

数字化时代对史学研究模式的影响

贾文龙^①

【内容提要】 随着计算机智能技术的发展,传统文献学面临着数千年来最大的变革,人类知识传承出现新载体,文献书写方式、收藏方式、阅读方式、出版方式都出现重大变革。这一时代变革对史学研究也产生深刻影响:改变了史料搜集的方式,改变了校勘与考据的概念,改变了索引、标点的方式,引发了研究思维的转变,改变了学术前沿的概念。历史学科研究生应该密切注意这种变化对学术研究的影响,努力提高个体信息素养,以使自己的研究跟上时代的进步。

【关键词】 历史学 文献学 数字化 史学理论

前言、古典文献学面临的“千年之变”

马克思说:生产工具促进生产力。马克思在经典著作中论述人类历史发展阶段划分时指出:历史发展阶段是以人类采用什么生产工具及相应的生产方法来划分的,所以分成石器时代、铁器时代等。推动历史发展,生产工具的变革是最积极、最首要的因素。人类的生产工具由手动发展到具备动力,马克思认为这是一次人类生产工具的大革命。现在计算机技术的发展,使人类的生产工具具有一定智能性,这是人类生产工具新的又一具有划时代意义的发展。

进入 20 世纪以来,计算机的软硬件技术对古典文献学产生了越来越深入的

^① 河北大学宋史研究中心暨历史学院,河北 保定,071002。

影响。一是出现了人类知识传承新的载体,在人类进入文明社会后,甲骨、金石、简帛、纸质文献构成了文献载体,而现在电子文献也成为人类知识的新载体,并具有容量、复制、保存诸方面的科技优势。二是人类书写方式出现新的变革,中国传统的书写工具有刀锥和毛笔两类,并与印刷术相结合,产生了数量庞大的古籍文献,现中国古籍存世约有 15 万种。而计算机时代,键盘录入以及文字识别成为新的书写方式,既带来论著写作的便捷化,又使海量文字录入成为可能。三是实现人类收藏与检索文献方式的变革。人类曾试图把所有知识放在一幢房子里,因而出现了图书馆,并因检索的需要而出现目录学的专门知识。网路技术的出现和搜索引擎的发展,使人类的精神文化宝库可以不经集中收藏就可以联结成巨型文献库,即使是珍稀古籍版本也可飞入寻常百姓家,文献材料对每个研究人员而言在数量上都是平等的,并出现去物质化的趋势。四是实现人类阅读文献方式的变革。所有印刷出版物均是“线性”的,只能一页一页按顺序阅读,而网页组织则是“网状”的,读者可以直接阅读自己最需要的内容,人类从有序阅读变为无序阅读。2009 年王余光先生《我们不得不面临的时代》一文提出在中国历史上,第一次阅读转型是从手抄书到雕版印刷;第二次是机械印刷和新式教育的变革;而第三次就是我们今天所处的时代,电视、电脑和手机等电子媒介屏幕阅读成为主流。五是实现图书出版方式的变革,人类图书出版长期处于铅与火的时代,自近代进入光与电的时代,而现在自动排版成为可能,不需要出版社而由作者直接出版图书的“去中介化”趋势也已出现,网络首发也成为学术成果发表的认可方式。

古籍文献是史学研究的基础,文献学面临的“千年之变”及计算机技术的智能化的深入发展,自然也会对史学研究模式产生深入影响。

一、文献普及化

对治中国古史研究而言,常以占用尽可能多的历史文献作为得出结论的前提,占用绝大多数史料就可以下基本结论,而没有看到重要史料,只能就事论事、就书论书,因而这样的研究也难以产生影响。

所以对古时学者而言,是否看到史籍决定了同一专题学者群的大小。清代徐松等人利用修《四库全书》的机会,得以观览《永乐大典》,撰写了《宋会要辑稿》等辑佚作品,完成《唐两京城坊考》《登科记考》等考据性著作,成为有清一代研究相关专题的最著名学者。即使《宋会要辑稿》成书后,大多治宋史学者也无缘阅读,一直到民国影印以后,才得以广泛传播。

另如《四库全书》是清代乾隆年间官修的规模庞大的百科丛书,它彙集了从

先秦到清代前期的历代主要典籍,共收书 3400 余种,79000 多卷,分装 36000 多册,总字数约 9 亿。《四库全书》作为有清修书集大成之作,清政府在各地建阁而收藏,但普通人根本无缘翻读。民国期间,时任北洋政府教育总长的傅增湘到北平图书馆看《四库全书》,都是在两个带枪士兵监视之下才得以成行。陈垣少时读《四库全书总目提要》,曾自叹今生可能无望读到《四库全书》。1915 年,文渊阁《四库全书》搬运至北京京师图书馆。陈垣每天租一架驴车到图书馆阅读《四库全书》,十年每日往返,才成就其学术的渊博高深。上世纪 90 年代,漆侠先生购买台湾文渊阁四库全书,所需 15 万经费需省长特批,当时河北省内只有 3 套四库全书。

