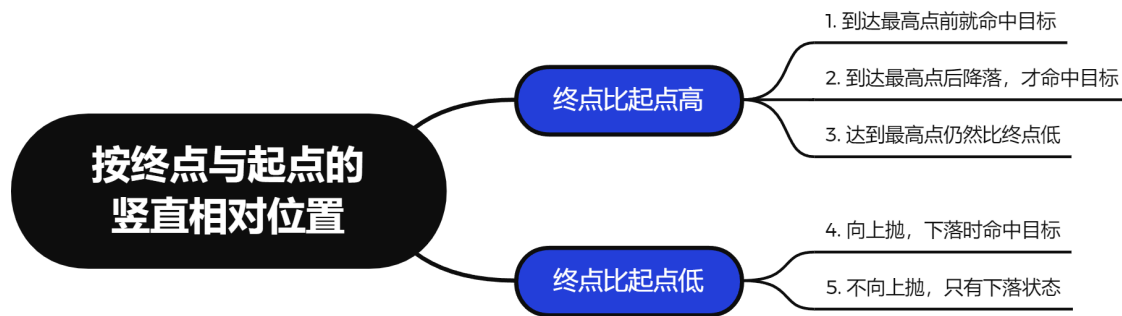
接【数学】Level 2.3 野外彩弹 CS 游戏，考虑重力和空气阻力影响，给定发射点坐标 (x_0, y_0, z_0) ，落点坐标 (x_1, y_1, z_1) ，空气阻力 f (矢量，以 (f_x, f_y, f_z) 表示)，及彩弹的初速度大小 v_0 (标量数值)，使用 numpy、math 库等工具求解满足给定起点、落点、空气阻力和初速的发射仰角。

提示 1：默认 $g = 9.8$ 为常数；如果初速度过小，解可能不存在

提示 2：弹道计算机的雏形

思路：

问题的关键是不是否**上抛**。从上抛到下落的过程会导致加速度变化，是这个问题的主要难点。在竖直方向，向上运动的加速度为 $g + fz/m$ (向下)，向下运动的加速度为 $g - fz/m$ (定向下)，须知空气阻力**不是动力**。其不可能导致原本向下运动的物体向上运动，所以可以认为向下运动的物体受到的 z 方向加速度仍然是向下的。



这里列出 5 种可能的情况, 我们发现情况 1 和 5 是比较特殊的, 情况 2 ~ 4 都是上抛再降落的情况, 特判讨论即可。

(1) 到达最高点前就命中目标

在竖直方向, 有 $z1 - z0 = vzt - \frac{1}{2}(g + fz/m)t^2$

解之, 得 $t = \frac{vz - \sqrt{vz^2 - 2(g + fz/m)(z1 - z0)}}{g + fz/m}$

此时, 对于 x, y 方向, 约束条件: $x1 - x0 = vxt - \frac{1}{2}(fx/m)t^2$, $y1 - y0 = vyt - \frac{1}{2}(fy/m)t^2$

为了操作方便, 将 t 作为一个辅助函数:

```
def t(vz, fz, z1, z0):  
    return ...
```

于是约束函数为:

$$L = [vxt - \frac{1}{2}(fx/m)t^2 - (x1 - x0)]^2 + [vyt - \frac{1}{2}(fy/m)t^2 - (y1 - y0)]^2 + [\sqrt{vx^2 + vy^2 + vz^2} - v0]^2$$

这里需要捕捉的异常为: t 是否是负数

(5) 不向上抛, 只有下落状态

在竖直方向, 有 $z1 - z0 = vzt - \frac{1}{2}(g - fz/m)t^2$

解之, 得 $t = \frac{vz - \sqrt{vz^2 - 2(g - fz/m)(z1 - z0)}}{g - fz/m}$

此时, 对于 x, y 方向, 约束条件: $x1 - x0 = vxt - \frac{1}{2}(fx/m)t^2$, $y1 - y0 = vyt - \frac{1}{2}(fy/m)t^2$

于是约束函数为:

$$L = [vxt - \frac{1}{2}(fx/m)t^2 - (x1 - x0)]^2 + [vyt - \frac{1}{2}(fy/m)t^2 - (y1 - y0)]^2 + [\sqrt{vx^2 + vy^2 + vz^2} - v0]^2$$

合适的话有解, 无解的情况比较复杂, 不做讨论了。

(2 ~ 4) 诸情况

上抛过程: $t1 = vz/(g + fz/m)$, 高度 $h = vz^2/2(g + fz/m)$

此时需要判断是否是情况 (3)。

下落过程: $z0 + h - z1 = \frac{1}{2}(g - fz/m)t2^2$, 可以求出 $t2 = \sqrt{2(z0 - z1 + h)/(g - fz/m)}$ 。

