

项目说明文档

1.环境说明

a.数据库

使用MySQL数据库，自主安装配置好数据库后，在终端输入指令：

```
mysql -u root -p
```

随后输入密码，并将database.sql代码复制到终端并回车

对于本项目，共设计了三个表：**帖子信息表（posts）**、**评论信息表（comments）**、**用户表（users）**

具体数据库表信息请见：[database.sql](#)

b.python依赖库

使用anaconda虚拟环境，请先下载anaconda，后创建新环境，安装本项目所需的python库

```
pip install -r requirement.txt
```

c.前后端

前端使用html，后端基于flask框架。

2.各代码说明

jrtt.py

爬虫代码，提供了特定（自选）主题的今日头条内容的爬取，对每个帖子的帖子号、标题、摘要、发布时间、发布人、相关评论（评论人、评论id、评论内容、评论发布时间等）以及发布人的粉丝概况等信息进行爬取。以下具体介绍

1. 类和初始化部分

```
class Jrtt():
    def __init__(self):
        super().__init__()
        self.file_name = os.path.basename(__file__).replace('.py', '')
        self.log = self.init_logger(self.file_name)
        self.headers = {...}
        self.cookies = {...}
        self.init_db()
```

- 以类封装爬虫，符合面向对象设计；
- 初始化日志模块，方便后续信息记录；
- 设置请求头和cookies（固定部分 + 动态部分）；
- 初始化数据库连接（MySQL，`pymysql`），后续用于数据存储。

2.日志初始化

```
@staticmethod
def init_logger(current_file):
    logger.add(f"{current_file}.log", format='{time} {level} {message}',
               level="INFO", rotation='00:00')
    return logger
```

- 使用了 `loguru` 进行日志管理，简洁且易用；
- 日志文件每日凌晨自动轮转，便于日志管理。

3.关键请求准备函数

- `get_nonce_id(params)`：访问搜索首页，获取动态 `__ac_nonce` 值，头条反爬中重要的动态参数；
- `get_search_id(url, cookies, params)`：发请求获取动态 `search_id`，在请求参数里必填，用于分页请求；
- 这些函数确保请求参数符合今日头条的动态加密要求。

4.核心爬取函数： `keyword_spider`

```
def keyword_spider(self, keyword, page):
    # 构造请求参数
    params = {...}
    # 获取 nonce
    nonce_id = self.get_nonce_id(params)
    # 更新 cookies
    cookies = self.cookies
    cookies["__ac_nonce"] = nonce_id

    # 读取本地JS文件，用execjs调用js函数，生成加密签名
    with open('toutiao_keyword.js', 'r', encoding="utf-8") as f:
        jscode = f.read()
    signtrue = execjs.compile(jscode).call('get_sign', nonce_id)
    cookies['__ac_signature'] = signtrue

    # 固定更新 ttwid，防止封禁
    cookies['ttwid'] = '固定字符串'

    # 获取search_id并写入params
    search_id = self.get_search_id(url, cookies, params)
    params["search_id"] = search_id

    # 请求搜索结果
    response = requests.get(url, headers=self.headers, cookies=cookies,
                             params=params)

    # 从html中提取json数据，并加载
    json_data = re.findall('>({"extraData".*?})</script>', response.text)
    datas = json.loads(json_data[0]).get("rawData", {}).get("data", [])

    # 数据入库 + 进一步爬取评论、作者
```

- 主要步骤：
 - 先获取必须的动态参数 `nonce_id` 和 `search_id`，通过调用本地JS动态计算加密签名，保证请求合法；

- 访问带有cookies和headers的搜索接口，拿到搜索结果JSON；
- 解析JSON后遍历帖子数据，写入数据库；
- 对每条帖子调用评论爬取和作者爬取，深度爬取。

5. 评论爬取: `comment_spider`

```
def comment_spider(self, url):
    # 组装评论接口参数和请求
    u_id = url.split('/')[4]
    params = {..., "group_id": u_id, "item_id": u_id}
    response = requests.get(url, params=params, headers=self.headers)

    # 遍历评论，写入comments表
```

