（一）进展情况

一、内容部分

1.前期数据收集

小组以 985、211 高校作为对象，选取哲学、经济学、管理学、法学、心理学、社会学、教育学、政治学、文学、历史、新闻传播、艺术共计十二个人文社科类专业，进行讲座预告数据收集，涉及如“新传小馆”“中国传媒大学”“中央美术学院人文学院”等数十家公众账号。小组通过各高校及其下属学院、研究所等机构组织的官网及微信公众号，收集讲座基本信息。并最终按照时间地点、主持人、主讲人、简介、参与方式、主办承办等部分进行分类，最终形成了以 1734 条来自各大高校讲座的数据，作为最终成果实现的数据基础。

2.公众号策划

（1）公众号设计

为给讲座资讯提供内容载体，小组创立微信公众号“今日人文社科”，并设计相应头像、二维码 logo 以及头图等，形成相对完整的公众号运营体系。小组将“今日人文社科”定位设定为人文社科类讲座信息共享平台，“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”是其创办理念。目前“今日人文社科”已发布 34 篇推送，拥有 28 名用户。

图 1 公众号 logo 与二维码设计

（2）专题推送

“今日人文社科”公众号开设“专题分享”模块，聚焦传媒领域前沿热点话题，为用户提供学界热点资讯、进行热门概念科普。目前该板块共有三篇内容，分别为《专题分享 | 什么是新基建？》《专题分享 | 什么是数字经济？》《专题分享 | 媒体融合的解构与建构》，对话题进行全方位、深层次的解读。如文章《专题分享 | 什么是新基建？》在对“新基建”给出严谨学界定义的同时，以直观图表形式介绍其“新”在何处。随后，推文分析了新基建如何被提出，并对其未来发展趋势进行了展望。整体多角度、多方位地呈现“新基建”带给社会的影响。

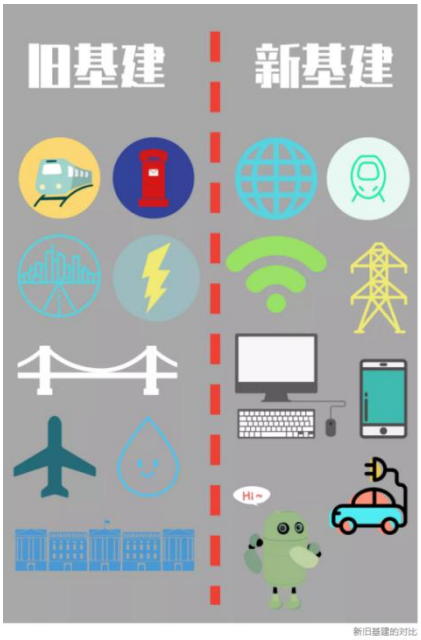

图 2 公众号推送《专题分享 | 什么是新基建？》截图

（3）讲座推送

目前“今日人文社科”公众号共发出讲座相关内容推送 31 篇，选题涵盖计算机、新闻传播、艺术等领域，聚焦学术前沿，为读者提供相应领域顶级学者的讲座记录或名校公开课笔记。如北京大学数学科学学院教授耿直讲座记录《因果推断——数据驱动的因果作用评价》，世界著名美术史家、芝加哥大学教授巫鸿讲座笔记《流通中的物与像，穿衣镜全球小史》，斯坦福大学 Coursera 公开课 CS228 概率图模型学习笔记，中国传媒大学中国通史系列、北京大学著名中国史学家叶炜兼顾总《魏晋南北朝》听课记录等等。讲座笔记的发布为读者提供多领域学术前沿追踪、顶级学术资讯，可增强用户黏性，从而为讲座推送自动生成后功能的不断完善提供用户基础。

图 3 公众号后台推送截图

二、技术部分

1. 数据获取技术进展

目前主要以高校官网公布的讲座信息为主要信息来源，使用 Python 进行公开讲座信息的收集工作，辅以人工手段对文本数据进行进一步处理（包括提取关键信息）。收集的数据目前已上传至 Github 仓库：*https://github.com/JJYDXFS/little-innovation/tree/master/Text_Materials*。

