# 数据库综合作业

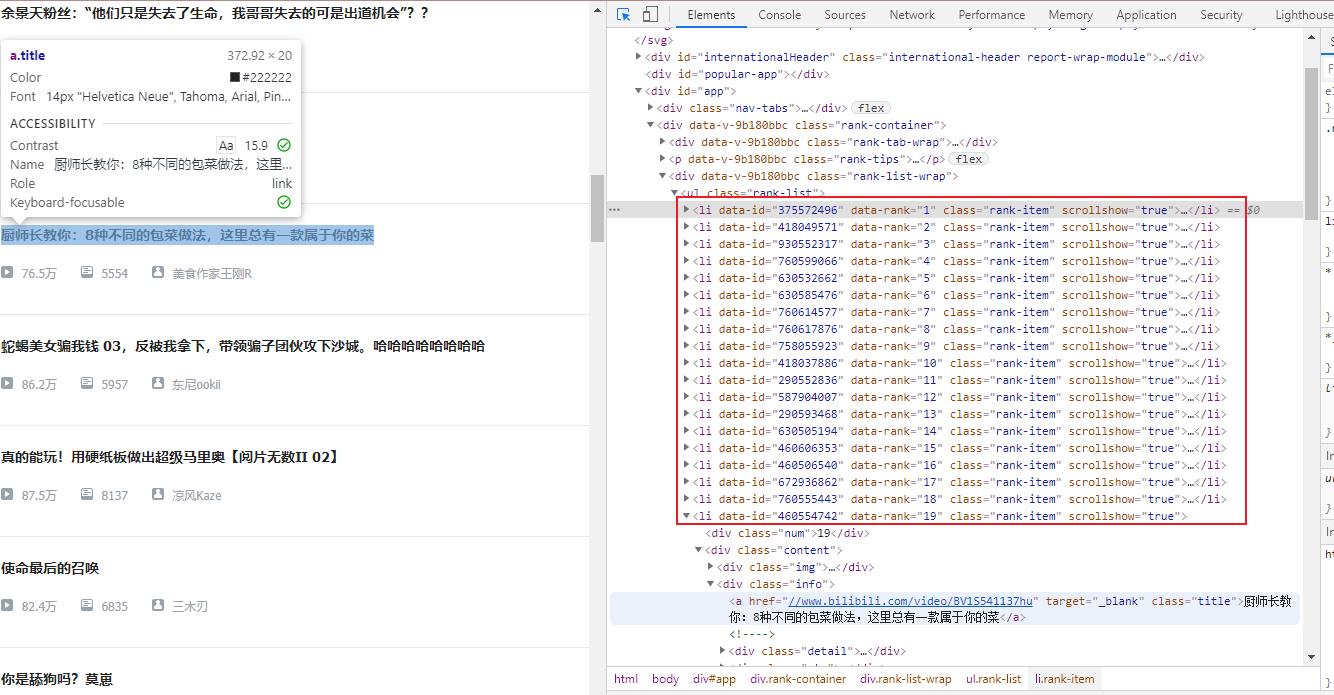
## 一、选题名称

B站综合排行榜数据爬取及数据分析可视化

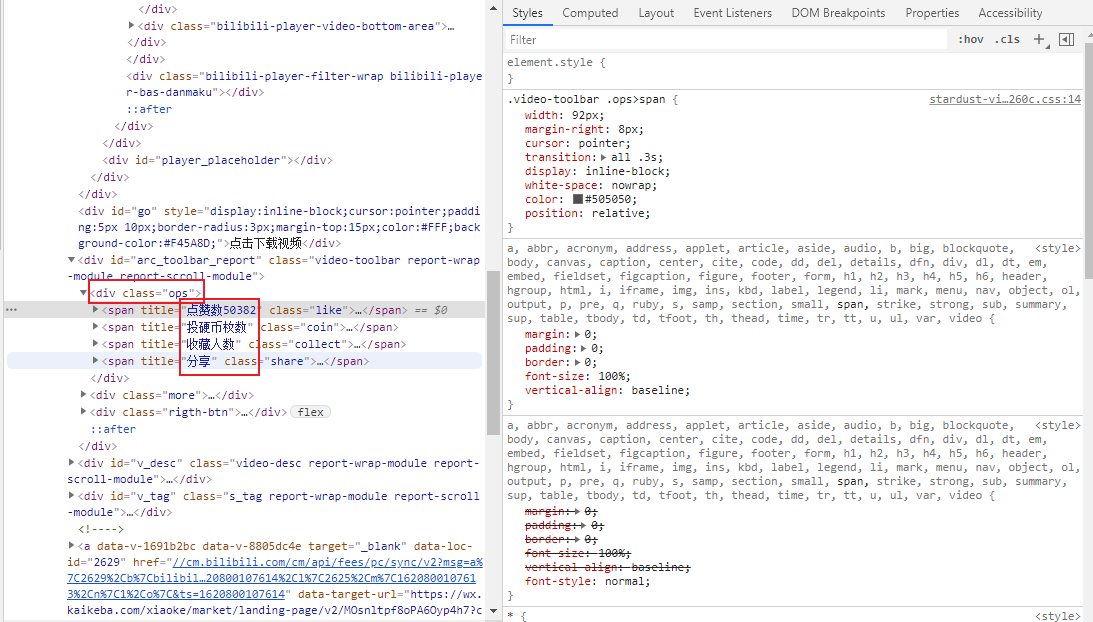
## 二、文档设计

### 1) 数据爬取过程：

首先分析哔哩哔哩排行榜（https://www.bilibili.com/v/popular/rank/all）的源码：



哔哩哔哩排行榜的内容只包括排名,视频名称、播放量、弹幕数、综合得分、作者、链接，并没有更加重要的投币、点赞、转发和收藏等关键信息。所以还需要对每一个页面内部进行爬取。

后续代码会根据这个页面的BV号进行更详细的爬取。BV代表的是哔哩哔哩每个视频对应的唯一的BV号，使用此bv添加网站后缀即可完成info页面的访问，在info页面中，可以进行进一步的爬取。

最后数据保存入csv文件中，预期的格式为



### 2)导入SQL server

由于播放量等信息带有“万”，不方便处理信息，后续需要转化为数字。接下来参考<https://www.cnblogs.com/nmsghgnv/p/11980663.html>，直接将python与sql server进行通讯和写入。

### 3) 数据分析

通过python端从sql server取出数据进行分析制作图表。预计使用python的matplotlib等模块进行可视化。最后进行文字分析。