

·旧锦新样

专业团队携手大众爱好者用智能技术助典籍“重生”

3万部古籍背后的“校书官”

北京大学数字人文中心基于自主研发的“识典古籍”数字化平台,于2024年发起了我是“校书官”——古籍大众智能整理计划,邀请高校师生、大众爱好者等,在智能技术的辅助下整理古籍,助力古籍“重生”。

在“校书官”计划发起后,平台收录、开放的古籍已突破3万部——古籍原本的影像触手可及,施以现代标点的文言文简洁易读,注疏典故让历史细节跃然纸上……志愿者中,不仅有相关文科专业的师生,还有互联网从业者、机场调度员等,他们凭借对典籍的热爱,突破专业所限,甘当一名“校书官”。

沉浸其中 8天整理19万字 互联网人钻进古籍里

我是“校书官”——古籍大众智能整理计划分阶段开展活动,包括初阶和进阶两类任务,定期通过“识典古籍”微信公众号招募志愿者。

利用假期时间,8天整理了19万字的古籍,是互联网大厂产品经理王唯一创下的纪录,至今仍经常被志愿者们提起。

“我平时喜欢看历史小说,了解到‘校书官’的活动后,想着下班后能换换脑子,而且还能做公益,就报名加入了志愿者团队。”王唯一说,有不少和她一样的“非科班”志愿者,大家在交流群中分享整理心得,互帮互助解答问题。

“我非常享受这种状态。就像玩游戏一样,原本只计划玩五六关,结果一抬头发现天色已经暗淡下去。”她说,过完假期,自己提交了19万字的整理结果,包含《安雅堂文集》《王荆文公诗》《全隋文 先唐文》等古籍。

赠人玫瑰,手有余香。让王唯一惊喜的是,整理古籍还给自己的本职工作带来了一些启发。

点开“识典古籍”平台中任意一部待整理的古籍,页面左侧是破碎泛黄的书页影像,右侧滚动着经AI(人工智能)识别的文字。这背后是文字识别、自动标点、实体识别三大核心技术。黑色文字代表算法识别的可信度比较高的字,一般不需要人工再去核对;蓝色则提示疑难字,得对照古籍影像上的字形再确认。志愿者需要做的,就是参考AI生成的结果校勘文字。

“为了方便整理,AI还会自动识别古籍页面的版式分布、图片信息,并对字迹进行分栏。然而,AI识别并非100%精准,每个人的整理习惯也不尽相同。当我连续整理了几页之后,发现AI自动省略了一些用不到的步骤,给整理工作提了效。”王唯一说,作为AI产品经理,自己在设计产品时经常会遇到“设定很完美,但用户体验感不佳”的问题。而“识典古籍”平台的贴心设定,能让人从背后看出复杂的产品设定逻辑,精准对接用户的需求。受此启发,王唯一开始以用户的视角去排查所设计产品的各个流程,发现了不少可优化的细节。

如今,王唯一仍习惯一下班就钻进古籍里。当整理的古籍通过审核在平台上架后,她都会收到一个提示——“您整理的书籍已经上架啦!您已被列为本书整理人”,“这时,我的名字会出现在页面上方,这是我最有成就感的时刻。”

解谜破案

为破解一个字 他们反复查阅资料推敲

志愿者中,不少人和王唯一一样虽然是“非科班”出身,但凭借着对历史古籍的一腔热爱,正在助力一部部古籍焕发新生。

李九术是一名机场调度员。白天,他在机场从事安全调度工作,下班后便走进书房,一头扎进古籍的海洋。闲暇时间,他还经常搜集各类古籍影像,打印成仿古经折本阅读。他说,自己对古籍的求知欲越来越浓了。