- 传入文章url，提取文章id后访问评论接口；
- 不需要复杂加密，直接请求，节省计算资源；
- 逐条写入评论表。

6. 作者信息爬取: `spider_spider_user`

```
def spider_spider_user(self, url):
    # 请求文章页，解析出作者json数据
    resp = requests.get(url, headers=self.headers, cookies={"ttwid": 随机字符串})
    html = etree.HTML(resp.text)
    content = ''.join(html.xpath('//div[@class="article-content"]/div[2]//text()')).strip()
    raw = ''.join(re.findall(r'ld\+json">(.*?)</script>', resp.text))
    jsdata = json.loads(raw)
    author = jsdata["author"]

    # 访问作者粉丝数和关注数接口
    stats = self.get_user_stats(author_url)

    # 入库 users 表
```

- 通过文章页内嵌JSON获取作者信息（包括作者主页url）；
- 访问作者粉丝关注接口获取更详细数据；
- 随机生成cookie防止封禁；
- 最后写入数据库。

7. 用户粉丝和关注数接口调用

```
def get_user_stats(self, author_url):
    token = author_url.rstrip('/').split('/')[-1]
    api = "https://www.toutiao.com/api/pc/user/fans_stat"
    data = {"token": token}
    resp = requests.post(api, headers=headers, data=data)
    ...
    return {"fans": ..., "follows": ...}
```

- 通过用户token访问官方粉丝统计接口；
- 返回粉丝和关注数；

- 方便统计分析作者影响力。

8. 运行入口与翻页

```
def run(self):
    for page in range(100000):
        result = self.keyword_spider("操作系统", page)
        if not result:
            break
```

- 以关键词“操作系统”为例，翻页采集；
- 遇到没有数据时结束循环；
- 支持超大页数，方便深度采集。

toutiao_keywords.js

反爬虫脚本，具体实现：

1. 代码上下文初始化部分

```
window = global;

document = {};
document.referrer = '';
document.cookie = '';
navigator = {};
navigator.userAgent = 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36'

location = {};
location.href = ''
location.protocol = 'https:';
```

- 这里定义了模拟浏览器环境的全局变量 `window`、`document`、`navigator`、`location` 等，给了部分属性值，给了一个伪装的 User-Agent 字符串。
- 一般用于在非浏览器环境（比如Node.js）运行这段代码，或者用来欺骗检测环境。

2. 核心函数定义：`_$jsvmprt`

```
(glb = "undefined" == typeof window ? global : window)._$jsvmprt = function(b, e, f) { ... }
```

- 这是一个赋给全局变量的函数。
- 函数参数 `b`, `e`, `f` 是外部传入的数据，`b` 是一个16进制字符数组或者类似的二进制数据。
- 这个函数里包含大量的内部函数和数据解码/解析逻辑。

3. 反射与继承检测代码

```
function a() {
    if ("undefined" == typeof Reflect || !Reflect.construct)
```

```

        return !1;
    if (Reflect.construct.sham)
        return !1;
    if ("function" == typeof Proxy)
        return !0;
    try {
        return Date.prototype.toString.call(Reflect.construct(Date, [],
(function () {}))),
            !0
    } catch (b) {
        return !1
    }
}

```

- 这个函数用于检测当前运行环境是否支持 `Reflect.construct`，以及 `Proxy` 等高级JS特性，判断JS引擎的能力。

4. 多层数据解析逻辑

```

function l(b, e) {
    var f = b[e++];
    var a = b[e];
    var d = parseInt("'" + f + a, 16);
    ...
}

```

- 这些函数实现了对16进制编码数据的解析，判断长度、偏移、内容等，作用是把二进制或16进制数据转成JS可操作的结构。

5. 动态执行代码的特征

这段代码中最关键的一点是：

- 有大量栈操作 `S[R]`，递归调用 `G` 函数，似乎是一个自定义的字节码解释器或者虚拟机。
- 通过对传入的二进制数据进行解码和执行，从而实现动态代码加载或执行。
- 结构类似某种基于栈的虚拟机，执行混淆过的字节码。

6. 加密或混淆关键

```

var r = 0;
...
for (P = i.q[z][0]; P < i.q[z][1]; P++)
    A += String.fromCharCode(r ^ i.p[P]);

