**选股宝、财联社7X24h新闻爬取说明**

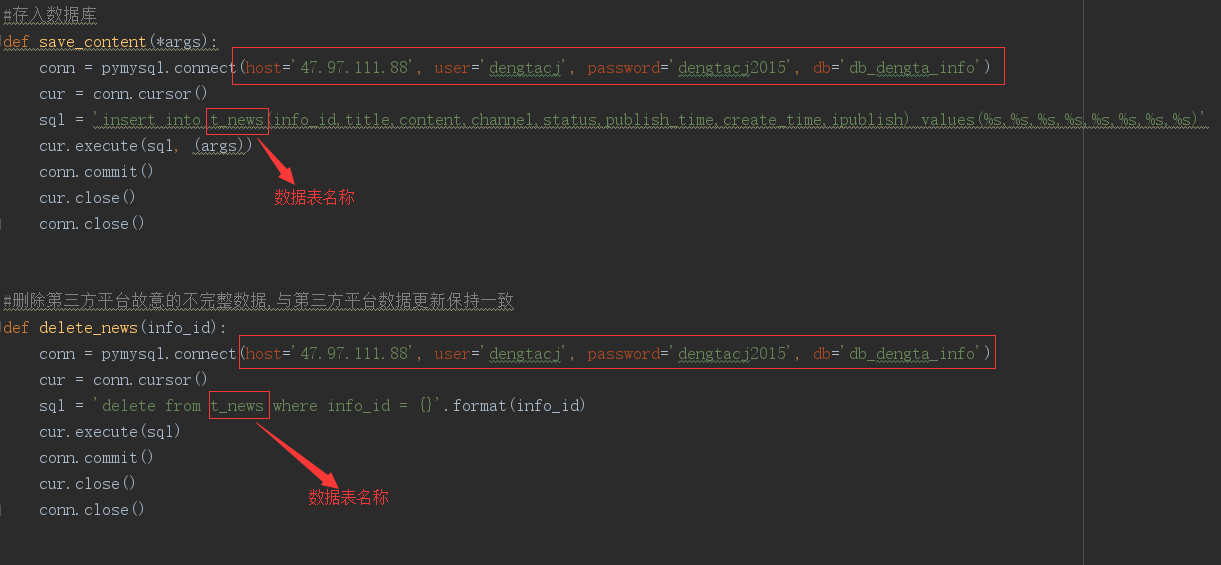
**Gitlab地址:** http://47.97.161.43/crawler/get\_7X24\_news.git

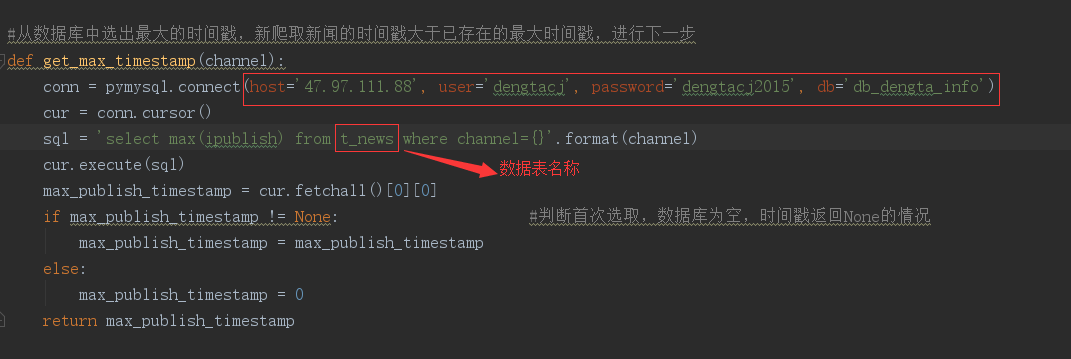
**虚拟环境安装依赖包：**requirements.txt

**运行7X24新闻爬虫：**nohup python3 -u run.py params1 > 7X24\_news.out 2>&1 &

**数据库配置更改：**

1. save\_to\_sql.py文件中save\_content、delete\_news、get\_max\_timestamp三个方法均需修改，改为正式使用的数据库和数据表：





1. rule.py文件中get\_info方法需要修改，改为正式使用的数据库和数据表：



**7X24\_news脚本说明：**

1. run.py为主程序脚本，运行该脚本即可启动爬虫程序；
2. get\_xgb.py为采集选股宝新闻的爬虫脚本，header\_xgb.py是采集选股宝新闻的头文件信息脚本；
3. get\_cls.py为采集财联社新闻的爬虫脚本，header\_cls.py是采集财联社新闻的头文件信息脚本；
4. rule.py为数据清洗规则脚本；
5. save\_to\_sql.py为数据保存至数据库的脚本；
6. logger\_error.py为记录程序异常的脚本；
7. newslog3.txt为记录程序异常信息的txt文档；
8. requirements.txt用于记录程序运行所需的依赖包及其精确的版本号，以便新环境部署；
9. 7X24news\_read\_me.txt是脚本说明文档。