北京书同文数字化技术有限公司于 1997~2000 年与香港迪志文化出版有限公司、上海人民出版社、微软北京研发中心、北大方正、清华大学计算机系、国家图书馆、香港中文大学等单位合作,开发完成了《文渊阁四库全书》原文及全文检索版,只占 7 个多 G 的硬盘空间,在保存了原书文献真实性的基础上,还提供完备的阅读及检索功能。北京书同文数字化技术有限公司又于 2001 年开发研制成了《四部丛刊》全文检索电子版。

四川大学古籍研究所较早完成“电子版《宋会要》”项目。2008 年与上海人民出版社、北京大学、河北大学、河南大学、武汉大学、上海师范大学也联合制作了《宋会要辑稿》研发项目。2013 年上海人民出版社出版《全宋文》数据库。现在一台计算机甚至一个智能 U 盘,就是可以随身携带的图书馆,跑图书馆查阅资料已经不是学者常做的事情了。

此外一些古籍珍本,也因大型综合性图书的出版而从馆内秘藏变为面向世人。如四川大学古籍研究所编纂《宋集珍本丛刊》就是从海内外各大图书馆收集而得的 960 种宋代珍本文献中精选而成。1996 年上海图书馆启动《宋元善本全文数据库》,并分期实现上网服务。1998 年国家图书馆启动“中国数字图书馆工程”,其中数字方志资源库、石刻拓片资源库、甲骨文献资源库、馆藏各类文献书目数据库、永乐大典资源库等六个子项目均是古籍项目。此外中华古籍善本国际联合书目系统、台湾国家图书馆善本丛刊影像先导系统、“学苑汲古——高校古文献资源库”、哈佛大学哈佛燕京图书馆藏善本特藏资源库、东京大学东洋文化研究所汉籍全文影像数据库、汉籍数字图书馆、长春图书馆“馆藏国家珍贵古籍数据库”等网站,也都可以查阅古籍。

在古籍数字化的时代,可以说古籍珍本已经飞入寻常百姓家,对于史学各个领域而言,其刚刚进入相关领域的研究者与已有多多年积累的研究者,其在文献占有的数量上几乎是相等的,再也没有文献占有的客观制约了。

二、史料综合化

古代典籍浩如烟海,学术上的突破,往往依赖于新材料的发现。从某种意义上讲,史学就是史料学的观点有相当的合理性,传统的史学研究在某种程度上,就是发掘史料功夫的较量。以傅斯年为代表的史料学派就是这样认为的:“只要把材料整理好,则事实自然显明了。一分材料出一分货,十分材料出十分货,没有材料便不出货”,他们宣称不读书只找材料,“上穷碧落下黄泉,动手动脚找东西”,对后来学者也有相当大的影响。

但在古籍数字化的时代,各种各样的古籍数据库的特点是通过检索来搜集史料,而且速度之快,几亿字古籍检索可以在分钟单位内完成,让每位初次使用数据库的学者都曾叹为观止。随着各类古籍数据库建设的精细化发展,其查全率和查准率都有相当提高,越来越可以放心使用。有学者举例:

有位学者研究中国古代的“矫制”,用传统方法查找资料,辛辛苦苦干了一个多月。文章写成后,请一位懂电脑的同行提意见,那位先生上网检索“矫制”,几分钟就检索完了。一对照,手工翻书比电脑查找还少找了一条资料。^①

现在面世古籍数据库大多以古籍为数字化对象,而不是以现在通行学科类别为区分标准,这就使得数字化古籍库模糊学科分类,事实上消解了历史、文学、哲学等学科间的材料隔膜。专治宋史的王曾瑜先生在撰写了科技思想史方面的文章后感叹道:“目前古籍电脑软件尽管还是处于原始发展阶段,有不少的缺点,却已显示了巨大的优越性,可以数倍,甚至数百倍、成千倍地提高查史料的效率。我常说,自己治史,过去是手工生产阶段,如今算是进入了机器生产阶段。我通过一些治思想史研究生的答辩,发现治思想史者常泛用宇宙观一词,才想到应为此写一篇文章。《从古“宇宙”词义谈古人的宇宙观》一文全靠古籍电脑软件检索,但还是与核查书本相结合。完工后,不免感慨说,如果未使用古籍电脑软件之前,凭自己的学力,这个题目确是连想都不敢想。就是在前辈学者中,只怕也只有张政烺先生方可胜任。如《明儒学案》一书,即使要浏览一遍,只怕也须用两三天,我使用古籍电脑软件检索,大约只花了三十分钟。”^②

