LOTTERY爬虫需求

2019年3月21日

版本：1.0

版本记录：

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 修改日期 | 添加需求 | 添加人 |
| 1.0 | 2019-3-21 | 基础数据表需求；  兼容sqlite、mysql；  脚本执行规范； | nigelaji |
|  |  |  |  |
|  |  |  |  |

## 双色球爬取需求

表名：ssq\_basedata（可自定义表数量或者名称，可交流）

爬取字段：

期号 issue varchar(30)

开奖时间 open\_time varchar(30)

红球1 red1 int

红球2 red2 int

红球3 red3 int

红球4 red4 int

红球5 red5 int

红球6 red6 int

蓝球 blue int

红和值 sum\_red int

连号 AA char(1) (二个及以上红连号，枚举值：0代表无、1代表有)

奇 odd int (红奇个数)

偶 even int （红偶个数）

大 big int （大于等于17个数）

小 small int （小于17个数）

## 大乐透爬取需求

表名：dlt\_basedata

爬取字段：

期号 issue

开奖时间 open\_time

红球1 red1

红球2 red2

红球3 red3

红球4 red4

红球5 red5

蓝球1 blue1

蓝球2 blue2

红和值 sum\_red

奇 odd int (红奇个数)

偶 even int （红偶个数）

大 big int （大于等于17个数）

小 small int （小于17个数）

## 数据库支持

mysql数据库

## 脚本要求

尽量执行脚本时不加参数，直接 python crawling.py run

命令行要输出一些信息，比如，爬取了多少条、用时、等相关信息，相当于一个运行记录。

## 日志要求

log/error.log 代码报错日志

log/record.log 脚本运行记录日志

## 表数据备份要求

文本文件备份（利用pickle模块）

backup/ssq\_basedata.pkl

backup/dlt\_basedata.pkl

## 表设计参考二

合并双色球和大乐透

字段：

期号 issue

开奖日期 open\_time

红球列表 red\_list

篮球列表 blue\_list

红和值 sum\_red

连号 red\_AA char(1) (二个及以上红连号，枚举值：0代表无、1代表有)

奇 odd int (红奇个数)

偶 even int （红偶个数）

大 big int （大于等于17个数）

小 small int （小于17个数）