

# 进阶实验

## ## 目标

- \* 爬取所有的帖子信息，存入 sqlite 数据库中（参照“实验指导.pdf”）
- \* 爬取所有的评论信息，存入 sqlite 数据库中（crawler 代码只完成了爬取帖子的操作，可仿照完成爬评论的操作）

## \* 注意：

\*\* 数据库名叫 wall，帖子的表名叫 post，评论的表名叫 comment，列名用 JSON 格式“名-值”对中的名即可

\*\* 要求可以增量爬取，第一次爬取的时候时间会比较久，以后爬取的时候只需要爬取没有存入到数据库中的内容即可。

## ## 获得帖子列表

### ### 请求网址

[http://weixiao.nickboy.cc/wall/content/gh\\_3852826aa4ab?&page=1&q=&cid=0&id=0&orderby=2](http://weixiao.nickboy.cc/wall/content/gh_3852826aa4ab?&page=1&q=&cid=0&id=0&orderby=2)

## \* 注意：

\*\* page 从 1 遍历到 4108（过几天可能会到 4109，4110.....）

\*\* 其余参数均可保持不变

### ### 相应数据（JSON 格式）

\* id: 帖子的 id

\* openid: 帖子发表人的 id

\* content: 帖子内容（unicode 编码）

\* picurl: 帖子内容中的图片 URL（可能由多个 URL 组成的列表）

\* nickname: 帖子发表人自己填写的昵称

- \* r\_count: 帖子的评论数
- \* time: 帖子发表的时间戳
- \* weight: 帖子是否置顶
- \* good: 帖子的点赞数
- \* premium: 帖子是否精华

```
[{"id":6278811,"openid":"oZpm3016UVonZw7jx-eJeAHwEXtc","cid":95497,"content":"\u6709\u5973\u751f\u60f3\u7684\u561b\u5f0c\u7537","picurl":"","nickname":"","r_count":0,"time":1644483220,"weight":0,"good":0,"premium":0}, {"id":6278774,"openid":"oZpm309pK74eoW56-hiFb9dEoZxY","cid":95497,"content":"\u4e3a\u4ec0\u4e48\u5c0f\u5587\u53ed\u6700\u8fd1\u7684\u4eba\u8fd9\u4e48\u591a\u5f0c\u90fd\u5feb\u53d8yaozi\u4e6","picurl":"","nickname":"","r_count":1,"time":1644482776,"weight":0,"good":0,"premium":0}, {"id":6278741,"openid":"oZpm30wa7GtRZvHMwLYPTWxJw0jl","cid":95497,"content":"\u8fd4\u6821\u5f0c\u662f14\u5929\u65e0\u65b0\u589e\u5c31\u53ef\u4ee5\u56de\u8fd8\u662f\u7b49\u5230\u53d8\u4f4e\u98ce\u9669\u4e6e\u4ee5\u540e\u518d\u7b4914\u5929\u624d\u53ef\u4ee5\u56de\u5f1f\u6211\u73b0\u5728\u5728\u4e2d\u98ce\u9669\u5f0c\u5df2\u7ecf\u6709\u4e24\u5929\u6ca1\u6709\u65b0\u589e\u4e6e","picurl":"","nickname":"\u533f\u540d","r_count":1,"time":1644482279,"weight":0,"good":0,"premium":0}, {"id":6278692,"openid":"oZpm30xXThgdYQADQDpsNDkdmR_c","cid":95497,"content":"\u60c5\u4eba\u8282\u7537\u53cb\u4eec\u7ed9\u6211\u9001\u4ec0\u4e48\u793c\u7269\u7269\u597d\u5462","picurl":"","nickname":"","r_count":0,"time":1644481452,"weight":0,"good":0,"premium":0}, {"id":6278688,"openid":"oZpm30xwHXBuZ5ojLMWIKrARyeTU","cid":95497,"content":"\u544a\u8beb\u5404\u4f4d\u59d0\u59b9\u5343\u4e07\u4e0d\u8981\u5728\u6708\u7ecf\u671f\u95f4\u548c\u590a\u4e2a\u2026\u4eca\u5929\u65e9\u4e0a\u4e0a\u5934\u4e6e\u6ca1\u63a7\u5236\u4f4f\u5f0c\u5934\u6655\u773c\u82b1\u5230\u73b0\u5728\u545c\u545c\u545c\u545c\u545c\u545c","picurl":"","nickname":"","r_count":7,"time":1644481403,"weight":0,"good":1,"premium":0}, {"id":6278682,"openid":"oZpm30-gC6ShpWzMG9vEBqPYTiW4","cid":95497,"content":"\u5927\u5bb6\u6709\u4ec0\u4e48\u671f\u5f85\u7684\u8868\u767d\u65b9\u6848\u561b\u83e\u5dd7\u60f3\u5411\u75"}]
```