2. 网站搭建技术进展

项目最后的呈现方式为：微信公众号定期提供讲座分享+Web 端网站提供讲座资源聚合服务及自然语言相关的应用，本部分主要侧重于说明 WebApp 的搭建进展。网站首页地址为 *http://111.229.84.244:5000/*。目前使用 html/css/js+Flask 的前后端不分离形式进行网站 demo 的搭建，数据库使用 MySQL8.0。前后端分工为由前端编写网页模板、后端 Flask 主要负责对模板进行渲染和对，动态路由也由 Flask 实现。

（1）服务器环境搭建

服务器使用腾讯云服务器标准型 S2，配置为 1 核 2GB 1Mbps，操作系统 Ubuntu Server 18.04.1 LTS 64 位，硬盘大小 50GB。进行过基础 LAMP（Ubuntu18+Apache2+MySQL+Php7）环境搭建，前期曾使用该方式进行过前后端分离的开发。目前在网站默认端口 80 也同步了网站主页。

（2）前端

采用传统的 html+css+js 技术进行开发，基于 os-templates 的开源前端模板，根据项目需求对界面和功能进行改造。使用 JQuery 与服务器进行数据交换、实现动态加载；使用 Vue.js 简化 DOM 操作，后期考虑采用 Vue 组件化开发对网站前端结构进行升级。目前前端共编写了 3 个网页模板：首页模板、学科页模板、搜索结果页模板（其中首页模板编写进度较快，其余两个模板只有简单 demo）。

2.1 前端功能设计

<1>首页

首页预计实现如图所示功能，包括提供各类搜索功能和讲座预告信息，最终是否引入用户注册登录功能暂时未确定。

图 4 主页设计

<2>学科页

学科页预计将提供学科词云和相关的预告、笔记。学科页功能设计如下。

图 5 学科页设计

<3>搜索结果页

搜索结果页将根据关键词给出搜索的讲座信息，搜索结果将按匹配程度排序，并提供部分内容的预览。

图 6 搜索结果页设计

2.2 模板实现进展

<1>网站首页模板

如下图所示，使用 GET 请求初步实现了导航栏的动态加载和点击跳转以及搜索框的自动跳转功能。

图 7 主页截图

<2>学科页模板

学科页暂未进一步细化模板，直接使用 POST 请求向后端调取对应数据显示在页面上。如图示结果为“艺术”学科页的当前效果图。

{ "title": "中国歌曲一百年", "organizer": "华南理工大学艺术学院", "introduction": "" }

{ "title": "“西而化之”的中国钢琴音乐创作——高为杰与高平对谈", "organizer": "华南理工大学艺术学院", "introduction": "" }

{ "title": "中国艺术之诗、书、画、印——识印\n", "organizer": "", "introduction": "本次讲座是图书馆推出的《中国艺术之诗、书、画、印》系列艺术讲座的首讲《识印》。“方寸之间，气象万千”，讲座为师生读者详细讲呈中国的印章艺术，深入介绍篆刻的历史延承、艺术流派和发展格局，精讲篆刻艺术的相关知识和技法，个人对篆刻的独特理解和阐释以及篆刻的美学文化。" }

{ "title": "贝多芬最后的奏鸣曲op.111的秘密", "organizer": "华南理工大学艺术学院", "introduction": "" }

{ "title": "中国民族音乐的时代创新与国际推广", "organizer": "华南理工大学艺术学院", "introduction": "" }

{ "title": "从专业技能型教育到社会主题型教育", "organizer": "华南理工大学设计学院", "introduction": "演讲将首先解读近年来被广泛倡导的学科交叉理念，接着从设计实践、设计学科自身规律，以及时代语境等角度，分析设计教育改革的必要性和路径。演讲还将结合设计职业趋势和设计影响力的拓展，结合实际案例，分析设计研究对象、设计方法和准则的变化。" }

{ "title": "法国动画片的美学内涵和技术创新", "organizer": "华东师范大学传播学院", "introduction": "" }

{ "title": "当代海派纪录片系列展映暨映后导演交流会", "organizer": "华东师范大学传播学院", "introduction": "本次展映共五天，播映五部当代海派纪录片，导演们亲临现场参加映后交流。" }

图 8“艺术”学科页截图

<3>搜索结果页模板

搜索结果页暂未进一步细化模板，直接使用 POST 请求向后端调取搜索结果（标题中含有关键词的讲座信息）显示在页面上。如图示结果为关键词“世界”的搜索结果。

{ "title": "不忘初心 创世界一流航空照明企业", "organizer": "华南理工大学材料科学与工程学院", "introduction": "", "field": "others" }

{ "title": "晚清时期的报刊阅读与思想世界", "organizer": "华南理工大学人事处", "introduction": "通过报刊阅读的理论探讨，扩展报刊“思想版图”与读者“思想世界”的认知。在阅读史的层面上理解报纸，就不仅仅将其视为传播信息的载体，而是广泛意义上的“知识纸”“思想纸”“政治纸”，报纸不仅提供新闻，它是一种系统的精神消费品，通过读者的阅读，报纸与周遭的世界建立广泛的联系，为读者建构了复杂的“意义之网”。“读书人”向“读报人”的身份转变，其背后往往存在着“古典”与“现代”“传统”与“时尚”“保守”与“先进”等理念的认知过程。报刊既为读者建构了“情感共同体”，也为读者提供了“思想版图”。因此，当现代报刊进入读书人的阅读世界，“读报人”必将成为观察晚清社会的一道绮丽的风景，“读报人”的所思所言所行，不仅是报刊阅读的历史轨迹，也是社会变迁的一个缩影。", "field": "journalism" }

{ "title": "纪录片《如影而形》 境内的第三世界", "organizer": "华东师范大学纪录影音档案馆、华东师范大学传播学院纪录片研究中心、华东师范大学图书馆联合主办", "introduction": "《如影而行》的纪录以民众戏剧工作者：钟乔为拍摄重点，内容涵盖“差事剧团”25年来在两岸底层民众间的戏剧演出及培力工作。概括来说，主要呈现如何在两岸的境内追寻第三世界身体的戏剧性.这是一件需要深思且发生在我们日常生活周遭的事情。\n本片的重要内容有一部分是北京“皮村”的农民工如何在改革开放的浪潮下，面对资本与发展代价的矛盾，无论就人或环境本身来探究，都与境内的第三世界发生密切的联想。而本片也以相同的观点，将目光移向岛内彰化台西村：一个隔着濁水溪5公里之遥，备受pm2.5污染却长久以来无人闻问的小村庄。如果，以剧场的身体行动出发，见到的将是：环境问题即是阶级问题。因为，留在污染前线的受害者，恒然是底层的民众…\n在这样的前提下，此项活动聚焦于如何让两岸的底层民众，看见剧场与民众发生关系时，我们如何进行两岸间透过剧场所展开的文化对话。\n2015年，纪录片导演黄鸿儒，在历经四年的拍摄，交出一张令人惊艳的成绩单：《如影而行》。这部纪录片不仅纪录到了当前钟乔的工作与活动，也清楚梳理了一个台湾左翼文化实践者所置身的时空脉络。", "field": "art" }

{ "title": "世界最大的对冲基金桥水投资公司的投资原则及行业现状与展望 ", "organizer": "中国传媒大学经济与管理学院、MBA学院 ", "introduction": "美国桥水投资公司成立于1975年，至今已有39年历史，是世界上最成功的资产管理公司之一，在2010年和2011年被评为全球规模最大且表现最优异的对冲基金，长期拥有极高的客户满意度评分。 ", "field": "economics" }

{ "title": "世界传播学发展与变革的态势 ", "organizer": "中国传媒大学新闻传播学部新闻学院主办", "introduction": "", "field": "communication" }

图 9 关键词“世界”搜索结果

（3）后端

使用 Python Flask 框架，目前实现了学科页和搜索结果页的动态路由、上述三类模板页的渲染以及三个接口：

a) GET：获取学科列表

b) POST：根据学科返回该学科包含的讲座信息

c) POST：根据关键词返回标题中含关键词的讲座信息

接口具体实现：

https://github.com/xiaochuang-JRRWSK/JRRWSK_web/blob/main/API/JRRWSK-2020-11-29.png

（4）数据库

数据库 ER 图如图所示，目前数据库中一共收录讲座数据 1048 条，涵盖学科 12 类。

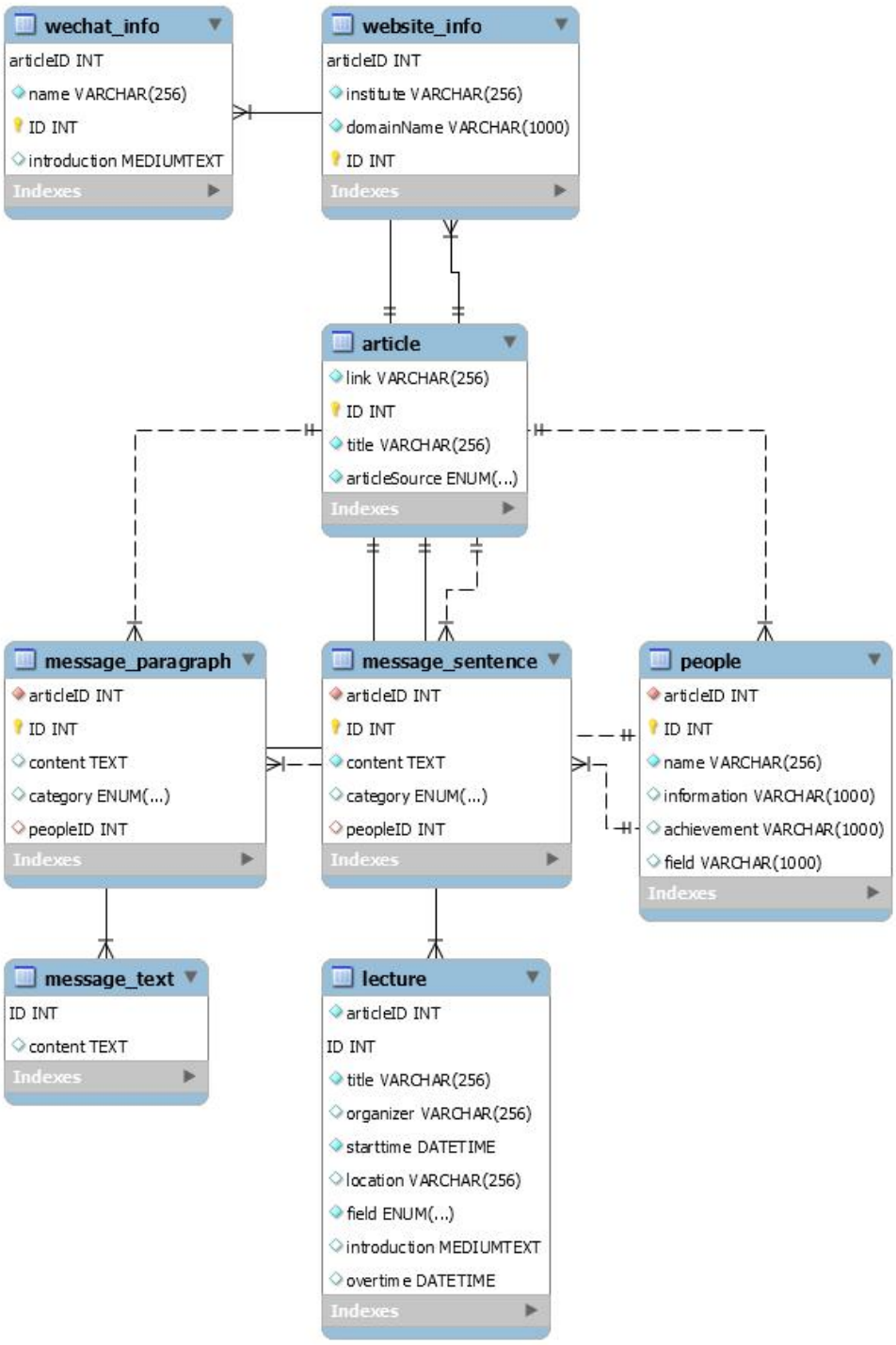

图 10 数据库 ER 图

（二）开支情况

小组购置了一台腾讯云云服务器，用来搭建项目网站。有效期从 2020 年 2 月 25 日至 2021 年 2 月 25 日，支出 99 元。到期后仍会续费。

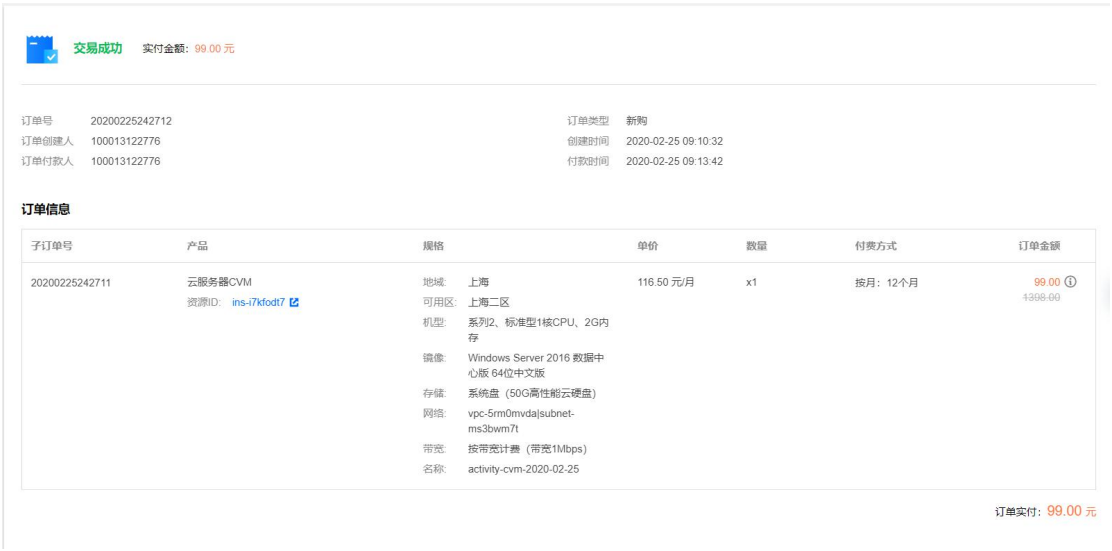

图 11 腾讯云服务器订单截图

（三）取得成果

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、主办方、承办方八个个部分，由此整理出了严肃类讲座预告的模板。

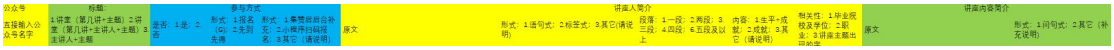

图 12 编码内容

一．标题

如果是有讲坛式的定期推送，形式有二大类。一是：讲坛名|第几讲（或主讲人姓名）+讲座主题；二是：第几讲|主讲人姓名+讲座主题。

如果不是专门的讲坛类，形式为：讲座预告|主讲人姓名+讲座主题。

二．时间、地点、讲座主题、主讲人、（主持人）

三．主讲人简介

人物简介：与主题相关、与学校相关

标签式

姓名、职务、荣誉称号

段落式

姓名、年龄、籍贯、学位、学术成就、荣誉称号、成果奖项、社会职务、科研成果

四．主持人简介

五．内容简介

结构分析式

段落式

六．推荐书目

图片、书名、作者、内容简介、推荐语

七．参与方式

参与的形式有三，一是（携学生卡）直接入场先到先得，二是点击阅读原文或扫二维码报名，三是转发集赞后台报名。说明报名截止时间和入场时间。

八．主办单位、承办单位、协办单位

图 13 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。

2. 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 14 公众号二维码

3. 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。

4. 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。

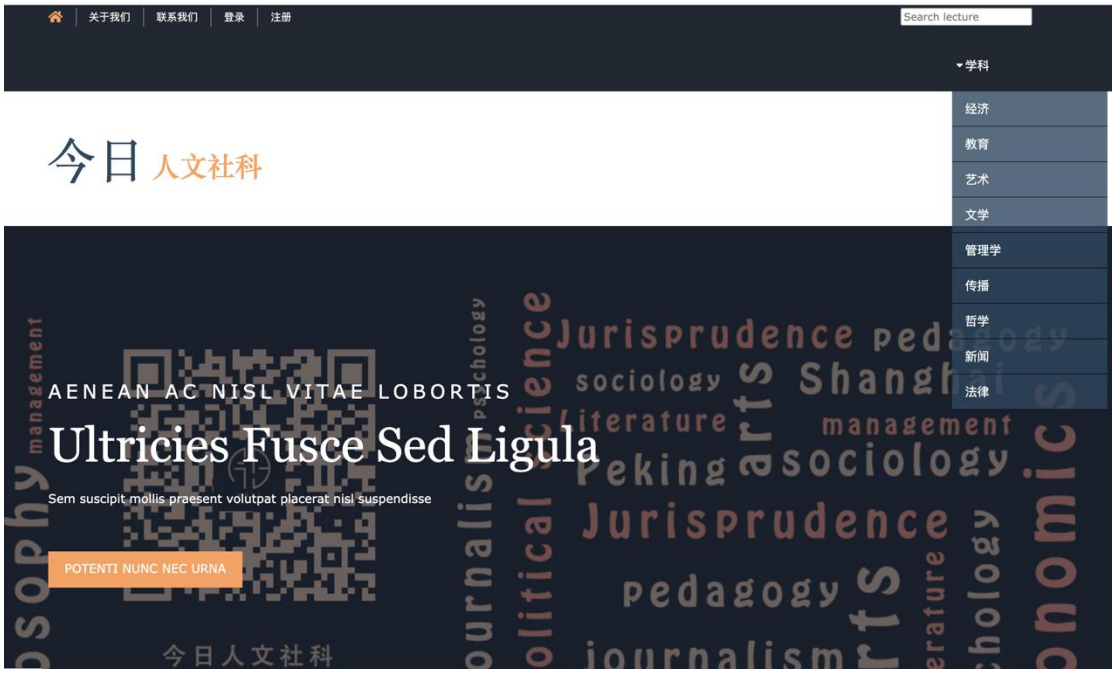

图 15 网站首页

（四）存在的问题和困难

一、问题和困难

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。

2. 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。

3. 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。

二、拟解决的方案：

1. 设计和整理更多的模板，并设计多种模板之间的组合方式，利用自然语言处理技术，对不同粒度的文本进行组织和排列。

2. 设计默认填充的算法方案，解决数据缺失的问题。优先使用完整数据，并在数据有缺失的情况下争取填充，使每一条数据物尽其用。

3. 扩大数据来源，保证数据源的时效性和稳定性，争取使用爬虫的方式，实现半自动定期更新数据。

（五）建议和要求

1. 当前工作重点：

（1）完成网站的开发

（2）完善公众号的运营细节

2. 工作组织方面：

（1）保持沟通。根据工作相关人的数目，灵活召开工作会。并保证工作内容每周跟进一次。

（2）有效分工。根据每个人的工作特点，合理分配任务，保证小组工作的有序推进。

3. 后期工作规划：

（1）首先把网站搭建好，再对后端包括但不限于文本生成算法进行优化。

（2）扩大数据库，找到更稳定的数据来源，减少人工成本。

（3）丰富公众号内容，传递有价值的信息，吸引更多用户。

（4）尽可能创造项目经费来源。