参与文字整理的基础工作后,李九术通过考核进入了志愿者进阶组——负责文字、标点的精校和审核等进阶任务,队友多是相关专业的高校师生。

有一次,他和队友遇到了一句诗——“小桃明淑?”,AI系统未能识别出问号处的字,前端的初阶志



古籍大众智能整理计划吸引众多志愿者成为“校书官”。

愿者也没能顺利解答这个问题。“古人写诗讲究对仗,‘小桃’是植物,后面也应匹配名词。”他和队友根据古籍影印图片上的字形看出了该字中含有“巢”字,根据上下文锁定了“巢”和“窠”字。翻阅《本草纲目》,他们印证了“窠”可表示植物丛生,再援引《和子由记园中草木》中“苍苔窠”的用法来佐证,最终才推定此处应为“窠”字。“这个过程就像解谜破案一样,很有意思,收获也很多。”李九术说。

参与我是“校书官”活动后,不少志愿者直呼:“长知识了!”志愿者卜瑞吉是北京一家科技公司的员工,前段时间认领了《阳泉县志》的整理工作,书中关于古时财税的内容,令他印象深刻。“以前我以为,银子的单位,两、钱、毫、厘已经到了极限,但读了古籍才发现,厘之后还有忽、丝、渺、蔑、微等七八个计量单位,粮食的最小单位在‘粒’之后也还有一个。”

这些意料之外的知识,也让她在忙碌的日常工作中,找到了浪漫的诗和远方。她曾整理过一本名叫《夏小正》的农书,其中有一句“八月丹鸟羞白鸟”,“我从中了解到,古人经常管小虫子叫‘鸟’,‘丹鸟’是萤火虫,‘白鸟’是蚊子。这句话的意思是,到了萤火虫吃蚊子的季节,就是八月了,用此句形容季节更迭很巧妙。”

精心守护

从纸端到指端 理科生用科技焕新古籍

选题、数据准备、分卷、审核……除了志愿者,“校书官”计划背后,还有北大数字人文中心工作人员的努力,他们负责团队管理、书目

管理、审订等工作。中心主任、北大信息管理系教授王军介绍,“识典古籍”平台由两部分组成,前端是阅读平台,后端是古籍整理平台。“就像是餐厅的前厅和后厨。中心工作人员和志愿者作为后厨,整理上线古籍资源,把数字化资源端上阅读平台,与全世界的读者共享饕餮盛宴。”

一群人,一条心,齐心协力让古籍重获新生。

作为古典文献专业的毕业生,北大数字人文中心工作人员李红雷在上大学时便做过古籍整理的工作,在接触到AI辅助校勘的方法后,他省去了一些繁琐的工作环节,集中关注更需要创造性思考的疑难问题。来到数字人文中心后,他将自己所学所思应用于“识典古籍”平台的建设中,并在“校书官”计划中负责选题、分卷、审核等多项工作。

“我们着重挑选了一些海外馆藏的珍贵善本,所选古籍内容都与大众生活紧密相关,期待大家能挖掘出古籍与现代生活相契合的内涵。”他说,团队还精心设计了古籍通识课程,邀请领域资深学者担任主讲嘉宾,广泛普及古典文献的基础专业知识和古籍智能整理实操方法。

郑振宇是一名比较活跃的志愿者,因为父母学的都是中文专业,所以他从小就接触古文。高考选择专业时,理科成绩更好的他选择了计算机专业,如今从事古籍保护工作,致力于用科技力量为古籍延寿。

“一开始接触的古籍保护工作十分繁杂,包括修复、扫描、图像处理等,手动操作效率很低。于是我开始自学编程,开发了一套工具,

把原本需要几周完成的工作压缩到几分钟。”他说,接触到“识典古籍”平台后,他被免费开放的AI识别功能所吸引,也加入到志愿者的行列中,如今已经在平台整理上架了11本古籍。他还把平台推荐给了同事,“识典古籍”平台已经成为大家工作中的“神器”。

目前,平台已累计3万多部古籍,包含《四库全书》《永乐大典》等大部头,面向海内外读者免费开放。王军说:“希望通过‘识典古籍’平台和我是‘校书官’活动,推动中华文明传承发展,向世界展示和传播我们的文化之美。”