在数字化时代下,已经有学者将电子文献列为文献学的重要研究内容,如张三夕先生主编的《中国古典文献学》中有“电子文献的检索和利用”的专题论述。^③杨琳先生《古典文献及其利用》中认为:“‘数字文献(Digital document)’是以二

① 王文涛:《怎样利用数字资料研究中国古代史》,《历史教学(下半月刊)》2003年第5期。

② 王曾瑜:《治辽宋金史杂谈》,《纤微编》,河北大学出版社,2011年版,第24页。

③ 张三夕:《中国古典文献学》,华中师范大学出版社,2007年第2版。

进制数字代码形式记录于磁带、磁盘、光盘等载体,依赖计算机系统存取并可在通信网络上传输的文本、图像、音频、视频等文献。数字文献也叫做电子文献。”^①

自 20 世纪以来,人类迎来知识爆炸的时代,人类知识总量急剧增长,至 21 世纪初,人类至少已经写作了 3200 万册书、7.5 亿篇文章,这种学术深化发展必然带来各种学科的细致分类,有学者认为 5000 年前,世界上只有三门学科:语言、图腾、技艺,2000 年前,已经增加到 5000 门,本世纪中叶应该有 20000 门学科。贝弗里奇讲:“‘不管一个人对一门学科(艺术、经济、宗教、政治、科学等)掌握多少,人们总不能看到这样一个令人惊讶的事实,即这些学科集中在一起可以形成某种别的东西。事实是,整体大于原来个体加在一起的总和,其性质不能从部分中预测到,这才是事实的关键所在。’独创性常常在于发现两个或两个以上研究对象或设想之间的联系或相似之处,而原来认为这些对象或设想彼此没有关系。”这些数以万计增加的新学科,决大多数都是交叉学科,而新产生的人文社会科学学科至少也会数以千计,都会产生大量的学术增长点。而综合性的数据库一定会为这些新的学术热点提供有力的技术支持。

三、数据个性化

在前数字化时代,中国传统史学的治学模式都是近似的:先是确定题目,然后阅读相关史籍,边作史料卡片,边进行理论思考。随着研究者对相关文献的系统翻阅,所选题目的历史发展渐渐清晰,进而找出问题的关键所在,结合时代背景提出作者的个人结论。因而史料搜集是中国传统史学研究的基本功,读书并做卡片是几乎所有导师对学生的基本要求。因为史料史料的翻阅与搜集是相当枯燥的事情,因而提倡“板凳要坐十年冷”的治学精神。可以说,史料卡片的多年积累,几乎是中国传统史学的不二法门。众多著名学者的成名之作都是建立在大量史料卡片记录的基础之上,如钱锺书先生的《谈艺录》《管锥编》就是在几大麻袋卡片的基础上完成的。文革期间,因为所搜集的众多史料卡片的丢失,迫使众多学者中断了学术研究,漆侠先生曾说过文革期间最大的损失之一,就是为写作《宋代经济史》而准备的几麻袋卡片被抄走。文革结束后,漆侠先生马上到天津图书馆等地重新查阅史料,还说“这样也好,可以让我重读一遍史料”,又经多年准备,才最终完成《宋代经济史》。

在数字化时代,为史料卡片的制作与搜集有带来相当大的便利性。现在流

^① 杨琳:《古典文献及其利用》,北京大学出版社,2010 增订本。

行的数据库,几乎都有文字复制功能,这就避免了大量手工抄写劳动,而进行“Ctrl+C”、“Ctrl+V”的操作就可以了。1997年~2000年香港迪志文化出版有限公司与北京书同文数字化技术有限公司等单位合作,开发的《文渊阁四库全书》全文检索版,就提供了文字复制功能,同时可以复制书名和卷数的史料出处,但要注意的是,在复制后粘贴时,注文不在原句之下,而是在复制史料的末尾之后。另一大规模数据库“中国基本古籍库”6.0版也提供出处复制。

此外,还有相当多的数据库提供卡片输出功能,既有检索结果,又有文献出处,使史料卡片的制作进入机器化时代。台湾中央研究院的《汉籍电子文献》、陈郁夫先生的“寒泉”检索系统都有分段显示并输出的功能。2010年北京国学时代文化传播公司与河北大学宋史研究中心合作研发的《宋辽夏金元史电子馆》也实现了卡片输出功能,检索宋辽夏金元时期近4亿字文献后,可以几秒钟内将结果输出为电子卡片。陕西师范大学袁林先生主持开发的“汉籍全文检索系统”,共近10亿字,也可以一次性输出为一个文本文件。

