取得的成果：

1. 整理模板

考虑到985、211高校的讲座资源较为丰富，根据全国985、211高校名单，小组成员寻找了各高校汇集讲座信息比较多的一两个公众号或官方网站，并收集这些网站和公众号2015年至2020年2月的讲座预告，再依据据表格对这些预告内容进行编码，编码内容包括标题、时间地点主题基本信息、主持人、主讲人、简介、推荐书目、参与方式、主办承办协办方八个个部分，由此整理出了严肃类讲座预告的模板。

C:\Users\H黄婧怡\AppData\Local\Temp\1607609556(1).png

图1.编码内容

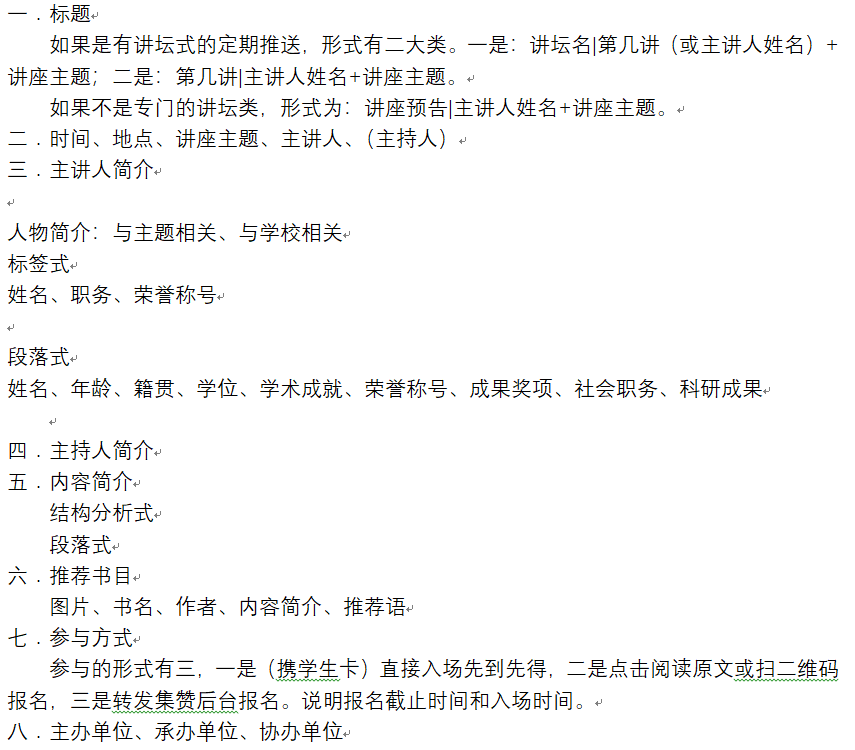


图2.模板内容

讲座类推送的内容一般分为讲座开始前的预告和结束后的总结两类，主要有严肃和活泼两种风格。考虑到讲座后的总结已经没有太大意义，受众更多地是想获取预告信息以便自己能参与讲座，所以我们选取了预告类；又因为严肃类较活泼类更容易整理出模板内容，小组成员先尝试从严肃类开始。

1. 公众号

公众号“今日人文社科”创立于2020年3月2日，前期已经设计好logo、二维码等品牌信息。公众号旨在成为一个人文社科类讲座信息共享平台，方便读者掌握讲座资讯，紧跟学科动态，聚焦顶尖高校，直击人文社科热点话题。目前公众号仍是人力写作与发布，最后将实现依据模板的自动生成内容。开创至今，公众号已发出34篇推文，内容包括话题和讲座总结两方面，话题关注新基建、数字经济、媒体融合；讲座总结则是团队成员依据自己关注的学科，每周听完学科内讲座后总结而成，领域包括计算机、新闻传播学和艺术学。



图3.公众号二维码

1. 数据库

小组已将爬取下来的讲座信息进行过数据清洗和人工预处理，并且已经将处理好的文本结构化，存储到服务器上搭建的数据库之中，数据库设计思路及数据之间的相互关系已在之前的ER图中给出。目前，数据库内存储了2015-2020年来1048条、12类不同学科的讲座信息，包括讲座题目、主办方、开始时间、结束时间、举办地点、简要介绍等；同时，对每场讲座的主讲人也进行了信息整理和存储，记录了每名主讲人的姓名、基本信息、主要成就、研究领域等。对非结构化的文本信息进行结构化存储，为后续的文本分类处理、自动讲座推送生成奠定了基础。

1. 网站搭建

目前，网站的搭建已经初步完成。网站首页上方的工具栏提供了登录界面和搜索界面，在登录界面上，用户可以登录自己的账户或注册进行操作；搜索方面，提供了两种搜索方式，一种是在搜索栏中通过输入进行关键词匹配搜索，进行搜索后，将会从数据库中返回存在相应关键词的对应讲座的题目、主办方、简要介绍等信息；另一种方式是通过学科下拉菜单选择对应学科的讲座信息，点击后将返回此类学科的所有讲座信息。未来将继续优化完善搜索界面，实现搜索结果的跳转功能，为每一条讲座信息设置单独的浏览界面，使用户能够直接通过点击讲座标题获取讲座具体内容信息。主页头图下方存有已整理好的讲座笔记和未来讲座预告信息。



图4 网站首页

存在的问题：

1. 进入公众号后的分类栏不知道该如何处理

打算在进入公众后最下面的分类栏设置“学科分类”和“关于我们”两部分，但是依据之前的人文社科学科分类，一共有哲学、经济学、管理学、法学、心理学、社会学、教育学、政治学、文学、历史学、新闻传播学、艺术学十二类，公众号分类栏无法显示这么多条分类，小组成员目前只想到通过“多次链接”的方法选课，但这种方法从用户体验角度来说稍显麻烦，不知道有没有其他解决方法。

1. 复杂内容无法整理出模板，可整理成型的模板单一

目前已经整理出的模板只有严肃类讲座预告，除此之外的多种讲座推送形式，由于风格多变、内容不一，无法整理出统一模板。也就是说，现有的成型模板形式单一。

1. 部分文本数据存在空缺，为后续处理带来了一定的困难

当前已经构建好的数据库中，不少数据项存在空缺，在使用模板进行文本的自动生成时会造成一定的困难，因此下一步需要考虑如何处理存在部分空缺信息的文本信息。

1. 未来讲座预告的整理

当前数据库中的数据均为已经结束的讲座预告信息，且对文本信息的处理都基于人工操作，面对未来大批量的新数据时处理仍有难度，需要寻找更为高效的方法；且目前仍缺少未来讲座预告的获取途径，需要进一步掌握并及时爬取数据新的讲座预告信息、进行处理后添加到数据库当中。