

2020 年中国排名前一百大学知识图谱

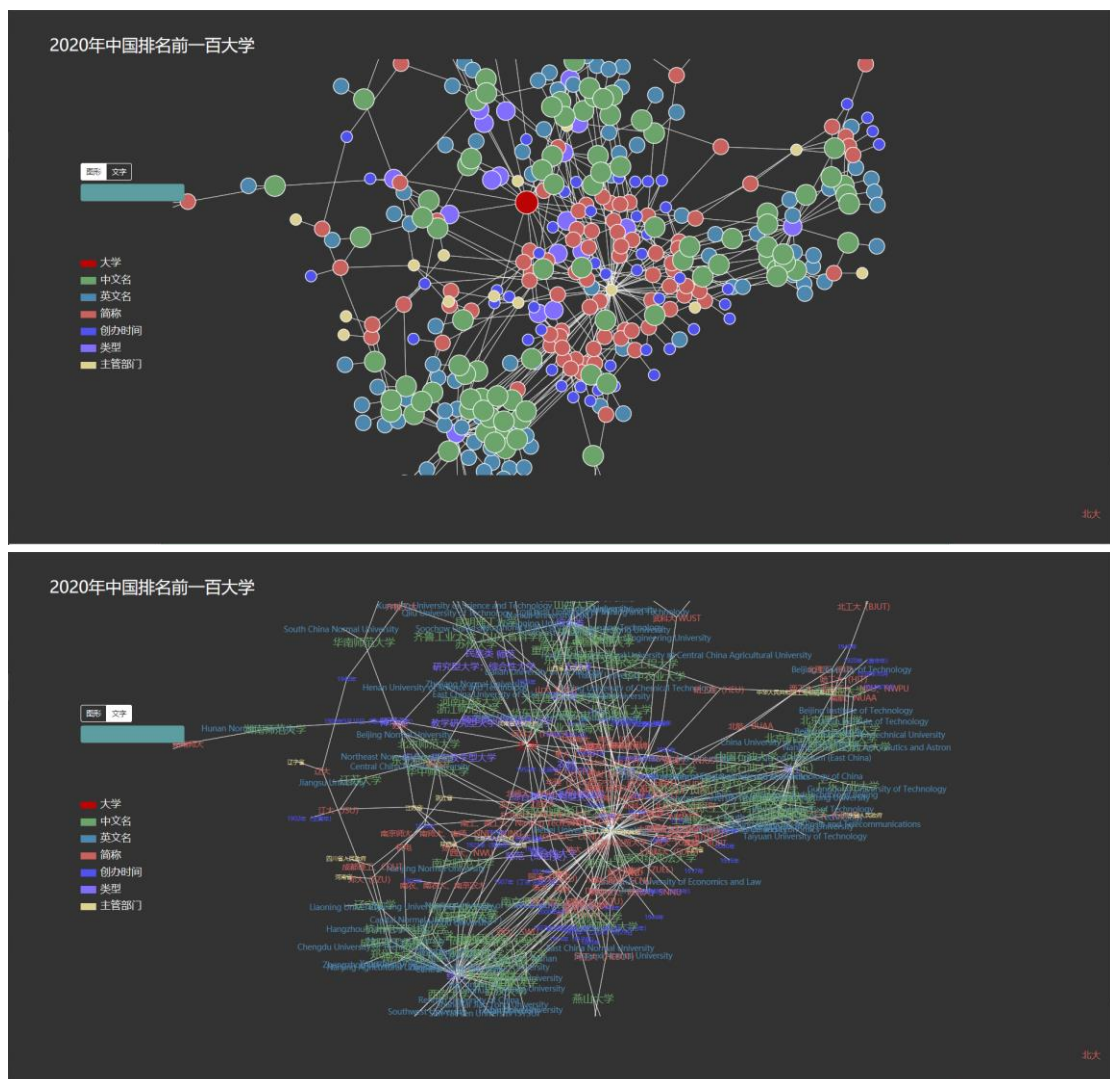
本项目是根据 2020 年校友会发布的中国前一百大学排行榜来获取中国百强大学的相关信息。提取的主要信息包括有大学的中文英文名称，大学的缩写，大学所属的类型，大学的主管部门和创办时间等等。

数据的来源主要是来自于百度百科的这一百所大学的相关词条，从网络上获取到 100 所大学的名称之后通过爬虫技术爬取百度百科资料，从而获得相关信息，主要是在“基本信息”这一项目板块中的信息。

中文名	浙江大学	博士后	科研流动站57个
外文名	Zhejiang University		科研工作站3个
简称	浙大、ZJU	国家重点学科	一级学科14个
创办人	林启		二级学科21个
创办时间	1897年5月21日	院系设置	7个学部、37个专业学院（系）、1个工程师学院、2个中外合作办学学院、7家附属医院
类别	公立大学	校训	求是创新
类型	综合	校歌	《浙江大学校歌》
属性	211工程（1995年入选） 985工程（1999年入选） 世界一流大学和一流学科（2017年入选） 九校联盟，环太平洋大学联盟 高等学校创新能力提升计划（2013年入选） ▼ 展开	校庆日	5月21日
主管部门	中华人民共和国教育部	地址	紫金港校区：杭州市西湖区余杭塘路866号 玉泉校区：杭州市西湖区浙大路38号 西溪校区：杭州市西湖区天目山路148号 华家池校区：杭州市江干区凯旋路268号 之江校区：杭州市西湖区之江路51号 ▼ 展开
现任领导	党委书记：任少波 校长：吴朝晖 [7]	院校代码	10335
专任教师	中国科学院院士：23人 中国工程院院士：23人 [8]	主要奖项	全国百篇优秀博士学位论文48篇 国家级教学成果奖58项（2001-2018年） 国家科技进步特等奖1项、一等奖6项、二等奖37项-十二五期间
本科专业	128个	知名校友	竺可桢、陈独秀、厉绥之、路甬祥、李政道、丁仲礼、段永平、史玉柱等
硕士点	学术学位授权一级学科62个 学术学位授权二级学科309个 专业学位授权类别27个		
博士点	学术学位授权一级学科59个 学术学位授权二级学科274个 专业学位授权类别4个		

在数据的爬取过程中，不可避免地会出现有部分学校有部分特征缺失，在这边我们使用“无信息”标签进行替代。最后获取到的数据中有综合、师范、理工、农林、财经、政法、民族等 7 个大类的大学。最后将所有爬取到的数据汇总，存入 json 和 csv 文件中。之后将所提取的特征分别取出，存入到不同的 txt 文件中，为之后建立 node 和 link 做准备。

接下来使用 D3.js 的模板进行数据的可视化。可视化首先就是需要建立 node 和每个 node 之间的 link，将每个特征的每一个值都提取出来将其作为 node，并且对于每个有联系的 node，存在关系的实体，我们使用 link 将其链接起来。在可视化的过程中对于不同类型的 node，使用不同的颜色进行标注，link 就是简单的使用线段来进行 node 之间的连接。同时既有图形显示，也可以选择文字信息展示，最后得到的结果如下图所示：



与此同时，此项目还能完成节点的详细信息以及其关联性息的查找，例如如果需要查询“浙江大学”的相关信息，可以在搜索栏中输入浙江大学，同时也能够将鼠标移动到相应的 node 上，这时候会自动出现相关联的信息以及本身

的信息。