可以说,在数字化时代,传统的史料搜集工作已经向数据收集转化了。这一时代特点,首先对人物与名物等带有明显检索关键词的选题带来便利。进行人物研究,检索历史人物的姓名、字号、别号、事迹、著述,花费旬日检索得来的史料基本上就可以扶持开始着手研究工作了。笔者要研究宋代词人笔下的“镜”意象,使用北京国学时代文化传播公司研制的《全宋词》数据库,检索得到镜意象共224处,65000字的卡片几秒钟内生成。随着研究的深入,所涉及的关键词可以不断增加调整,从而多单一检索变为多元检索,从定向检索变为关联检索,由静态检索变为动态检索,从而在更大范围内找寻出相关史料,进行分类整理之后,就可以建立起属于研究者相关选题的个性数据库。

戴伟华先生《地域文化与唐代诗歌》为例来说明。唐代文学研究中地域文化视角得到了应有的重视,归纳起来大致有六个层面:(1)以本贯、占籍为切入点;(2)以隶属阶层为切入点;(3)以南北划分为切入点;(4)以文人的移动路线为切入点;(5)以诗人群和流派为切入点;(6)以文化景观为切入点,等等。那么,面对这样丰硕的成果,要从诗歌自身来讨论地域文化问题有相当的困难。如果以诗歌创作为本位切入地域文化与唐代诗歌研究,可能会更贴近唐诗的实际。正是基于这样的考虑,戴先生从唐诗创作地点切入,花费大量时间创建两个必备的数据库,一是《唐文人籍贯数据库》,一是《唐诗创作地点考数据库》。在此基础上,戴先生将过去主要以诗人籍贯为主的分析,转换为以诗歌创作地

点为主的地域文化与文学的研究。^①

此外,在论著参考方面,也可以建立个人数据库。如超星数字图书馆中,用户可以编辑专题的虚拟图书馆。另外“人大复印报刊资料”等中文全文期刊数据库,也按学科领域或专题进行编选而形成的系列专题资料。数字化时代,不论史料收集,还是论著参考,都可以以数据形式,形成个性化、专题化的数据库。

四、解读碎片化

在数字化时代,建立属于研究者的专题数据库成为可能,史料搜集有向数据收集转化的趋势,这就使得史学研究更多从史料汇编开始进行,而史料阅读从有序阅读更多转向为无序阅读进行。

传统史学研究中的史料卡片制做,是建立在有顺序读书的基础之上。中国古代史籍,不论是纪传体、编年体,还是总集、别集,其内容编排都有内在的逻辑顺序,其作者及成书都有固定的文献背景,而边读书边做卡片就是在这一逻辑框架内展开的,因而几乎每张史料卡片中都要注明时间、地点、人物信息,以便与其他卡片建立联系,最后用大量卡片完成所选题目的全貌拼图。

而在数字化时代,选题之后通过检索生成的数据卡片,其所依据的所查询数据库的子库编排顺序,其最大缺点是史料脱离了当时的历史背景,人物常孤立于时代群体之外,事件常不明其背后利害关系,制度常难解其演变过程,地点常断裂于自然区域,总而言之,检索史料脱离了其所在的活生生时代背景。

因此,数字化时代检索生成史料汇编后,语境回追成为重要环节,需要回到原文中阅读而理解上下文语境,让每条史料回归当时的史源出处和历史背景。如朱瑞熙等先生合著《辽宋西夏金社会生活史》时,引清代学者钱大昕《十驾斋养新录》之说,认为妇人自称“奴”始于宋代。王曾瑜先生认为此说无十分把握,使用《四库全书》电脑软件检索,不料竟有两三万个“奴”字,王先生凭藉以前的阅读和使用经验,单查《太平广记》和《全唐诗》两书,果然在前一书中找到了唐代女子自称“奴”者。^②

如果说,少量史料重新回到文献语境中再次阅读是比较容易,而让海量检索信息回归历史(如包伟民先生所言对历史完整场景的理解)则有相当的难度。比如《四库全书》电脑软件,“朱子”、“晦庵”、“朱熹”的检索结果就有 25817 卷、78864 个匹配。至于“孔子”,更是高达 23757 卷、111641 个匹配。检索“镜”字,

① 转引自吴夏平:《古籍数字化与文献利用》,尹小林主编《第一届中国古籍数字化国际学术会议文集》,北京:五洲传播出版社,2009 年版。

② 王曾瑜:《治辽宋金史杂谈》,《纤微编》,河北大学出版社,2011 年版,第 24~25 页。

结果有 19210 卷、50335 个匹配。复制、阅读、理解这些数目庞大的检索结果,是相当花费功夫的。更重要的是,还可能出现研究者淹没于海量信息中,而难以完成史料碎片的历史拼图。