```

- 这段代码中，`r` 是一个异或掩码，对数据进行了异或操作，常用作简单的加密/混淆。
- 这种手法可以防止直接用文本方式分析代码。

7. 错误检测和防篡改

```

if ("HNOJ@?RC" != I)
    throw new Error("error magic number " + I);

```

- 对输入的 b 数组头部有“magic number”校验，确保传入数据格式正确。
- 这也是一种防止代码被恶意修改或错误调用的机制。

特点	说明
伪造全局浏览器环境	让代码在非浏览器环境能执行，或者欺骗环境判断
复杂16进制字节码解析	读取16进制数据，解析成指令和参数
自定义字节码解释器	实现类似虚拟机的运行机制，动态执行代码
异或加密混淆	对数据做简单异或处理，防止明文分析
Magic Number 校验	确保数据格式正确，防止篡改
反射与特性检测	判断运行环境能力，做不同处理
动态执行代码	实现代码的动态加载和执行，防止静态分析

pie.py、post_viz.py、times.py、viz.py

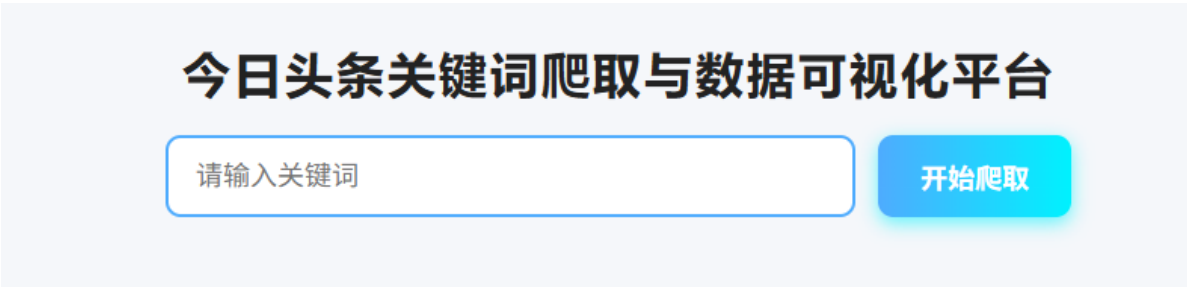
根据已经爬取并写入到本地数据库中的数据进行可视化分析

crawler.py

为主程序设计的爬虫代码，基本与jrtt.py一致。实现了用户界面自定义关键词输入功能。

app.py

主程序，采用了B/S架构，可运行特定关键字的今日头条信息爬取，在终端输入python app.py后，打开本地测试ip后，结果如下：



本代码集成了今日头条的内容爬取、对保存到本地数据库的爬虫数据的可视化分析，具体包括近30日发布的相关主题帖子的数目分析、爬取的帖子发布者的粉丝量排行信息、爬取的帖子的内容的摘要词云图、帖子评论内容的词云图。