成效

“众包模式”让转化提速

据统计,我国现存古籍20多万种,5000多万册,多数只完成了初步的影像扫描,只有少数真正实现了文本数字化。人工智能等新技术的应用和发展,让古籍能以另一种形态得到长期保存和传承。越来越聪明的AI,可以进行识别、排序、校对、结构整理等工作,在此基础上再进行人工整理、核对,就能大幅提高古籍整理出版的效率和精度。

目前“识典古籍”平台文字识别的准确率已经达到了96%,自动标点准确率达94%。北大数字人文中心表示,邀请大众参与古籍整理,这种“众包模式”胜在参与人数多,效率高,通过志愿者的广泛参与,能够较快地将纸质文献转化为可供检索的数字化文本资料。只有在文献信息的储备量足够庞大后,古籍内容的深度开发利用才有可能。

何蕊

·新作推介

《心向光明:邹韬奋传》: 以笔为刃投身文化抗战

邹韬奋为什么能够成为文化抗战的一面旗帜?韬奋精神如何在当下焕发时代价值?中国出版协会全民阅读工作委员会主任聂震宁所著《心向光明:邹韬奋传》给出了答案。

“在那个风雨飘摇、社会动乱的时代,有人是不战而屈,邹韬奋是战而不屈。韬奋先生不仅是爱祖国、敢斗争的人,也是尚高节、重感情的人。”聂震宁说,创作过程中他突出了传记的文学属性与人文温度,期待吸引青年读者阅读,激励更多人传承韬奋精神,在时代浪潮中坚守初心、勇担使命。

20世纪三四十年代,在国家危亡、民生凋敝的黑暗年代,邹韬奋以笔为刃,与生活书店的同仁们为了宣传进步思想、唤起民众觉醒、推动全民抗战,作出了艰苦卓绝的努力。《心向光明:邹韬奋传》记录他的生平事迹和光辉精神,以时间为经,以事件为纬,通过丰富的历史细节,真实再现邹韬奋创办生活书店,以笔为刃捍卫民族尊严,将民主政治和抗日救亡的民族精神广泛传播以唤醒民众的人生历程。中国作家

协会副主席、评论家李敬泽评价该书为“新时代传记文学的一个重要收获”。

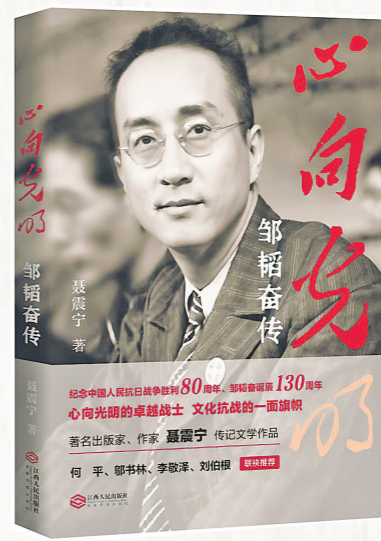
谈及邹韬奋,离不开“六刊一报一书店”——《生活》周刊、《大众生活》、《生活日报》、《生活星期刊》、《抗战》三日刊、《全民抗战》、《大众生活》香港版以及生活书店。这批以抗战为主题的刊物和生活书店出版物,发出团结御侮、抗战到底的最强音,在广大读者中唤起巨大的爱国热情,也为奋起反抗的军民鼓舞与呼,把一批批青年引上革命道路,以实际行动勇敢战斗在民族解放第一线。

书中回溯了邹韬奋文化抗战多个关键节点。比如,1926年邹韬奋接任《生活》周刊主编,以犀利之笔,力主正义舆论,抨击黑暗势力。九一八事变后,他主编的《生活》周刊以反内战和团结抗战御侮为根本目标,成为国内媒体抗日救国的旗帜。1932年7月成立生活书店,邹韬奋任总经理,团结一大批进步的作者,短短几年,使其在全国各地分支机构扩展到56家,先后出版数十种进步刊物和包括马克思主义译著在内的千余种图书。