37\u670b\u53cb\u8868\u767d\u0c\u6765\u4e2a\u53cc\u5411\u5954\u8d74\u01",  
,"picurl":"","nickname":"","r\_count":0,"time":1644481304,"weight":0,"good":2,"premium":0},{  
"id":6278662,"openid":"oZpm309VXwqJnnsTvzbUXgOUrzX4","cid":95497,"content":  
"\u8f6c\u4e13\u4e1a","picurl":"","nickname":"","r\_count":0,"time":1644480988,"weight":0,  
"good":0,"premium":0},{  
"id":6278646,"openid":"oZpm30xXThgdYQADQDpsNDkd mR\_c","cid":95497,"content":  
"\u534e\u5983\u4e34\u6b7b\u4e4b\u524d\u0c\u7504\u5b1b\u4e3a\u4ec0\u4e48\u8fd8\u8981\u53bb\u770b\u5979\u01f\u96be\u9053\u62f\u5fc3\u8f6f\u4e86\u01f","picurl":"","nickname":"","r\_count":2,"time":1644480827,"  
weight":0,"good":1,"premium":0},{  
"id":6278626,"openid":"oZpm30-7L2Uc8MwanZfQ0 QYToGUo","cid":95497,"content":  
"\u4eca\u5929\u76f4\u63a5\u83bd\u5230\u4e86\u857\u9053\u529e\u0c\u8054\u7cfb\u5230\u4e86\u4e3b\u4efb\u0c\u4f46\u4ed6\u4eec\u53ea\u8ba9\u6211\u7559\u4e86\u7535\u8bdd\u8bf4\u7ea6\u597d\u65f6\u95f4\u7ed9\u6211\u6253\u0c\u8fd9\u662f\u6ca1\u5e0c\u671b\u7684\u610f\u601d\u5417\u01f\u8981\u662f\u6ca1\u7ed9\u6211\u6253\u7535\u8bdd\u4e0b\u5468\u4e00\u8981\u4e0d\u8981\u518d\u5230\u8857\u9053\u529e\u01f\u08\u4e3b\u8981\u65f6\u95f4\u4e0d\u5927\u591f\u4e86\u2026\u5bc4\u01f","picurl":"","nickname":"","r\_count":0,"time":1644480644,"weight":0,"good":0,"premium":0},{  
"id":6278516,"openid":"oZpm307N97\_P\_lpQl7WhERPe4zpM","cid":95498,"content":  
"\u6211\u597d\u60f3\u540e\u5165","picurl":"","nickname":"","r\_count":12,"time":1644479247,"weight":0,"good":3,"premium":0},{  
"id":6278417,"openid":"oZpm30zVlhYV76lragGDIwMM\_Zsw","cid":95497,"content":  
"\u600e\u6837\u624d\u80fd\u4e0d\u8ba9\u53d1\u73b0\u81ea\u5df1\u51fa\u8f68\u2026\u9a6c\u4e0a\u5f00\u5b66\u4e86double\u4e5f\u4e48\u64cd\u4f5c\u554a\u83d\u2d","picurl":"","nickname":"","r\_count":5,"time":1644478197,"weight":0,"good":1,"premium":0},{  
"id":6278410,"openid":"oZpm306XWWR-Ygx PIP5pWtbXih74","cid":95497,"content":  
"\u01f\u0c\u81ea\u5bb6\u95e8\u53e3\u90fd\u4e0d\u6562\u4e3a\u81ea\u5df1\u63a7\u8bc9\u593a\u56de\u91d1\u724c","picurl":"","nickname":"","r\_count":0,"time":1644478132,"weight":0,"good":0,"premium":0},{  
"id":6278362,"openid":"oZpm30z20Chlkr-vR7QLWtyeoPd0","cid":95497,"content":  
"\u19\u53f7\u665a\u4e0a\u70b9\u843d\u5730\u4ece\u9996\u90fdT3\u56de\u5b66\u6821\u6

