#### **指导**

以Bilibili热榜为例，练习协程与非关系数据库的使用

1. 获取某一分区视频的排行榜信息，并从中解析出前 10000 条视频信息。可使用协程爬取数据。注意协程应用的方式：发送http请求后，可以跳转到数据处理的函数中，而不要跳到发送另一个http请求的函数中，因为由于b站的风控较为严格，短时间大量的请求可能会导致ip被封禁。

2. 将视频信息存入MongoDB数据库，并记录此视频当时的排名和数据创建时间。

3. 爬取同一分区一周之后的热榜，并从中解析出前 10000 条视频信息。对比两次结果并更新数据库，要求第一次结果需从数据库中取出。根据以下规则更新数据库：对于仅在第二次排行榜中的视频，将信息存入数据库；对于仅在第一次排行榜中的视频，将信息从数据库中删除；对于两次排行榜都存在的视频，更新数据库，保留创建时间，增加更新时间字段，并更新排名。

4.（附加）可否从数据库中查询出两次排行榜中上升和下降排名前10的视频信息？

5.（附加）由于数据量比较大，数据库处理的时间可能较长，可以考虑使用SMTP协议在数据库处理完成时通过邮件的形式通知你。

相关链接：

- b站分区编码一览表：https://github.com/SocialSisterYi/bilibili-API-collect/blob/master/video/video\_zone.md （如果你没有用到这个网址，大概率本次作业爬取的数据不符合要求）

- 视频热榜API：https://s.search.bilibili.com/cate/search?main\_ver=v3&search\_type=video&view\_type=hot\_rank&copy\_right=-1&cate\_id={分区编码}&page={页码}&pagesize={分页大小}&time\_from={开始时间}&time\_to={结束时间}

example：https://s.search.bilibili.com/cate/search?main\_ver=v3&search\_type=video&view\_type=hot\_rank&copy\_right=-1&cate\_id=194&page=1&pagesize=20&time\_from=20211210&time\_to=20211217

窗体顶端

窗体底端