此时, 对于 x, y 方向, 约束条件: $x1 - x0 = vxt - \frac{1}{2}(fx/m)t^2$, $y1 - y0 = vyt - \frac{1}{2}(fy/m)t^2$

于是约束函数为:

$$L = [vxt - \frac{1}{2}(fx/m)t^2 - (x1 - x0)]^2 + [vyt - \frac{1}{2}(fy/m)t^2 - (y1 - y0)]^2 + [\sqrt{vx^2 + vy^2 + vz^2} - v0]^2$$

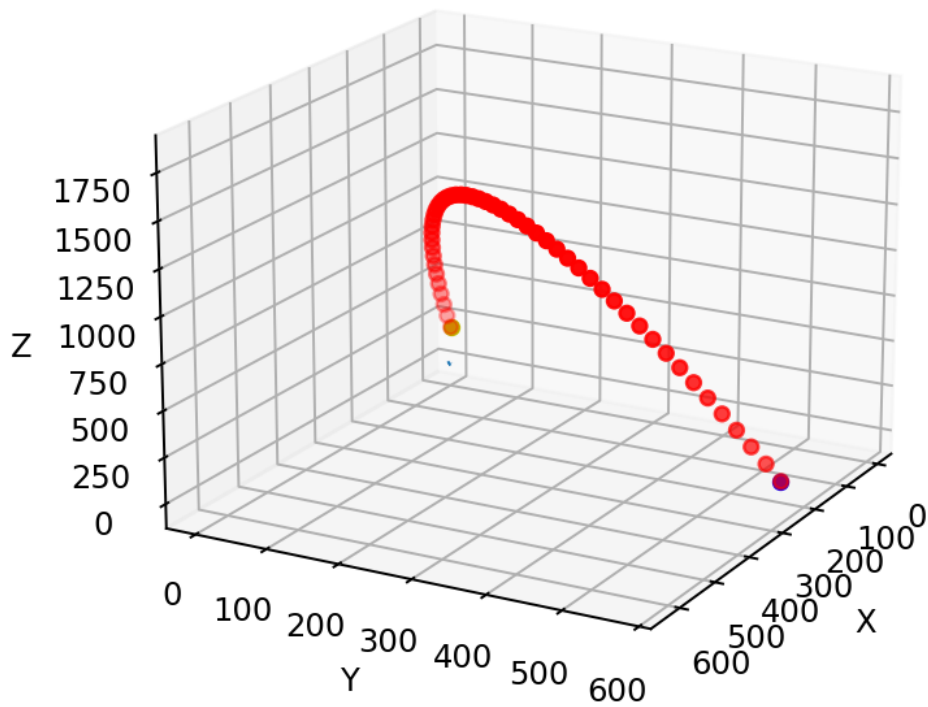
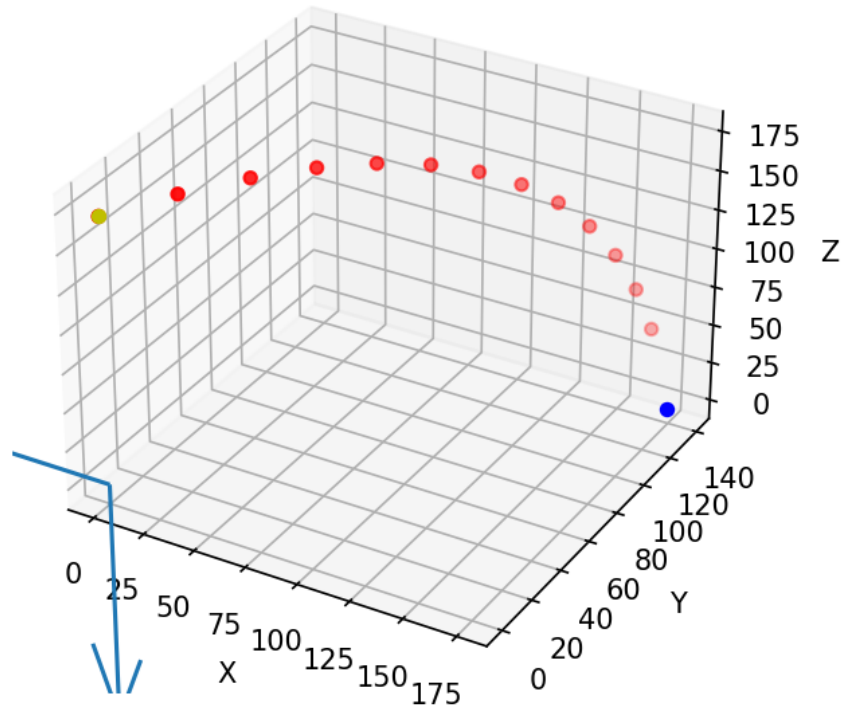
这里 $t = t1 + t2$, 合适的话有解, 无解的情况比较复杂, 不做讨论了。