汪圣铎先生对电脑检索和史料汇编的关系有如下认识:1 避免重复;2 考察文献成文年代和所反映情况和关系上的考证;3 是文字校点核校方面的努力,因而自信它的功能决不是电脑检索所能取代的。^①

因为利用数据库检索的史料是脱离历史背景的,因而常“只见草茎,不见根须”,对学术研究的进一步开展当然有深刻的不利影响。李华瑞先生评价宋史界博士论文时说出这种弊端:“现代信息数据技术的高度发展,为年轻学者,甚至为非历史专业初入宋史方向的博士生,在很短的时间内掌握大量论文所需资料成为可能,那种以掌握史料多寡作为衡量史家能力高低的时代已一去不复返,这大致也是现今博士学位论文在短短三年间字数动辄逾数十万言的主要原因。但是由于相当多的青年学生过分依赖数据库,而缺乏对基本史料的阅读理解,缺乏对宋代历史发展的总体把握,论文叙述很难把握问题本质与现象之间的联系,丰富的历史内容失去了多彩的颜色,变得单调而孤立。只见树木,不见森林的现象颇为严重。”^②

在数字化时代,我们既要利用检索手段获得新材料,而又要避免臆想检索词和查询结果脱离历史背景的弊端,因而研究生要特别加强这方面的训练。2012 年上海华东师范大学“e—考据与文史研究”研修班,其意图就是代表了这样的努力:“以清代名臣福康安之史事与传说作为主要的研究个案,选择此一目前仍混沌不明的有趣题材为案例,透过讲演与实例演练,尝试引导学员如何切入一个陌生的课题,发掘有意义的问题,规划可行的研究路径,搜索有帮助的证据,处理不兼容的材料,并探索该如何逼近历史的原貌,供学员能在短时间内切入庞杂史料,从中拓展、组合、拼凑关键的数据,进而张开研究的视野与深度。”

五、考证工具化

对史料进行考证,是史学研究的基本功。在数字化时代,考证出现工具化的趋势。这是因为,古籍数据库全库检索本身就能成为考证重要手段,如袁林先生举例:

中学语文课文有《乐羊子妻》一篇,需考证:乐羊子是否战国时乐

^① 汪圣铎:《两宋货币史料汇编·编后语》,中华书局,2004 年版。

^② 李华瑞:《近三十年来国内宋史研究方向博士学位论文选题取向分析与思考》历史教学(下半月刊),2009 年第 6 期。

羊?此“子”是否类如“孔子”为男子尊称?在我们所做“汉籍全文检索系统”软件10亿字古籍中,分别检索含有“乐羊子”、含“乐羊”但不包含“乐羊子”的全部资料,并予以归类,发现两组材料特征完全不同,仅《封神演义》将此二者混同,由此可得结论:“乐羊子”非“乐羊”,此“子”为姓名一部分,非男子尊称。

更重要的是,因为古今各类工具书的大量数字化,如《汉语大辞典》《汉典》的网络版,《瀚堂典藏》数据库(原称《龙语翰堂》)中对中国传统小学工具及古代类书的数字化,以及百度知道、新浪爱问、雅虎知识堂、天涯问答等网络工具书的发展,许多考证问题也成为省时省力的技术性操作。比如研究生初读《续资治通鉴长编》卷一的史料,利用网络工具书可以做简单考证,以利于理解古文:

丁巳,命宗正少卿郭玘祀周庙及嵩、庆二陵,因诏有司以时朝拜,著于令。(玘,未见。)

【有司:指官吏。古代设官分职,各有专司,故称有司。例:《史记·廉颇蔺相如列传》:召有司案图。诸葛亮《出师表》:宜付有司论其刑赏。王安石《答司马谏议书》:授之于有司。】

【著 zhù ◎ 显明,显出:~名。~称。显~。昭~。卓~。◎ 写文章,写书:~述。编~。~书立说。◎ 写作出来的书或文章:名~。巨~。遗~。译~。~作。◎ 古同“贮”,居积。】

目前,已经出现古籍校勘智能化的趋势,如书同文公司开发了“校得快、校得准、校得精”软件,另有公司开发了黑马校对系统。北京国学时代文化传播公司开发的古籍自动比对技术,更有助于古籍考证。以两个数字化版本的比对重新标点《文渊阁本二十四史》,现将《宋史》第一九九卷《刑法志》中,具有明显不同之处择要举例如下:

GX:禁于已然之谓敕禁于未然之谓令设于此以待彼之谓格使彼效之之

BZ:==≈★=====≈★=====≈=====

DB:禁於未然之谓敕禁於已然之谓令设於此以待彼之谓格使彼效之之

GX:刘一止言法令具在吏犹得以为奸今一切用其所省记欺蔽何所不至

BZ:==★=====★=====

DB:刘一正言法令具在吏犹得以为奸今一切用其所省记欺蔽何所不至

GX:日小事五日三年诏御史台鞠徒以上罪狱具令尚书丞郎两省给舍以

BZ:=====★=====

DB:日小事五日三年诏御史台鞠徒以上罪狱具令尚书丞郎两省给舍以

GX:日乃候进止裁处轻重必当其罪咸平四年从黄州守王禹偁之请诸路

BZ:=====★=====

DB:日乃候进止裁处轻重必当其罪咸平元年从黄州守王禹偁之请诸路

GX:改法诏以强盗计赃应绞者并增一倍赃满不伤人及虽伤人而情轻者

BZ:=====★=====

DB:改法诏以强盗计赃应绞者并减一倍赃满不伤人及虽伤人而情轻者

GX:捕系罢其职奉元丰二年成都府利路铃辖言往时川峡绢匹为钱二千

BZ:=====★=====

DB:捕系罢其职奉元丰二年成都府利路铃辖言往时川陕绢匹为钱二千

考证以上不同,唐律中勅为刑法,追溯以往,当为“禁于已然”;《宋史》中有《刘一止传》,刘一正为误;《续资治通鉴长编》卷四三咸平元年十二月甲寅载:“(刑部郎中、知制诰王禹偁)落职知黄州”,因而当为咸平元年。

此外,人民出版社的“金典引文比对系统”,也可以校证引用马恩列斯毛语录是否正确。

随着古籍数字化的发展,甚至可能实现电子标点古籍。现在国学公司利用多年积累而成的句型库、语法库、参考文本库,已经实施电脑重新点校二十四史。李铎先生提出的让计算机自主学习而标点《四库全书》的设想,在将来是有实现可能的。

在数字化时代,台湾新竹清华大学黄一农教授所著《两头蛇》一书充分利用网络文献和数字化古籍来考察明末清初的第一代天主教徒,所利用资料多达1099种,进而提出“e考据”的概念,认为“我们有机会在很短时间内就掌握前人未曾寓目的材料,并填补探索历史细节时的许多隙缝,或透过逻辑推理的布局,迅速论断先前待考的疑惑或者矛盾。事实上,一个有机会孕育‘e—考据学派’的时代或已出现。”

六、观点理论化

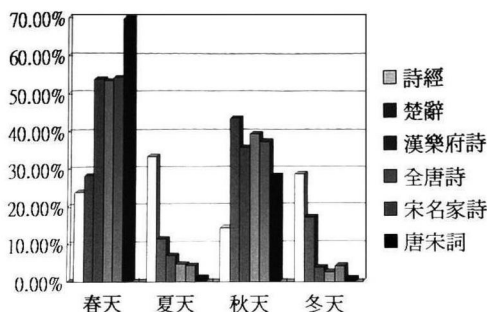
在数字化时代,数据库检索将原来花费精力的史料搜集工作变为技术性操作,对于专业研究者,还是业余爱好者,这种检索都很容易完成。这就使得一些题目的学术价值相对降低了,如传统的“寻章摘句”的乾嘉学派考据式的研究就在此中之列。如一则民间故事所言:乾隆南游经顺德府与高僧对话:“常念佛经,上面有几句阿弥陀佛?”老僧反问:“熟读四书,上面有几个子曰?”这样的问题在

今天已经难以称为学术。只有由于这种考据式的研究在思维方面的省力性特点,在当前研究中的实用性思想的主导下,仍然有大量这方面的选题。

在史料搜集便捷化的趋势下,研究者的精力当然要投入到更深层次的理论层面。刘家和在《〈崔述与中国学术史研究〉序》中说道:“记得从前有一位学术前辈说过:‘上穷碧落下黄泉,动乎动脚找材料。’这已很不容易,而现在是要‘上穷碧落下黄泉,以求有所新发现’,这就更难上加难了。”^①

在数字化时代,史学研究应更追求理论化思维。本文以两个方面为例:一是数据库的检索功能,其查询结果实质也实现了统计功能,可以促进量化式研究的新进展,从而为计量史学提供新的应用。二是在数据库中检索,其关键词很容易建立相互间联系,关系网络式研究的理论也会得到新发展。

在前数字化时代,很多统计只能凭籍感性认识,如明人孔尚任讲中国古代诗歌与季节的关系:“秋之气高洁爽朗,合于诗人之致,于秋得其五,于春得其三,于夏冬仅得其一二。”这种说法完全建立在作者感性认识之上,而缺少数据统计支持。现在有学者如罗凤珠先生利用数据库对“诗词中的四季与景物”进行了精确统计:^②