709\u62fc\u8f66\u7684\u5bb6\u4eba\u561b","picurl":"","nickname":"","r\_count":0,"time":1644477696,"weight":0,"good":0,"premium":0},{ "id":6278345,"openid":"oZpm303QOqttl4MMCBRq\_EdvFIEM","cid":95497,"content":"\u4eba\u5927\u7684\u8d22\u91d1\u548c\u5317\u5927\u5149\u534e\u5dee\u522b\u80fd\u6709\u591a\u5927\u554a\u540d","r\_count":7,"time":1644477495,"weight":0,"good":7,"premium":0},{ "id":6278327,"openid":"oZpm303wAXEssHyg8qihAJopeF6Y","cid":95497,"content":"\u4e0d\u7406\u89e3\u6709\u4e9b\u7537\u7684\u79c1\u5e95\u4e0b\u8ddf\u4f60\u66a7\u6627\u804a\u5929\u7ea6\u4f60\u7136\u540e\u5bf9\u5916\u8bf4\u81ea\u5df1\u6ca1\u6709\u559c\u6b22\u7684\u4eba\u83d\u05\u771f\u7684\u5f88\u8d25\u574f\u597d\u611f\u76f4\u63a5\u4e0b\u5934","picurl":"","nickname":"\u533f\u540d","r\_count":6,"time":1644477338,"weight":0,"good":3,"premium":0}]

### 截图

# 人大墙 (匿名社区)

全部

表白

吐槽

秘密

闲置

Q 搜索

说点什么呢~



表白

昵称

发布

我发

我回

我赞

新回

新发

最热

精选

#61617 · 匿名 | 表白

4分钟前

有女生想y的嘛, lz男

👍0

💬0

#61616 · 匿名 | 表白

12分钟前

为什么小喇叭最近y的人这么多, 都快变yaozi了

👍0

💬1

#61615 · 匿名 | 表白

20分钟前

返校, 是14天无新增就可以回还是等到变低风险了以后再等14天才可以回? 我现在在中风险, 已经有两天没有新增了

👍0

💬1

#61614 · 匿名 | 表白

34分钟前

情人节男友们给我送什么礼物物好呢

👍0

💬0

#61613 · 匿名 | 表白

34分钟前

告诫各位姐妹千万不要在月经期间和npy那个...今天早上上头了没控制住, 头晕眼花到现在呜呜呜呜呜呜

👍1

💬1



## ## 获得帖子评论列表

[https://weixiao.nickboy.cc/wall/comments/gh\\_3852826aa4ab?cid=6278646&page=1](https://weixiao.nickboy.cc/wall/comments/gh_3852826aa4ab?cid=6278646&page=1)

### \* 注意:

\*\* page 从 1 遍历往后遍历, 直至没有数据

\*\* cid 填写想要查看评论的帖子标号, 上面 URL 可以查看到帖子列表 JSON 数据中黄色高亮那条帖子对应的评论

\*\* 其余参数均可保持不变

### 相应数据 (JSON 格式)

\* id: 评论的 id

\* openid: 评论发表人的 id

\* content: 评论内容 (unicode 编码)

