

“超大规模古籍数据库”时代来临

2013-08-29 09:02:00

来源: 中国社会科学报—中国社会科学网

中国社会科学网讯(记者 孙妙凝) “古籍数字化是人类文明载体的一次大转移——从纸媒体向电子媒体的转移。在这一过程中,如抗拒时代潮流,就必将被历史淘汰。”首都师范大学电子文献研究所所长尹小林说。

8月16—18日,由首都师范大学电子文献研究所、首都师范大学中国诗歌研究中心、亚洲大学资讯学院共同主办的“第四届中国古籍数字化国际学术研讨会暨第六届文学与资讯技术国际研讨会”在京举行。与会学者就“古籍数字化在文史研究领域的应用”、“专业古籍数据库的建设”、“古籍数字化技术与发展方向”、“古籍数字化政策与法规”等议题进行了研讨。

超大规模古籍数据库初具雏形

我国宏富浩繁的文化典籍,是中华文明和民族精神的重要载体。近年来,古籍数据库迅速发展,大约3万种古籍品种完成了数字加工,数十亿字的古籍被数字化,出现了以《四库全书》、《国学宝典》、《中国基本古籍库》等较为成熟的大型古籍数据库。

尹小林表示,古籍数字化是现代学术研究的基础。利用计算机及网络技术进行深入的古籍整理工作,在当今数字化时代势在必行,它必将大幅度地提高研究中国古代文化的效率,将学者的时间和精力从艰苦而繁琐的爬梳、翻检工作中解放出来,推动人文学术研究的发展。古籍数据对文化传播、历史研究,甚至是思维方式都有重大的、革命性的影响。

北京大学古文献研究中心主任廖可斌也认为,古籍数字化是非常有意义的事业,对整个古典文献研究、古代文化研究都将产生巨大的影响,对学术研究方法的影响亦相当深远。

谈到古籍数据库的发展趋势,尹小林认为,十亿字以上的“超大规模古籍数据库”建设将是未来的发展方向。古籍数据库可以分为几个级别:一百万字到一千万字的小规模数据库,一千万字到一亿字的中等规模数据库,一亿字到十亿字的大规模数据库,如《四库全书》。近几年,自动比对、自动标点、自动排版“三大核心技术”的研发成功,是古籍数字化取得的重大突破,标志古籍整理从传统手段迈向现代化。在“大数据时代”,数据资源和技术都取得了较大进展,十亿字以上的超大规模数据库已经初具雏形。

国家古籍整理出版规划小组办公室原主任许逸民则表示,古籍数据库可以起到索引的作用,但它更应该是一个史料库,读者可从中直接获取所需的资料。他认为,应当提倡古籍资源的整合,古籍数据库的资料越多越好,但是对于基本古籍还是要精耕细作。就目前来看,很多基本古籍的整理和数字化工作并不充分,如《十三经注疏》等,这需要引起重视。此外,要通过高科技手段做到使用电子文本比纸质文本更方便,比如加入正文检索、注文检索、专门词汇的检索等功能。总之,超大规模数据库建设与基本古籍的精耕细作应齐头并进。

近几十年来，我国的古籍整理事业在不断向前推进，积累了一定的成功经验，也确立了一些规范。古籍数字化和古籍整理成果如何接轨？这其中也衍生出一些新问题。

“比如，古籍数据库即使取得了整理本的授权，是否需要加入校勘记呢？如果版本已经改动，却不附校勘记，就不符合古籍整理的规范。选用什么底本也是需要考虑的问题。另外，古籍数字化中的繁简字转换也很值得探讨。事实上，将繁体字转换为简体字是对古籍的一种改造。有几种情况是不能简化的，一些书名如《四书贖言》中的“贖”，以及一些专业术语如“校讎”的“讎”，都不能随意转化。”许逸民说。

许逸民认为，总的来说，要明确古籍数字化的性质，它是古籍整理众多方式中的一种，需要符合古籍整理的规范。当然，它也有特殊性，对计算机技术的要求很高。我们需要依赖技术的进步来完善古籍数字化的工作，但是说到底技术只是古籍数字化的一种支撑，而不是对古籍数字化的领导。此外，古籍数字化整理是把双刃剑，对于读者来说，它提供了极大便利，但对于研究者而言，反而可能受到限制。电子化的典籍已经被加工过，研究者还是应当回到原典，提出自己的看法。

北京大学古文献研究中心主任廖可斌表示，古籍数字化在最开始只是文献的扫描录入，如今又不断有检索系统出现。其中一个很重要的问题是，如何将古籍整理的一些概念，如版本、校勘应用进来。没有注明版本的古籍文献是无法为研究者所用的。另外，古籍数据化产生的一个弊病是：一些计算机技术非常熟练的年轻人变得过于依赖电子文献，而疏于阅读。“应当在问题的引导下检索资料，而问题从原始文献中来。在古籍数据化的浪潮下，研究者应趋利避害。”廖可斌说。

制定相关法律法规保障古籍数字化发展

相对于古籍数字化的快速发展，有关古籍数字化出版物的法律法规还相对滞后，由此也不断凸显出一些新问题。学者认为，应当在立法上保障古籍数字化的发展。

安徽省社会科学院文学研究所教授陈友冰表示，古籍数字化工程浩大，其读者与生产者之间、电子产品与纸质产品、古籍与数字化古籍之间涉及问题很多，而且会不断产生新问题，如古籍的“文献保真”和“文献整理”的界定和处理原则；典籍已有的纸质产品的电子产品的著作权界定；如何体现数字化过程中古代典籍在知识产权保护中的特殊性等。这些新问题无成规可循，必须通过制定相关法规加以解决。

古籍数字化极大地方便了读者阅览古代文献，但在古籍资源的利用上与传统出版社产生了分歧。近年来，我国频频发生涉及“古文点校”成果电子化的著作权纠纷。华东政法大学教授王迁认为，以“复原”古文原貌为目的的“古文点校”虽然可能包含独创的智力成果，但基于事实和对事实的思想观点不受保护等著作权法的基本原理，其并不能作为作品受到著作权法的保护。

但这绝不意味着这类智力成果不值得保护。王迁表示，我国的古代文献资料极为丰富，而绝大多数人只能通过点校版本阅读，因此古代文献的点校者和出版者应当受到保护。“可以借鉴国外著作权立法，增加一项‘对特定版本的邻接权’，既不违反著作权法基本原理，又能使点校者和出版者受到合理期限的保护，是很好的立法对策。”王迁认为。

责任编辑：田华

文档附件：

隐藏评论

用户昵称： (您填写的昵称将出现在评论列表中) 匿名

请遵纪守法并注意语言文明。发言最多为2000字符（每个汉字相当于两个字符）

3578

发表

中国社会科学院电话：010-85195999 中国社会科学网电话：010-84177865；84177869 Email: skw01@cass.org.cn
投稿邮箱：skw01@cass.org.cn 网友之声信箱：skw02@cass.org.cn 地址：中国北京建国门内大街5号
版权所有：中国社会科学院 版权声明 京ICP备05072735号