主题：bilibili视频简要数据管理

内容：通过爬虫对bilibili视频的简要数据进行收集，其中包括视频av号，视频标题，视频up主，投稿时间，视频播放数，视频收藏数，视频弹幕数，视频硬币数量，视频tag。

开发模式：BSS

开发环境：lnmp，linux + nginx + mysql + python

开发配置：

|  |  |
| --- | --- |
| linux | Linux deathsea-PC 3.13.0-24-generic #46-Ubuntu SMP Thu Apr 10 19:11:08 UTC 2014 x86\_64 x86\_64 x86\_64 GNU/Linux |
| nginx | nginx version: nginx/1.4.6 (Ubuntu) |
| mysql | MySQL Ver 14.14 Distrib 5.5.41, for debian-linux-gnu (x86\_64) using readline 6.3 |
| python | Python 2.7.6 |

开发工具：vim or PyCharm

框架选用：web.py ，其中数据库操作部分不使用框架本身功能进行，而是使用mysql自家的mysql-connector-python进行。

部署方式：fastcgi

具体配置：待填

数据库具体设计：

数据库名：DBFP\_av\_info，保存视频信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 中文名 | 类型 | 允许为空 | 备注 |
| id | ID | MEDIUMINT | 否 | Primary id  auto increatment |
| av | av号 | int(8) | 否 |  |
| title | 视频标题 | char(100) | 否 |  |
| up\_id | up主id | Mediumint | 是 | DBFP\_up\_info外键 |
| con\_time | 投稿时间 | DATETIME | 否 |  |
| play\_times | 视频播放数 | int(8) | 否 |  |
| collect\_times | 视频收藏数 | int(8) | 否 |  |
| dan\_count | 视频弹幕数 | int(8) | 否 |  |
| Bi\_count | 视频硬币数量， | int(8) | 否 |  |
| tags | 视频tag | SET | 是 | 存放tag的id |

数据库名：DBFP\_up\_info，保存up主信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 中文名 | 类型 | 允许为空 | 备注 |
| id | ID | MEDIUMINT | 否 | Primary id  auto increatment |
| uid | 用户id | INT | 否 | b站唯一id |
| name | Up主名字 | char(10) | 否 |  |
| lvl | up等级 | SMALLINT | 否 |  |
| sign | up主签名 | char(100) | 是 |  |
| birth | up主生日 | DATE | 否 |  |
| reg\_date | up主注册时间 | DATE | 否 |  |
| follow\_count | 关注数 | INT | 否 |  |
| fans\_count | 粉丝数 | INT | 否 |  |

数据库名：DBFP\_tag，保存所有已有的tag

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 字段名 | 中文名 | 类型 | 允许为空 | 备注 |
| id | ID | MEDIUMINT | 否 | Primary id  auto increatment |
| name | tag名 | char(30) |  |  |

难点：爬虫，不排除bilibili有反爬虫的策略。