1935年8月回国后,邹韬奋积极参加抗日救亡运动,在上海创办《大众生活》周刊,年底发行量就达20万份,打破当时中国杂志发行纪录。其间,邹韬奋担任上海各界救国会与全国各界救国联合会的领导工作。1936年11月,因积极宣传抗日,他同救国会的沈钧儒等其他领袖被国民党当局逮捕,成为“救国会七君子”之一。

“这部回答新时代新命题的传记,让我们得以跨越百年时光,触摸那颗滚烫的爱国之心,思考当下新闻实践,学习借鉴韬奋在战火中淬炼的新闻理念,在今日舆论场焕发新的生命力。”中华全国新闻工作者协会书记处书记殷陆君说。

许畅



《大历史小故事——让孩子站在世界看中国》: 在世界文明坐标系中 走进中国历史

把中国历史放在世界历史演进的大坐标中,通过横向比较,可以更好地理解中国历史发展,使我们既不自尊自大、故步自封,又不自卑自卑、缺乏自信。这正是北京大学出版社出版的《大历史小故事——让孩子站在世界看中国》传达出的编写理念。

该书由著名学者楼宇烈担任主编,精选160个小故事,巧妙串联起中华民族5000年发展历程,将中国历史精彩瞬间融入世界文明坐标中。该书不只讲述中国历史朝代更迭的传统故事,而是以中国历史为基点,以世界历史为参照,把同一时期不同国家和地区的历史展示出来,让小读者既能站在中国看世界,又能站在世界看中国。

譬如,在讲述唐代历史时,书中不仅描绘了唐代的繁荣景象和开放包容的文化氛围,还将其与同时期的阿拉伯帝国、拜占庭帝国等异域文明进行对比,展现了中国和这些国家的文化交流以及各自在世界历史中的地位;在讲述郑和下西洋的故事时,除了展现明朝的航海技术和海上力量的强大,还通过与同时期欧洲远洋探险活动的对比,让孩子理解不同文明在探索未

知世界时的异同。

该书的结构设计匠心独具。全书共分8册,每册开头栏目是“本册大事记”,列出该册对应时期内中国历史上的重要节点;“放眼看一看”栏目,简要梳理这一时期世界其他地区发生的重大事件,特别强调了中国与世界各地区的交往和相互影响;“读前热身”栏目,是进入每个历史故事前的阅读引导;紧接着的“开讲啦”栏目,则是故事的核心内容。

最后一个栏目“延伸思考”,是该书的特色板块,设计了200个延伸思考问题,既有对历史现象背后原因的思考,也有对历史故事细节的追问,还有对历史事件发生背景、原因的延伸思考,更有对未来的展望。这些问题需要孩子通过延伸阅读和研究性学习寻找答案,而且很多问题也是开放性的,并没有标准答案或唯一答案。显然,这样做的目的,是试图帮助孩子理解历史的丰富性和复杂性,培养孩子的综合思维能力和批判性思考能力。

值得一提的是,作为写给孩子的读物,这套书在视觉呈现方面也下了功夫。该书精心挑选700幅历史图片,如同一部视觉史书,让孩子在文字之外,能够通过直观的图像感受历史的真实与厚重。这些图片涵盖文物、遗址、人物肖像、地图等多种类型,提供了多元化的学习素材。比如,通过呈现世界各地的古迹,让孩子意识到人类文明的多样性和相互关联性,增加了身临其境的感受。

除历史图片,《大历史小故事——让孩子站在世界看中国》还有800幅彩绘漫画贯穿全书各部分,进一步增强了这套书的趣味性和亲和力。这些漫画将历史故事以更生动形象的方式呈现出来,有的使用AI绘制,给孩子留下丰富的想象空间。

立足中华文明,培养孩子胸怀世界的大格局和大历史观。这是《大历史小故事——让孩子站在世界看中国》的初衷和愿景。

周雁翎



“校书官”计划为志愿者准备的纪念品文创。