3.具体运行

无用户可视化界面运行：

在终端输入 `python jrtt.py`

通过修改jrtt代码中的keyword的字段进行特定主题的内容爬取，具体效果如下：

```
result = self.keyword_spider("keyword",page)
```

```

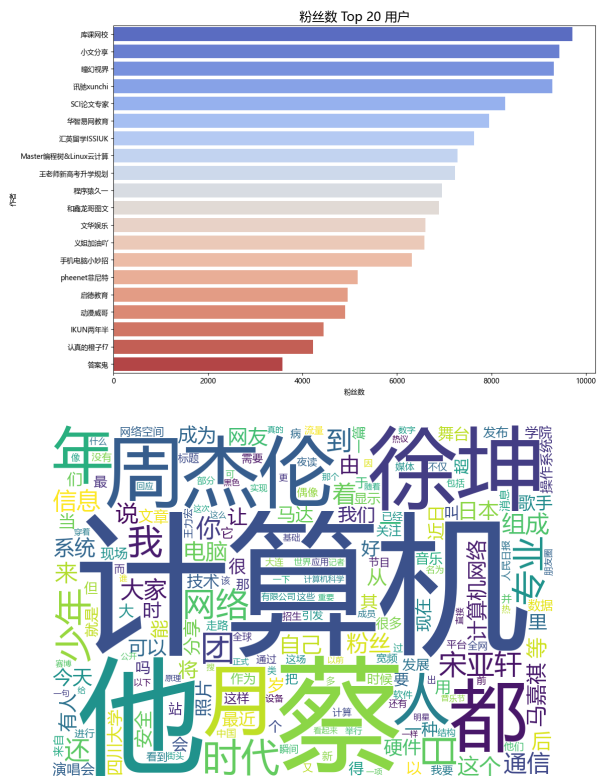
PS D:\6220> & C:/Users/sjj/Anaconda/python.exe d:/6220/jrtt.py
2025-06-29 16:54:26.948 | INFO | __main__:keyword_spider:70 - 当前采集 计算机组成, 第0
页
20250629165428D247E1D43B8E394246CA
2025-06-29 16:54:30.736 | INFO | __main__:spider_spider_user:284 - 作者 IT档案馆 信息写
入完 成
2025-06-29 16:54:30.738 | INFO | __main__:keyword_spider:146 - 计算机是由哪些部件组成的
? 如 何组装一台自己的电脑? ,https://toutiao.com/group/7097534679590404623/?channel=&in ogs=&
in tfs=&original source=&source=news&traffic source=&utm medium=wap search&utm source=, IT档
案馆
2025-06-29 16:54:31.193 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.198 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.201 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.205 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.212 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.216 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.221 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.224 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.228 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.232 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.236 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.241 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.245 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.249 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.253 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.260 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.264 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.269 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.273 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:31.276 | INFO | __main__:comment_spider:191 - 评论 信息写入完成
2025-06-29 16:54:32.616 | INFO | __main__:spider_spider_user:284 - 作者 CSDN 信息写入完

