**项目说明**

**姓名：张毅轩 学号：1120211228 班级：软件工程2021研**

2021年，据《晚点 LatePost》获悉，视频网站Bilibili（B 站）日活跃用户在近日已突破6500万，超越优酷成为仅次于爱奇艺与腾讯视频的第三大长视频平台。在用户使用时长上，B站已经稳定在80分钟以上。

B站最具备特色的就是弹幕文化，围绕着观看视频所发送的弹幕，B站的用户自发形成了独特的社区。根据业内统计，视频的观看量和发送的弹幕数量的比例在70：1到100：1之间，弹幕是具备很高信息量的数据。

基于弹幕具备自发区分性和高信息量的特点，我开发了针对B站弹幕的爬虫分析程序。程序由三部分组成：爬虫程序；词云图程序；LDA主题模型程序。其中爬虫程序用来获取B站的弹幕数据；词云图依据词频将弹幕数据可视化，观看更为直观醒目；LDA主题模型用来分析弹幕的主题内容。