根据罗先生统计,《唐宋词》中以春天为主题的约占 70%,以秋天为主题的约占 28%,而以夏、冬为主题的各约占 1%。而《全唐诗》以春天为主题的约占 53%,以秋天为主题的约占 39%,而以夏为主题的约占 5%,而以冬为主题的约占 3%。《宋名家诗》以春天为主题的约占 54%,以秋天为主题的约占 37%,而以夏、冬为主题的各约占 4.5%。证明中国古代诗人题咏,春季的体裁实多于秋季,而夏季和冬季的体裁比例更少,这种统计数据是更有说服力的。

但是这种数字统计之后,还要进行理论分析,否则只能描述现象,而不能给出历史原因。笔者研究宋代词人笔下的“饒”意象,检索得知,吴文英使用 58 次,

① 邵东方:《崔述与中国学术史研究》,人民出版社,1998 年版。

② 2009 年 8 月 18—20 日“第二届中国古籍数字化国际学术研讨会提交论文”(北京)会上发言。

贺铸使用 28 次,陈允平使用 19 次,晏几道使用 18 次,陆游使用 16 次,辛弃疾使用 16 次,周密使用 16 次,吴潜使用 12 次,李曾伯使用 12 次,刘辰翁使用 10 次,秦观使用 9 次,周邦彦使用 7 次,苏轼使用 4 次,姜夔使用 3 次。但是何原因,使不同作者对“镜”意象的使用出现这样的偏差,则还需进一步理论分析。

袁林先生曾用模糊数学中的模糊聚类分析和模糊相似优先比分析方法,对数据库检索进行分析,提出前资本主义社会主要的被剥削被压迫阶级中有一些游离于斯大林定义的奴隶和农奴之外,并不能为这两个概念所涵盖。^①北京大学李铎先生研发的《全唐诗 & 全宋诗分析系统》,以智能“分析”替代全文“检索”,在数据深层挖掘和知识发现方面具有开创性意义。将《全唐诗》57000 多首,《全宋诗》254240 首作为检索对象,可以直接完成重出诗提取、格律诗标注、字、词组和诗歌格律的频率分布统计等人力无法完成的课题。

古籍数据库的使用可以促进“群体传记”等关系网络类研究。群体传记学(Prosopography)兴起于 20 世纪二三十年代对欧洲历史的研究,它是指通过对一群人物生平的集体性研究,来探讨他们共同的“背景特征”,其采用的方法是:

先建立一个研究范围,然后询问一组系统的问题:关于出生与死亡、婚姻与家庭、社会出身与其所继承的经济地位、居住地、教育、个人财富的数量与来源、职业、宗教、仕历等等。之后将这一范围内所有个人的各种信息对比、组合,并探析其重要的变量。研究者会分析这些信息的内在相关性,及它们与其它类型的行为与活动的相关性。^②

英国学者郝若贝教授《750 至 1550 年间中国的人口、政治和社会转型》一文是“群体传记”典型成果,文章提出:宋初的统治集团,是一群所谓的开国功臣;到 10 世纪末,统治人群便被一群职业官僚所替代;而到了 11 世纪晚期,这群人又被所谓的“地方精英”所取代^③。现在哈佛大学与北京大学等单位建设的《中国古代人物传记资料库》(CBDB)就尤为注重输入人物的各种关系,从而为深入研究提供数据支持,可以实现学者的相关设想:“以某一作家的生平家世为焦点,可以迅速排列比较同时代和异时代的多个、甚至多组作家的类似背景资料,从而发现其中各种关联和变化的脉络。^④”

① 袁林:《中国古代奴婢性质问题的模糊数学分析》,《河北师院学报》1993 年第 3 期。

② Lawrence Stone, "Prosopography", Daedalus 100.1 (1971), p 46.

③ Robert M. Hartwell, "Demographic, Political and Social Transformations of China, 750—1550", Harvard Journal of Asiatic Studies, December 1982, pp. 405—416.

④ 李铎、王毅:《数据分析时代与古典文学研究的开放性空间——兼就信息化工程与古典文学研究之间的互动问题答质疑者》,《中国文化研究》2006 年第 2 期。

七、思维双向化

古籍数字化时代,不仅为研究者提供了前所未有的史料检索的方便,更重要的是,还可以促进了学者群体的思维转化。

首先,数字化时代的学者可以实现从“一意求之”到“无意求之”的选题转变。人类习惯于直线式思考,是以 $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$ 的顺序依次排下去的逻辑,我们通常会顺着这些线条来收集史料,沿着线条进行思考。苏东坡在《又答王庠书》里说:“书富如海,百货皆有。人之精力,不能兼收尽取,但得其所求者尔。故愿学者每次作一意求之。如欲求古今兴亡治乱,圣贤作用,但作此意求之,勿生余念。又别作一次,求事迹故实;典章文物之类,亦如此。他皆仿之。”这种“一意求之”读书法实际上也是集中精力的研究路径:“每当我们把零星的断片(即史料)按照一定规则重新排列、组合以后,都会有一种豁然开朗的感觉,因为我们从中发现了那些资料在原有脉络之中难以解读出的字面之外的第二甚至第三重含义,以及它们之间的各种内在关联,我们对这些含义和关联作进一步的分析或综合,往往会有新的发现和解读,这就是人文学术研究的一般过程。^①”