```

可以看到数据库中保存的爬取数据

561 rowselect * from users;						
id	author_name	author_url	fans_count	follow_count	content	
243	灼见	https://www.toutiao.com/c/user/token/MS4wLjABAAAA1P83cdm95b1XE1Rb1CV7z7GLOJF0gVur1C48w7SA/	26	0	NULL	
244	雅思天后刘薇	https://www.toutiao.com/c/user/token/MS4wLjABAAAA729gcltETQlqJleUfpyR2Ux96mCQp5S3XedbeVIE/	85	0	NULL	
245	新京报	https://www.toutiao.com/c/user/token/MS4wLjABAAAAHj3N8o7smE9De61Y8tzeXORMzCvX8a7NixQhSFg/	1182	0	NULL	
246	中肯留学	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA8reCC-OxkhBT87SgJFFlyU-Q1ws8gouDr19xql_9kSM/	408	0	NULL	
247	全球网安经济商	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAgqmELATQJk1Bv8Shd5d377mCqJ3cdhCtTmUd/	631	0	NULL	
248	美商网商校	https://www.toutiao.com/c/user/token/MS4wLjABAAAA52at3EFMFz4h1Qdp3Y0S8DUMF-1_gp2YekBpQGDz0/	4	0	NULL	
249	老李说留学	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAHkTePAG1JggitCytpDz_SF8NwV4ps4jbl_9u4vB8xhM1GayOqgWjKazTgdn1P1Gr/	1148	0	NULL	
250	YoYoStudy学术辅导	https://www.toutiao.com/c/user/token/MS4wLjABAAAA64tDUN53r53zq1_AeNbTM1d5d7FvNCLOJj_LFyJ-89SL58zxCnIS9IXLumhoV0r/	15	0	NULL	
251	新华网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAP9Lr1x61xP1mrgG8D9q-5om9Lans_kuTz_ipAGRE/	2896	0	NULL	
252	新华网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAP9Lr1x61xP1mrgG8D9q-5om9Lans_kuTz_ipAGRE/	2896	0	NULL	
253	汇英留学 TSTUUK	https://www.toutiao.com/c/user/token/MS4wLjABAAAAc5B0b81Eh_AlyxyqbdyGhVtMvMv13y3vUdUcTgTtVt/	7627	0	NULL	
254	秒拍新闻	https://www.toutiao.com/c/user/token/MS4wLjABAAAAHFRvzdqg37Z3B_huYq4h5F615f0J5wMkLw9D8m/	341	0	NULL	
255	美理留学	https://www.toutiao.com/c/user/token/MS4wLjABAAAAa0un7Gy1muel81B3d6ky0k3Cn8nTq9p5SPF02QhX54/	1245	0	NULL	
256	人民日报	https://www.toutiao.com/c/user/token/MS4wLjABAAAAa1a7J2T8oPvg1BMD3k105u7ATP0NthYLT4j2wZ1k1u/	2428	0	NULL	
257	中国日报网	https://www.toutiao.com/c/user/token/MS4wLjABAAAA_3z2pd1bwTt8ZLUPLSLshugPcU3fhZ3BTPIrSnphqC/	482	0	NULL	
258	环球网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAa2zhMceCo3Me9T13b1L231ACBGk3dcrd_1ZFwZ140/	3238	0	NULL	
259	选读新报	https://www.toutiao.com/c/user/token/MS4wLjABAAAAADc2ZIMEFT81Te-H9YGT38uQWqYQpVhTAl_vmx1X0gY/	1149	0	NULL	
260	人人都是产品经理	https://www.toutiao.com/c/user/token/MS4wLjABAAAAEXFV97Zt5C4_PRTTRWmXkyS105JugzwBuZQeUQ/	183	0	NULL	
261	人人都是产品经理	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA519AcenFWCgeNnaQ1N9pST_yoMaSNEVLS1hTF4b23I31880C1UC11kc5Mfbv9/	188	0	NULL	
262	人人都是产品经理	https://www.toutiao.com/c/user/token/MS4wLjABAAAAEXFV97Zt5C4_PRTTRWmXkyS105JugzwBuZQeUQ/	183	0	NULL	
263	中国网	https://www.toutiao.com/c/user/token/MS4wLjABAAAA8qdhBw07b7A7f1sd1zjsw9TKaqMULxopq1M_GZzhRSA/	841	0	NULL	
264	中国网	https://www.toutiao.com/c/user/token/MS4wLjABAAAA8qdhBw07b7A7f1sd1zjsw9TKaqMULxopq1M_GZzhRSA/	841	0	NULL	
265	中关村在线	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	149	0	NULL	
266	曹科技	https://www.toutiao.com/c/user/token/MS4wLjABAAAA1UPCwHtRfCP8b8kx10uV7UQ12m8BpAukuyU88/	69	0	NULL	
267	电脑爱好者	https://www.toutiao.com/c/user/token/MS4wLjABAAAA1jwM6oG7q7GOG7M3w6ee13h9FV_MtzbBV1KvVq5Cq/	33	0	NULL	
268	瑞智评测	https://www.toutiao.com/c/user/token/MS4wLjABAAAAVjp3b8cFr2hF336HsAUB97h7q-ouH8GogvYj2wZCq/	35	0	NULL	
269	曹科技	https://www.toutiao.com/c/user/token/MS4wLjABAAAA1UPCwHtRfCP8b8kx10uV7UQ12m8BpAukuyU88/	89	0	NULL	
270	曹科技	https://www.toutiao.com/c/user/token/MS4wLjABAAAA52zhMqMs_1V5HvDkH_sYF1FSN19vHnEzA_60RdaMa/	94	0	NULL	
271	钛金局	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAT-MtfdiaUPML-LAM-2v31y0MXJF9ggZk5N1E1_Bnc/	653	0	NULL	
272	前海海客梅	https://www.toutiao.com/c/user/token/MS4wLjABAAAAW52c2wZp2k1_AJus8_av0b0c-c20vwm2Atc2tNP1I/	7	0	NULL	
273	三星生活	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA2w_2AF1w30_ccnSt8l8p-aSF6K3mFC9P8p6-w4M/	5	0	NULL	
274	速途网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAD9t5VhW-DHn17X51akFvud1JFVdpNP3et33cT0K/	2	0	NULL	