如果经过以上三种情况判断后都没有解，程序就判断没有解。

按照计算出的 (vx, vy, vz) 来做时序模拟，所得结果如下所示。黄色点是初始位置，蓝色点是末了位置。

这里时序模拟用到了 `matplotlib` 和 `numpy` 库。

3d_image_show



Level 4

Transformers

Transformers 提供了数千种预训练模型来执行文本相关任务，如分类、信息提取、问答、摘要、翻译、文本生成等。github 中提供了相应的api，可以在给定文本上快速下载和使用这些预训练模型，在自己的数据集上对它们进行微调。

情感分析

关键代码如下：

```
from transformers import pipeline
import csv

c = pipeline('sentiment-analysis')
# 股票明天下跌
print(c("Stocks will fall tomorrow."))
# 明天肯定又会出太阳
print(c("It's bound to be sunny again tomorrow."))
# 日复一日做一样的事情想到明天还要重复做肯定很崩溃
print(c("every single day It must be overwelming to think tomorrow I have to do it over again."))
```

测试结果：

```
[{'label': 'NEGATIVE', 'score': 0.9988724589347839}]
[{'label': 'POSITIVE', 'score': 0.9996645450592041}]
[{'label': 'NEGATIVE', 'score': 0.9987743496894836}]
```

测试结果基本符合事实。

有道翻译API (request 库)

向有道翻译的 API 发起 request 请求，实现汉译英效果，就像在 GUI 页面进行翻译业务那样。

输入：

```
if __name__ == '__main__':
    src, tgt = translate('股票明天将会下跌')
    print(src) # 输入
    print(tgt) # 结果
```

输出：

```
股票明天将会下跌
The stock will fall tomorrow
```

Selenium

这个库的使用已经在 Level 2 中试过，这里不再赘述。

Level 5

基金吧舆论情感走势与基金净值走势的相关检验

背景：随着支付宝、东方财富网购买基金功能的完善，越来越多的大学生加入了炒基金的行列。和传统炒股、炒基金模式不同的是，现在购买基金的 APP 往往提供“评论区”功能，用户可以发帖、回帖、分享自己的见解。Level 5 中，我将综合利用爬虫、情感分析和其他图形库技术，对东方财富网基金吧内舆论进行情感走势与基金净值走势的相关检验。

Step1: 分析基金吧网页结构

基金吧 url: `http://guba.eastmoney.com/list,of{ID}_{Page}.html` , 其中 `{ID}` 表示基金代码, 如果不是第一页, 则需要加上 `{Page}` , `{Page}` 表示当前页号。

最大页号 xpath: `//*[@id="articlelistnew"]/div[82]/span/span/span/span`

标题 xpath 构成:

- 第一个帖子的 xpath: `//*[@id="articlelistnew"]/div[2]/span[3]/a/text()`
- 最后一个帖子的 xpath: `//*[@id="articlelistnew"]/div[81]/span[3]/a/text()`

很容易看出遍历帖子所需要的变量。

发帖时间 xpath: `//*[@id="articlelistnew"]/div[2]/span[5]` , 这里的 2 处为帖子位置。

Step2: 分析基金净值网页结构

爬基金净值的网页 url: `https://www.dayfund.cn/fundvalue/{ID}.html` , 其中 `{ID}` 表示基金代码。

日期 xpath: `//*[@id="his_nav_table"]/tbody/tr[1]/td[1]`

增长率 xpath: `//*[@id="his_nav_table"]/tbody/tr[1]/td[4]`

Step3: 封装基金吧网页、基金净值网页对象

先简单将翻译功能封装成 `translate` 模块, 这里不赘述。

基金吧网页支持如下方法:

```
class talk_bar:
    url = ''
    max_page = 0

    def __init__(self, id):
        # 按基金编码来初始化url, 初始化翻到第一页
        # 获取最大页数

    def change_page(self, page):
        # 翻到基金吧第page页

    def get_title(self, num):
        # 获取第num个标题内容和时间
        # 返回值是一个二元素列表, 第0个元素是时间, 第1个元素是内容
```