\* picurl: 评论内容中的图片 URL (可能由多个 URL 组成的列表)

\* nickname: 评论发表人自己填写的昵称

\* time: 评论发表的时间戳

```
[{"id":7046963,"openid":"oto4e0_JdlRI-LCVQoAjVa6YMkN8","content":"\ud83d\udc7f",
"picurl":"","nickname":"","time":1588649262},{id":7046961,"openid":"oto4e0-u2vghPO
Ro3pHI5HwRmUDQ","content":"\u603b\u6709\u4eba \u5728\u6df1\u591c\u91cc
\u54ed\u6ce3~","picurl":"","nickname":"","time":1588649212},{id":7046959,"openid":"
oto4e08Kzh3GTEjRqOdp_3DnYKB8","content":"\u6709\u4e9b\u975e\u4e3b\u6d41\u5
4e6\uff5e","picurl":"","nickname":"","time":1588644818},{id":7046939,"openid":"oto4e
0-Mp1_oBV2QhUgovXAlO5Q","content":"\u6709\u554a","picurl":"","nickname":"","ti
me":1588637454},{id":7046924,"openid":"oto4e02t9cMMlITvKddHYjnJOXUg","conten
t":"\u53ef\u76d0\u53ef\u751c\u7684\u4e5f\u597d\u554a","picurl":"","nickname":"","ti
me":1588606352},{id":7046923,"openid":"oto4e0zR-Y5fUWOiGoVSq1QR0RNM","cont
ent":"\u94c1\u6c41 \u975e\u5e26\u662f\u7f8e\u7537\u5417
\u53ef\u7231\u6b63\u592a\u4e0d\u884c\u5417","picurl":"","nickname":"","time":158
8606112},{id":7046854,"openid":"oto4e0y8OQaLbKxG8tlCKbT2fQOQ","content":"\u59
73\u670b\u53cb\u4e0d\u597d\u73a9\u83c\udf1d \ud83c\udf1d \ud83c\udf1d
\u5355\u8eab\u597d\u4e45\u597d\u4e45\u4e86","picurl":"","nickname":"","time":158
```

8565954},{\"id\":7046852,\"openid\":\"oto4e0zR-Y5fUWOiGoVSq1QR0RNM\",\"content\":\"\\u6765\\u627e\\u6211 \\u6211\\u53ef\\u5bb9\\u6613\\u5f97\\u5230\\u4e86\\u54c8\\u54c8\\u54c8\",\"picurl\":\"\",\"nickname\":\"\",\"time\":1588561902},{\"id\":7046808,\"openid\":\"oto4e07r-OpXKx2DZTXVHKDyicoA\",\"content\":\"\\u672a\\u6765\\u4f1a\\u6709\\u66f4\\u597d\\u7684\",\"picurl\":\"\",\"nickname\":\"\\u8d77\\u98ce\\u4e86\",\"time\":1588514630},{\"id\":7046400,\"openid\":\"oto4e00CFxOGj8ecLRY5Xd3\_I0OQ\",\"content\":\"\\u60f3\\u60f3\\u5427\\uff0c\\u563b\\u563b\",\"picurl\":\"\",\"nickname\":\"\",\"time\":1590655272}]

### 截图

#61610 · 匿名 | 表白

1小时前 :

华妃临死之前，甄嬛为什么还要去看她？难道是心软了？

👍1

💬2

回点什么呢~



添加图片

昵称

评论

点击下面回复的内容可@评论者

#2 · 匿名 (1号)

54分钟前 :

去气死华妃

#1 · 匿名 (2号)

1小时前 :

不是不是，因为甄嬛知道华妃不死心一心要见皇上，皇上见了一心软就不会死了，华妃当时不死心啊还沉迷在爱情里，她去就是要告诉华妃欢宜香的真相是诛心去了，让华妃心死心伤自尽

▽ 更多评论