275	安企中国	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAGQF5ScnOrFRqQK0K6p1i0uH-13H03v9b9foolp8eNc/	34	0	NULL	
276	腾讯数码精英	https://www.toutiao.com/c/user/token/MS4wLjABAAAAV115x-X19b8a1kva1XhpD7QbA-nig5UPBPATiEg10Ma4y-jgitwdaAKe/	3	0	NULL	
277	彼得斯技术公司	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAFGvghL9FOASc1Txc1K1xplw1_PcltpoQ9d_NEOXPC/	182	0	NULL	
278	全球迈	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAQ3CkRoFUA8p7T0nV2oz-dtITqUyUcZRTSL58z3S5omYua6_YjyAgjvYUlg6/	3	0	NULL	
279	机智通	https://www.toutiao.com/c/user/token/MS4wLjABAAAA3n-1nsR3EDwXv1U8B8b8d8n8-q0Jw4B3JgTAM/	5	0	NULL	
280	中关村在线	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	149	0	NULL	
281	趣康科技	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAQCLFumC2G3mQcCnA83s_2AYW1lFyH8qG9W4rpass9M21fhK3pjabixCT13Vj/	8	0	NULL	
282	中关村在线	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	149	0	NULL	
283	青石古道	https://www.toutiao.com/c/user/token/MS4wLjABAAAAATsidmNhgUNhFN3-xvPC4Hr5XFJd8JyGCK169FUMFzC53ZURACkdvCJvVqdwg53z/	1	0	NULL	
284	当代通商	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	28	0	NULL	
285	贝塔斯利曼	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAQ3CkRoFUA8p7T0nV2oz-dtITqUyUcZRTSL58z3S5omYua6_YjyAgjvYUlg6/	2	0	NULL	
286	美商网商校	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	2493	0	NULL	
287	长洲数人	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAGF5ScnOrFRqQK0K6p1i0uH-13H03v9b9foolp8eNc/	1	0	NULL	
288	做个靠谱的坏人	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAGF5ScnOrFRqQK0K6p1i0uH-13H03v9b9foolp8eNc/	2	0	NULL	
289	站长的爹	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAGF5ScnOrFRqQK0K6p1i0uH-13H03v9b9foolp8eNc/	6	0	NULL	
290	中国日报网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	482	0	NULL	
291	英语天才talk	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	15	0	NULL	
292	必克英语	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	18	0	NULL	
293	爱范儿	https://www.toutiao.com/c/user/token/MS4wLjABAAAAAQ3CkRoFUA8p7T0nV2oz-dtITqUyUcZRTSL58z3S5omYua6_YjyAgjvYUlg6/	58	0	NULL	
294	96后探泊大叔	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	121	0	NULL	
295	中国日报网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	482	0	NULL	
296	李书航团队	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	29	0	NULL	
297	美商网商校	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	57	0	NULL	
298	美商网商校	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	2493	0	NULL	
299	徐体APP	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	84	0	NULL	
300	徐体APP	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	687	0	NULL	
301	中国日报网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	58	0	NULL	
302	中国网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	841	0	NULL	
303	什么值得买	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	106	0	NULL	
304	陈一梅的网络日志	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	3	0	NULL	
305	江西网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	6	0	NULL	
306	PCOnline太平洋科技	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	97	0	NULL	
307	启得教育	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	4959	0	NULL	
308	索思Daniel	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	781	0	NULL	
309	微媒体APP	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	84	0	NULL	
310	fordic学习力	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	4	0	NULL	
311	中国日报网	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	482	0	NULL	
312	芥末留学指南	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	2559	0	NULL	
313	徐体APP	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	687	0	NULL	
314	芥末留学指南	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	2559	0	NULL	
315	芥末留学指南	https://www.toutiao.com/c/user/token/MS4wLjABAAAAA3qprtr1pJg1y5k8-eJSSPLDvFfYyR7K765KAB_Tk5b8/	3	0	NULL	

对爬取数据的可视化，即可直接分别运行pie.py、post_viz.py、times.py、viz.py代码即可，得到如下可视化信息：



有用户可视化界面运行：

在终端输入 `python app.py`

点击本地测试ip后，输入想爬取的相关主题内容即可，具体效果如下：

