**一：关键词段落:（数据量需求，不限制，根据接口信息，设置个每天定时更新爬取）**

**关键词段落总体思想：主要是为了找配资行业的，原创文章，包含改文章关键词的段落。**

**关键词段落目的：这个是针对行业文章的一个重要内容，主要针对原创文章的关键词段落进行抓取。**

**关键词段落应用：形成文章的关键词段落**

**数据源：(站点 包含配资的栏目文章)**

[https://www.peizitoutiao.com/](https://www.peizitoutiao.com/gppz/)

https://www.magic-sunshading.com/

<https://www.blogcms.cn/>

<https://www.shmeishida.com/>

https://www.anxinsc.com/

信息要求：

1：先判断文章关键词是否包含配资，才执行下面的步骤

2：去掉文章内容第一段，和最后一段。从中间段落，爬取包含文章标签词的段落，要求段落字符数大于100,小于500.

3：标注对应段落的文章关键词(就是文章标签词)

4：标注对应段落的替换关键词为配资

5：开头是链接词的，替换连接词

Post地址:

http://121.40.187.51:8088/api/key\_paragraph\_api

key ： md5('datapool'+时间戳年月日)

content 关键词段落内容

keyword 关键词

rekeyword 替换关键词

**二：关联段落:（数据量需求，不限制，设置个每天定时更新爬取）**

**关联段落总体思想：主要是为了找跟股票相关的一些段落，主要是哪些股票原创文章或者股票知识性的文章的一些包含对应关键词的段落。**

**关键词段落目的：这个是为了让文章的相关性更高，需要跟关键词段落结合使用。**

**关键词段落应用：形成文章的相关性段落**

**数据源：**

http://www.southmoney.com/zhishi/gprm/

信息要求：

1：从数据源爬取段落，要三天内的。

2：段落文字数需要大于250个字符，小于500个字符。

3：段落需要包含以下有股票关键词的文字（有匹配到一个就行）：

个股,股市,A股,港股,新股,美股,创业板,证券股,股票,炒股,散户,短线,操盘,波段

4：爬取的段落，标注对应匹配股票关键词。

5：不能重复

6：每个数据源，写好自动化执行的脚本。定期自动更新，每隔1小时执行一次。

2021年8月27日新增筛选：

7：段落包含:“图”这个文字的。过滤

8：段落包含股票名字，或者股票代码的。过滤（这个要另外爬取个股票名字跟代码表，只需要A股，A股会不定期更新，所以也需要做个定期更新这个表。一般一个星期一次）

2021年9月6日新增筛选：

9：包含日期的段落过滤

10：指向性比较强的段落过滤

11：开头是链接词的，替换连接词

Post地址:

http://121.40.187.51:8088/api/relation\_paragraph\_api

key ： md5('datapool'+时间戳年月日)

content 关联段落内容

**三：缩略图:（数据量需求，不限制，设置个每天定时更新爬取）**

要求:

1:金融类通用图片

2:改md5

3:将图片进行比例处理,按照121x75(长宽)的比例进行裁剪

4:命名用接口信息命名

Post地址:

http://121.40.187.51:8088/api/thumbnail\_api

key ： md5('datapool'+时间戳年月日)

file 缩略图文件

**四: 内容图:（数据量需求，不限制，设置个每天定时更新爬取）**

要求:

1:金融类通用图片

2:改md5

3:命名用接口信息命名

Post地址

[http://121.40.187.51:8088/api/contentimgs\_api](http://121.40.187.51/api/contentimgs_api)

key ： md5('datapool'+时间戳年月日)

file 内容图文件

**五：传视频**

**视频的主体思想：主要是找哪些当天的，原创的，知识性讲解类型或者炒股技巧类型的视频，如果有配资，或者杠杆融资比较相关的视频也可以。**

**例如:**

为什么散户容易亏钱？就是忽略了止损的重要性！简单止损一学就会

被套不用怕，调仓换股的最佳操作方法，看懂战胜90%的股民

改良超强指标cci，仅用30秒选出强势龙头！股民学会轻松翻倍！

**关键词段落目的：主要生成股票通用型文章，供发文章使用。**

**关键词段落应用：伪原创网站需要用到的内容**

**数据源：**

抖音：搜股票相关内容，每隔1小时更新，筛选2分钟到5分钟的视频。

快手：搜股票相关内容，每隔1小时更新，筛选2分钟到5分钟的视频。

筛选:

1:不要有时间性的,和个股评论的（例如: 9.28、午评、早评

）要那些炒股技巧 知识类型的。

2：包含股票名称，股票代码的也不要。

3：包含指数的也不要

**六：传点评数据**

**点评数据主体思想：主要是找哪些股票相关的评论。**

**关键词段落目的：为了增加网站的一些相关性信息更新。**

**关键词段落应用：网站的点评更新。**

**数据源：**

**例如: <https://t.10jqka.com.cn/> (不要用这个数据源,已经用了)**

**<https://xueqiu.com/>**

**找股票用户评论的哪些网站的交流发言数据**

**筛选:**

1：要当天的数据

2：评论内容大于50字符，小于500字符

3：筛选包含这些词的.个股,股市,A股,新股,股票,炒股,散户,短线,操盘,波段,杠杆

4：不要资讯类型的

5：将个股,股市,A股,新股 这些词 替换为 股票