演绎思维要求从客观实际出发,搜集大量的事实材料,将此作为出发点(起点),然后遵循传统的逻辑规则,沿着归一的或单一的方向进行严谨周密的推理论证,进行分析、综合、抽象、概括,揭示客观事物的本质及其规律性,必须一环扣一环地展开研究,特别重视因果链条,要求实事求是,不允许用联想和想象代替推理和论证,更不允许出现跳跃。

在数字化时代,学者对所遇问题没有研究也可以试着检索,从而实现“无意求之”的改变。数字化时代的史学研究,有的问题可以边研究边查史料,王曾瑜先生写《宋朝兵制初探》时,还没有古籍数据库,但对史料分布有深入,也是一边写一边找材料,成为宋代军事史研究的重要参考著作。而在数字化时代,使用模糊查询或渐进式查询,更有可能如此进行研究。更重要的是,以前学者多将精力耗费于史料搜集工作,对于一个问题的诸多答案、办法和方案中,只能选择一个最好的。而“无意求之”,则可以在研究中多思路地进行思考,不断探索其他的可能性,从内心深处把齿轮从单数变成复数,从而把自己从一个问题的解决者转变为一个问题们的解决者,或称为多问题的解决者。

“无意求之”的研究路径,使学术研究在进行分析之前,可以先进行假设,“以果为始”,以最终的结果作为思考的开始,以最初需要的数据或者信息为这个循环思维的终点。在思考一个重要问题,或者做出重要结论前,先确定真正的问

① 史睿:《试论中国古籍数字化与人文学术研究》,《国家图书馆学刊》1999年第2期。

题,思索一下所有的相关因素,进行清理。然后针对重点提出假设,并明确支持或者推翻假设需要的数据,然后明确目的地,收集数据。这种“假设—检验”的思维模型,使胡适先生讲的“大胆设想,小心求证^①”成为可能。

演绎思维与以果为始思维各有缺点,前者因为依托逻辑,只是沿着一条固定的思路走下去,容易使我们的思考受拘泥、被定型、局限住;后者往往主题先行,即先有题目再去论证,容易造成伪命题的出现。但是在数字化时代,可以兼用两种思维,相向思考,势必大于裨益于学术创新。

尾 语

数字化时代的另一面是网络的迅速发展,网络的无与伦比的交互性发展,现在已在很大程度上取代传统的论著索引工具书,而成为发布学术成果的主要途径。这体现在两个方面,一是研究动态近期化,要求对研究问题进行的学术史回顾到最近时期;二是研究情报多样化,除了论著之外,学术会议、学位论文、调研报告、讲演发言、网络博客等非正式成果也可列入参考文献。中国学术期刊全文数据库是国内较通用的全文库。超星电子图书馆为目前世界最大的中文在线数字图书馆。万方数据是国内第一家以信息服务为核心的股份制高新技术企业。随着各类研究成果的发展,穷尽相关研究参考文献也成为可能,学术创新的难度也相应提高,有的问题可能进入微创新时代。

2000年1月美国公布了《高等教育信息素养(informationLiteracy)能力标准》,确定了一个具备信息素养的人应达到的目标是:“确定所需信息的范围;有效地获取所需的信息;鉴别信息及其来源;将检索出的信息融入自己的知识基础;有效地利用信息去完成一个具体任务。”在数字化时代,学会利用数据库和学术信息搜索都成为研究生必备的研究素养。

当然,数字化时代对学术研究也会带来负面影响,一是会有“以检索代替研究”、技术伪装学问的诟病^②。二是只重史料,而不愿理论思考,从而实质上重复前人的研究成果,漆侠先生认为这样的研究终为辕下之驹,这是我们应该避免的。

【作者单位】 河北大学

① 胡适:《清代学者的治学方法》,欧阳哲生编:《胡适文集》第二册,北京大学出版社,1998年版,第285、302页。

② 吴夏平:《古籍数字化与学术异化》,2011年8月16—18日“第三届中国古籍数字化国际学术研讨会提交论文”(北京)提交文章。