基金净值网页支持如下方法:

```
class value:
    url = ''

    def __init__(self, id):
        # 按基金编码来初始化url

    def change_page(self, page):
        # 翻到第page页

    def get_rate(self, num):
        # 获取表格第num行的内容
        # 返回值是一个二元素列表, 第0个元素是时间, 第1个元素是当日净值增长率
```

这里列出的也是**伪代码**。实际情况中因为对每一页的访问量, 并且 API 翻译需要时间, 因此可以对每一页的 html 页面进行缓存, 以减少 `request` 次数, 这里没有列出具体方法, 实现详情见代码。

这样, 只需要调用


```
import url

test = url.talk_bar('161725')
test2 = url.value('161725')

for i in range(1, 61):
    pr2 = test.get_title(str(i))
    pr = test2.get_rate(str(i))
    if pr2 is not None:
        print(pr2)
    # if pr is not None:
    #     print(pr)
```

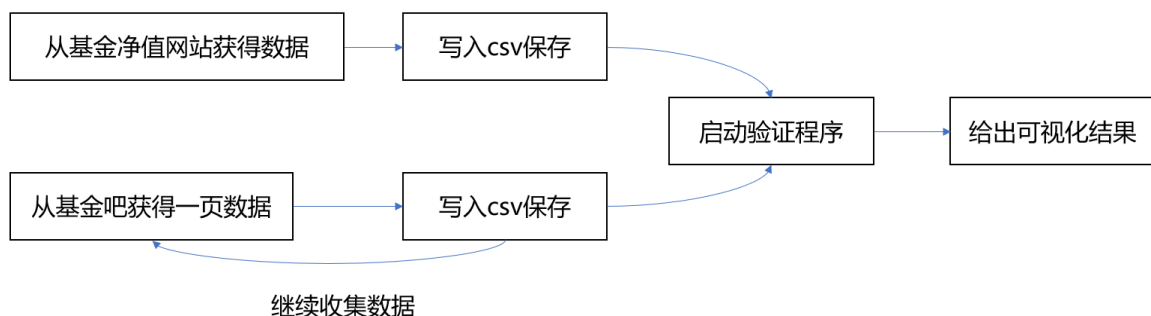
就可以相当方便的给出我们需要的数据。

```
['06-06 06:15', '用行动赢取“苹果华为”礼包']
['06-05 17:21', '【讲故事瓜分2w】我和我的新能源故事']
['06-03 11:51', '【30000惊喜好礼】鑫元中短债夏日派对']
['06-06 11:17', '为什么说这两个板块有望迎来爆发行情?']
['06-06 11:04', '策略:确定了,明天A股走势推演已出炉,请尽快查收!']
['06-06 11:03', '重磅美股纳指大跌超2%,下周A股何去何从?(附策略)']
['06-06 11:00', '策略]端午节后A股会涨吗?直接告诉你']
['06-06 10:42', '市场反弹后感受也好了,而不买会踏空吗?告诉你布局逆']
['06-06 11:30', '0606猴哥点评:科创量价齐升再度大涨近3%,锂电板块持']
['06-06 11:30', '白酒今天百分百大跌[大笑][大笑][大笑][大笑][大笑]']
['06-06 11:29', '给大家的一封信:']
['06-06 11:28', '$白酒基金LOF(SZ161725)$跑不赢大盘还']
['06-06 11:27', '观点鲜明:2022年6月6日,各板块基金,今天的操作建议']
['06-06 11:27', '新能源带涨创业板,踏空的朋友要追吗?']
['06-06 11:27', '拍断大腿,触底反弹又是新能源,你们别买让我来[兴奋]']
['06-06 11:26', '不要和趋势作对,趋势向下买易亏。趋势向上卖易飞,顺']
```

```
['2022-06-02', '-0.77%']
['2022-06-01', '-0.53%']
['2022-05-31', '3.21%']
['2022-05-30', '3.29%']
['2022-05-27', '0.52%']
['2022-05-26', '0.04%']
['2022-05-25', '-0.52%']
['2022-05-24', '-1.78%']
['2022-05-23', '-1.52%']
['2022-05-20', '3.96%']
['2022-05-19', '-0.91%']
['2022-05-18', '-0.87%']
['2022-05-17', '0.87%']
['2022-05-16', '-0.82%']
['2022-05-13', '-0.36%']
['2022-05-12', '1.01%']
```

Step4: 各种库模块封装以及数据流传送设计

数据流传送图:



可见写入和验证程序的读取是比较固定的模式，把这个操作封装成库（与 numpy 及 matplotlib 等库结合）。

数据命名规范：

- 基金净值：{ID}_jz_{create_time}.csv，其中 ID 为基金代号，create_time 是一个 yyyy-mm-dd 形状的导出时间。
- 基金吧数据：{ID}_tk_{page}_{create_time}，其中 ID 为基金代号，page 是导出时此页为第几页，create_time 是一个 yyyy-mm-dd 形状的导出时间。

这样，用户就可以无需看到复杂的内部结构，直接初始化（这里以招商中证白酒指数分级基金：161725 为例）

```
from time import sleep
import mycsv

test = mycsv.collector('161725')

test.get_jz()
sleep(5)
test.get_tk('1')
```

就可以获得下面的结果：

	A	B	C	D			A	B	C	D	E	F	G
1	date	rate				1	date	title					
2	2022/6/2	-0.77%				2	2022/6/6 6:15	用行动赢取“苹果华为”礼包					
3	2022/6/1	-0.53%				3	2022/6/6 14:33	【30000惊喜好礼】鑫元中短债夏日派对					
4	2022/5/31	3.21%				4	2022/6/5 17:21	【讲故事瓜分2w】我和我的新能源故事					
5	2022/5/30	3.29%				5	2022/6/6 16:17	策略 端午节后A股会涨吗？直接告诉你					
6	2022/5/27	0.52%				6	2022/6/6 16:16	刚要入睡，证券市场又传重磅利好，6月准备迎接大行情					
7	2022/5/26	0.04%				7	2022/6/6 16:05	重磅美股纳指大跌超2%，下周A股何去何从？（附策略）					
8	2022/5/25	-0.52%				8	2022/6/6 16:00	A股行情普涨突破3200点，应证提示小心踏空					
9	2022/5/24	-1.78%				9	2022/6/6 14:21	6.05复盘：A股估值处于历史低位，白酒军工新能源怎么看？					
10	2022/5/23	-1.52%				10	2022/6/6 16:36	周末消息面总结解读和下周A股操作思路					
11	2022/5/20	3.96%				11	2022/6/6 16:36	A股出现逆转信号，牛市的要来了吗，下半年准备扬眉吐气					
12	2022/5/19	-0.91%				12	2022/6/6 16:33	百万基金实盘：小高理财6月6日投资分享，成功预判节后上涨行情					
13	2022/5/18	-0.87%				13	2022/6/6 16:33	生活中不可避免的思维盲点——归因谬误（上）					
14	2022/5/17	0.87%				14	2022/6/6 16:32	复盘：确定了，A股开始加速上涨，千万要踏准节奏！					
15	2022/5/16	-0.82%				15	2022/6/6 16:32	继续优化仓位结构，A股形成向上突破的趋势雏形					
16	2022/5/13	-0.36%				16	2022/6/6 16:29	我用的闲钱，不挣十几点不会走，现在又不是高位[呲牙][呲牙]					
17	2022/5/12	1.01%				17	2022/6/6 16:29	周线级别金叉已经形成，大盘波段反弹已经开启					
18	2022/5/11	1.40%				18	2022/6/6 16:28	A股基金收评 A股6月上涨预测成功 下一步我们这样做					
19	2022/5/10	1.55%				19	2022/6/6 16:26	假期过后大盘又迎来开门红，沪指突破3200点后该如何走					
20	2022/5/9	-2.11%				20	2022/6/6 16:22	今天想必大家是开心的，又是一个吃大肉的行情，特别是新能源和					
21	2022/5/6	-3.80%				21	2022/6/6 16:22	收评 市场如期迎来“报复性”反弹，可从防守逐步转向进攻？市场行					
22	2022/5/5	0.17%				22	2022/6/6 16:19	不要和趋势作对，趋势向下买易亏。趋势向上卖易飞，顺势而为利					
23	2022/4/29	0.55%				23	2022/6/6 16:17	百万实盘：大头6月6日投资分享，科创板有望迎来牛市？					
24	2022/4/28	2.66%				24	2022/6/6 16:16	盈亏日记 今天基金盈利接近1.8万，节后有望延续反弹					
25	2022/4/27	2.53%				25	2022/6/6 16:16	真正穿越牛熊的中长期优质基金！（附详细名单）					

拿到数据，就可以来做数据检验了。

这里不把翻译后的数据放进来，是因为翻译的结果要消耗大量请求，极易发生错误，本人认为保存原始数据的操作越安全越好。

Step5: