



扫二维码 看科学报

扫二维码 看科学网

扫二维码 科学周末

扫二维码 医问医答

总第 6818 期
2017年6月23日 星期五

今日 8 版

官方微博

新浪: <http://weibo.com/kexuebao>
腾讯: <http://t.qq.com/kexueshibao>—2008

国内统一刊号:CN11-0084 邮发代号:1-82

大数据寻诗访古

■本报记者 张文静

大数据与文学,一个是理性工具,一个是感性思维,看起来似乎不沾边。但如今,二者的联系却日渐紧密起来,也由此引发了不少争议。

谁是最爱往外跑的诗人?

唐宋时期最爱往外跑的诗人是谁?答案可能是苏轼。

打开“唐宋文学编年地图”,点击苏轼的名字,地图上立刻显示出密密麻麻的足迹,西到雅安,东到蓬莱、青浦,北到定州,南到海南南部的陵水。从青年时代开始,苏轼的脚步就一直没停过,在他 58 岁那年甚至一口气走了 31 个地方,堪称中国古代一股“行走的力量”。

“苏轼一生的轨迹信息高达近万余,遍布全国各地,是我们录入信息最多的一位唐宋诗人。”“唐宋文学编年地图”的建立者、中南民族大学教授王兆鹏告诉《中国科学报》记者。

“唐宋文学编年地图”在今年 3 月上线,是王兆鹏主持的国家社科基金重大项目“唐宋文学编年系地信息平台”的一个研究成果。这个

电子地图的形成花费了五年的时间,100 多人的团队参与其中,负责数据的整理编写,融合了地理信息系统、测绘、计算机、文学等多个领域的内容。

目前,这份地图已录入了 100 多位唐宋诗人的行迹信息,范围北至蒙古乌兰巴托,南至越南清化。点击任何一位诗人的名字,地图上就会出现他一生的行走路线图,再点击任一地点,这位诗人在此创作的诗歌作品就会展示出来。如以地点或年份为关键词进行检索,则会看到某地在某个时间段共有多少位诗人来过、留下了哪些作品。于是从地图上,我们可以看到,李白出生于西域碎叶城(今吉尔吉斯斯坦托克马克市),王维出生于晋中,12 岁的孟浩然在襄阳居乡读书,45 岁的宋之问则从郑州来到洛阳又到西安,写下了多首诗作。“这张地图的最大亮点就是打通了时空维度。”王兆鹏说。

地图上线后,火爆程度让王兆鹏大吃一惊,他没想到这样一份文学地图引起了大众这么多关注。“本来项目是年底结题,我们还没着急,没想到一下子火了,上线第一天的点击量超过了 100 万,两天就到了 220 万。大家一直在问,为什么没有某某诗人,所以我们现在必须要加班加点,尽快将所有诗人的信息传上去。”王兆鹏说。

用大量数据来展现唐宋诗人的故事,不仅有文学专业的教授在做。今年 3 月,一篇名为《计算机告诉你,唐朝诗人的关系到底是什么样的?》的文章刷爆朋友圈,很快达到了 10 万+的阅读量。这篇文章来自一位普通的程序员“前进四先生”之手,发布于他的个人微信公众号“前进日志”中。

在对四万多首唐诗进行了数据整理后,“前进四先生”发现在唐朝,两位关系最好的诗人不是李白和杜甫,也不是白居易和元稹,而是陆龟蒙和皮日休。这两位诗人互相提到对方的次数都在百次以上,中国文学史上的第一本唱和诗集《松陵集》也是他俩的作品。从排名前 30 的引用关系来看,白居易绝对是唐朝诗人朋友圈中的明星。

大数据与小阅读

用数据统计的方法来做文学研究,王兆鹏早在 1992 年就开始了。“当时我是系里主管研究生工作的副主任,偶然一次机会在杂志上看到一篇定量分析研究生学位教育的文章,我就想古代文学史的研究能不能也用定量分析的方法来做。后来我写了一篇文章《宋代词人历

史地位的分析》,就是用量化数据来描述词人的地位,比如什么叫地位很高、比较高或一般。从那时起,我就开始了这个领域的研究。”王兆鹏介绍说。

在国外,也有展现中国历代人物生平资料的数据库,比如由哈佛大学、北京大学、台湾“中研院”合作开发的 CBDB 数据库,通过字号、亲属关系、生卒年份等数据,展现人物的社会关系网。

对于大数据手段与文学研究的关系,在华东师范大学教授金雯看来,使用电脑算法来分析文本,不是让电脑复制人脑的功能,或者更大规模地完成人脑擅长的任务。人脑和电脑在阅读文本的时候所用的方法和关注的重点不一样,读出来的东西也可能截然不同。不过人脑和电脑在阅读阐释文字的时候也往往可以互为体用、互补短长,文学大数据分析和学者个人的“小阅读”之间存在着许多交融和合作的可能。正因为如此,借助电脑进行文本分析是近年来不断升温的“数字人文”的一个重要分支。

几年前,王兆鹏利用数据分析进行过另一项研究——唐宋词排行榜,曾招来过不少争议。

2011 年,王兆鹏出版了《唐诗排行榜》一书,运用统计学方法得出了唐诗前 100 名排行榜,排在榜首的是崔颢的《黄鹤楼》,其次是王之涣的《凉州词》、杜甫的《登高》、王之涣的《登鹳雀楼》和张继《枫桥夜泊》等,被大众熟知的陈子昂的《登幽州台歌》等诗作则名落孙山。2012 年,他又出版了《宋词排行榜》,将《念奴娇·赤壁怀古》列为宋词第一名。

“这是通过对历代选本、评点、论文、网络链接总数等指标综合计算而来的。”王兆鹏说,自己的数据采集分为三个方面,一是作家数据,包括生卒年月、创作起始时间、出生地、去世地、活动地点和在社会上扮演的身份等;二是作品数据,即作品的分类、版本、编年、系地等;三是读者数据,包括普通型读者、专家型读者和作家型读者三项。

两本书出版后,立刻有人质疑:“对古典文学的艺术鉴赏也能列排行榜?”“甚至有人说,是不是因为你是湖北人,所以把唐宋词第一名都给了写湖北的?”王兆鹏说。但在他看来,这两个排行榜是将现代科学手段引入到古典文学作品的研究赏析中的尝试,是严肃的学术研究。“而且,我评价的不是一首诗的好坏,而是评价它的影响力和知名度。”

但在中国科学院数学与系统科学研究院研究员安鸿志看来,“影响力”仍然是个模糊的指标。“要给唐宋词做排名,目的不同,指标



不同,得出的结论就可能截然不同。如果我们要办的是跑步比赛,首先就要确定跑一万米还是一百米,男子还是女子,个人还是接力,一旦确立指标,比如男性 100 米个人短跑,那么世界第一就是唯一解。可是对于唐宋词,会有唯一解吗?同样是看影响力,如果指标是‘哪首唐诗被现代人记住的最多’,那有可能是《静夜思》;如果把指标设为‘哪些诗句在日常生活使用频率最高’,答案又可能会是‘粒粒皆辛苦’。话说回来,没有唯一解就不能去研究吗?”安鸿志说,“当然也不是,这项研究是有价值的。但你要承认,指标不唯一,得出的结论也不唯一,这才是科学方法、科学态度、科学结论。”

正确利用大数据技术

在王兆鹏看来,哪首诗词排第一名倒也在其次,这些诗词为何能从古至今一直为人们所熟知,这其中有何传播规律,更能引起他的兴趣。“比如,除了诗本身之外,故事对作品的传播有着非常重要的助推力。”王兆鹏解释说,“崔颢的《黄鹤楼》在古代非常有名,为什么?一个重要原因是李白很欣赏。李白到黄鹤楼后想写一首诗,后来因为‘崔颢题诗在上头’便搁笔未写。这在宋代还成了一句俗语‘莫因崔颢题诗’。所以历代唐诗选本,没有不选《黄鹤楼》的。王之涣的《凉州词》也有‘旗亭画壁’的故事流传至今。”

此外,这些唐宋词大数据还能告诉我们一些其他的信息。“一般认为,北宋王朝的毁灭代表了中国经济文化重心南移的真正分野。但通过大数据,我们可以看到,中国诗坛重心的南移始于晚唐五代,完成于北宋。唐宋诗歌版图南移的时间和社会经济重心南移的时间并不完全一致,这打破了我们一些固有观点。”王

兆鹏说。通过数据整理,王兆鹏还统计出,唐宋诗人创作的优秀作品大多是在落后地区完成的,在失意受贬谪的地区更可能诞生名篇。“而且优秀作品的作者大多不是本地人。比如黄州、惠州打名片爱说苏轼,可苏轼是四川人;写黄鹤楼那么多名篇,可没有一个人是湖北人写的。”

王兆鹏正在做着更多探索。他计划将更多历史信息融入到现有的电子地图中去,比如某年哪些人中了进士,哪些人做了官、他们之间有何相互影响等。“数字技术正在改变我们观察文学、思考文学的方式。以前由于纸本的局限,我们每次只能看一位人物或一个时期。现在利用信息技术,我们可以把大大小小的人物同时放在一个舞台上,就像把树与树之间隔着

的围墙打破后展现出一片森林一样,把历史的一个个横截面完整地呈现出来。”“大数据技术作为人类认识客观世界的一种工具,给我们带来的进步是难以想象的。从研究者的角度来说,以前我们要到图书馆去查阅很多资料,现在只要坐在屋子里,摆上一台电脑,连接上网络,海量的资料就送到了眼前。”安鸿志说,“大数据资料既可影印,又可编辑,我们可以从中查询关键词,提取信息,进而进行修改、识别、对比。大数据技术是时代的象征,各行各业都要适应这个潮流,在文史哲研究领域当然也是如此。”

安鸿志同时也强调,大数据技术是客观的,但一旦有人使用它达到某种目的、得出某种结论时,就有了主观性,涉及到方法、目的以及表达等问题。“在这个时代,谁往大数据库中‘增砖添瓦’都是一种贡献,但对得出的结论要慎重。你也可以对某位研究者得出的结论提出质疑,但不能将其归罪于大数据技术。这两件事情要分清。”

周末聊吧

文明礼让 制度先行

■胡琅琦

最近,河南驻马店一名女子在斑马线上被自行车撞倒,后遭二次碾压的新闻受到了许多公众的关注。实际上,被称为“生命线”的斑马线没能给行人一个安全的空间,在国内一直是普遍存在的问题。

据公安部交管局最新的统计数据显示,近三年来,全国共在斑马线上发生机动车与行人的交通事故 1.4 万起,造成 3898 人死亡,其中机动车未按规定让行导致的事故占了总量的 90%。

《道路交通安全法》中明确规定,“机动车行经人行横道时,应当减速行驶;遇行人正在通过人行横道,应当停车让行。”但对很多司机来说,这只是交规的考试内容而已。

归根结底,国内机动车司机主动避让行人意识不足,很重要的原因在于,他们把行人当做和自己一样的交通对象对待,而不认为行人在道路中其实是弱势群体。

而真正考量一个社会文明程度的,常常是它对弱势群体态度。这种态度当然一方面也是要求司机文明礼让,可在现实中,这种自觉的养成首先需要依靠制度设计。

既然行人在道路上是弱势一方,尤其是在步行人数较多的地方,相关部门是否制订了向行人倾斜的交通安全设施的设置目标和计划。比如,增加人行横道的绿灯时长;扩充按钮式信号灯;在没有红绿灯设置的斑马

线地点也有显眼的指示装置;在一些没有信号灯的路口,给司机设置停止线,要求其必须停下,先观察路口情况等等。在美国和日本,“停一停”的指示设置,为交通事故的下降产生了十分积极的影响。

因此,对弱势群体做出特别保护不能仅仅停留在倡议和笼统的规定上,而是必须在制度和管理层面真正做到科学化、人性化和精细化。

除此之外,像新加坡这样以严厉的处罚制度闻名的社会,事实上也提供了解决文明缺失问题最有效的一种方法——让每一个不遵守规则的人付出最昂贵的代价。

它的前提是有严格细致的法律条文明确规定处罚的规则,它不能只针对机动车司机,还得包括行人的行为。在新加坡,行人闯红灯同样属于违法行为,有从罚款、强制教育到监禁等不同程度的处罚内容。

重罚之下必有成效,酒驾查处就是国内最好的例子。目前,北京地区已经开展了路口秩序环境综合治理,规定机动车遇人行横道不避让行人,将罚款 200 元、记 3 分。类似的举措在成都、杭州、深圳也已经纷纷施行。如果我们还无法让人性都能产生尊重和

看点

公共服务领域英文译写国家标准发布

国家质检总局、国家标准委 20 日联合发布《公共服务领域英文译写规范》系列国家标准。这是保障公共服务领域英文翻译和书写质量的基础性标准。

标准规定了公共服务领域英文译写的普遍性原则和要求;交通、旅游、文化、娱乐、体育、教育、医疗卫生、邮政等 13 个服务领域英文译写的原则、方法和要求,并为各领域常用的公共服务信息提供了规范译文。

标准规定,公共服务领域应当针对实际需要使用英文,不应过度使用英文,译写时应通俗易懂,便于理解,避免使用生僻的词语和表达方法;译写应用语文明,不得出现有损我国和他国形象或有伤民族感情的词语,也不得使用带有歧视色彩或损害社会公共利益的译法。该标准于 2017 年 12 月 1 日起正式实施。

全国 338 个城市水环境质量将排名

环保部日前正式印发实施《城市地表水环境质量排名技术规定(试行)》,这意味着水环境质量排行榜将在不久的将来与公众见面。

环保部将对全国 338 个地级及以上城市水环境质量、变化情况每年进行 4 次排名,即第一季度、上半年、1—9 月及全年,分别于 4 月、7 月、10 月及次年 1 月公布地表水环境质量较好、较差的 10 个城市及水质改善、恶化程度相对较大的 10 个城市名单。

城市地表水环境质量排名包括两部分:一是城市地表水环境质量状况排名,反映的是城

市地表水环境质量现状,可作为评价城市地表水环境质量优劣的依据。

城市地表水环境质量状况排名按照城市水质指数(CWQI)从小到大的顺序进行,排名越靠前说明城市地表水环境质量状况越好。二是城市地表水水质变化规律排名,反映各地水污染治理的成效,排名基于城市水质指数变化程度,排名越靠前说明城市地表水环境质量改善程度越高。

我国将创建 100 个稻渔综合种养“国标区”

从今年开始,农业部将用三年时间创建 100 个国家级稻渔综合种养示范区,力争将全国稻渔综合种养面积达到 3000 万亩以上。

据了解,稻渔综合种养作为一种高效、生态环保的新技术,在我国已推广 10 年,正被各地广泛接受,这种技术的参与让稻田效益大幅提升。湖北农业部门实地测产验收表明,稻渔综合养种的效益是稻田单作的 3 至 10 倍。

目前农业部已启动相关工作,并公布了相关标准,要求示范区周边无工业污染源,近 5 年没有发生重大污染事故或生态环境破坏事件,近 3 年没有发生水稻和水产品重大质量安全事件;种养区土壤环境质量要达到二类以上、水质达到三类以上标准;比如平原地区需集中连片,面积达到 3000 亩以上,水稻亩产不低于 500 公斤,沟坑面积不能超过总面积的 10%,化肥、农药用量需比水稻单作减少 50% 以上,亩均利润要比水稻单作提高一倍以上,且亩均增加利润不低于 2000 元等。

苏州纳米所入选国家第二批大众创业万众创新示范基地

6 月 21 日,国务院办公厅印发《关于建设第二批大众创业万众创新示范基地的实施意见》(国办发[2017]54 号),苏州纳米所作为江苏省唯一的科研院所入选国家“大众创业、万众创新”示范基地。

作为院地共建的创新型研究机构,苏州纳米所在建所之初即确立了“致力纳米科技创新,引领新兴产业发展”的建所理念。经过 10 年的创新发展,研究所基本形成了由新材料、纳米器件及应用、纳米生物医学、纳米仿生等研究方向组成应用基础研究体系,由纳米纤维与薄膜中心、GaN 衬底开发中心等工程中心组成产业研究与推广体系,由纳米加工平台、测试分析平台、纳米生化平台等向社会全面开放的公共服务平台构成的技术支撑体系,成为苏州纳米产业发展的技术策源地。在院地双方的大力支持下,十年来,苏州纳米所累计引进培育了 300 余家高新技术企业,纳米技术公共服务平台服务企业超过 1000 余家,培训人员达 1.4 万人次,为苏州纳米技术产业发展提供了强有力的支撑。

获批全国“双创”示范基地,将进一步激发苏州纳米所的创新创业活力,也为苏州区域经济社会发展带来新的机遇。苏州纳米所高度重视,已成立推进“双创”工作领导小组,完善组织体系,加快双创各要素集聚,优化双创生态系统,进一步聚焦苏州产业发展重点方向,借力江苏、苏州产业资源、政策资源,聚集创新要素,提升苏州产业层次,服务城市转型升级,打造“硅谷式”的一流双创服务体系,为“两聚一高”工作作出应有的贡献。(周天整理)

蒋志海制图